STATISTICA™

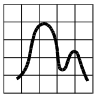
Том I:
ОСНОВНЫЕ СОГЛАШЕНИЯ И СТАТИСТИКИ I

1. Основные соглашения 1001
2. Панели инструментов и Строка состояния 1049
3. Окно таблицы с исходными данными 1123
4. Окно таблицы результатов Scrollsheet 1245
5. Окно текста/вывода 1273
6. Кнопки автозадач 1293
7. Управление данными 1315
8. Основные статистические понятия 1425
9. Основные статистики и таблицы 1439
10. Быстрые основные статистики 1573
11. Непараметрическая статистика и распределения 1601
12. Множественная регрессия 1653
13. Общая ANOVA/MANOVA 1709

Приложения 1827

Литература 1897

Том I:	СОГЛАШЕНИЯ И СТАТИСТИКИ I
Том II:	ГРАФИКА
Том III:	СТАТИСТИКИ II
Том IV:	ПРОМЫШЛЕННЫЕ СТАТИСТИКИ
Том V:	ЯЗЫКИ: BASIC и SCL



StatSoft®

Глава 1:
ОСНОВНЫЕ СОГЛАШЕНИЯ

Оглавление

Интерфейс пользователя системы *STATISTICA* (обзор) 1005
Настройка операций и внешнего вида пакета *STATISTICA* 1019
Интерфейс клавиатуры 1031
Соглашения по использованию мыши 1034
Другие диалоговые окна и параметры 1041
Сервер файлов системы STATISTICA 1045
Предметный указатель 1047

Подробное оглавление

ИНТЕРФЕЙС ПОЛЬЗОВАТЕЛЯ СИСТЕМЫ *STATISTICA* (ОБЗОР) 1005
Основные характеристики 1005
Индивидуальный режим работы 1005
Различные способы доступа к одним и тем же возможностям; пользовательский стиль работы 1005
1. Интерактивный пользовательский интерфейс 1006
Введение 1006
Модульная структура 1006
“Сеанс” интерактивного анализа 1008
Автоматическое создание отчетов (текстов и графиков) и автоматизация интерактивного анализа 1009
Окна документов в системе *STATISTICA* 1011
2. Макрокоманды 1012
3. Командный язык системы *STATISTICA* (*SCL*) 1013
Ввод и исполнение *SCL*-программ 1013
Пользовательские расширения языка *SCL* 1014
Пользовательский интерактивный интерфейс для *SCL*-программ 1014
SCL-программы “под ключ” 1014
4. Работа с пакетом *STATISTICA* из среды других Windows-приложений 1015
Кнопки автозадач - всплывающая или зафиксированная панель инструментов 1015
НАСТРОЙКА ОПЕРАЦИЙ И ВНЕШНЕГО ВИДА ОКНА ПРИЛОЖЕНИЯ *STATISTICA* 1019
Настройка пользовательского интерфейса 1019
Работа в одном окне приложения системы *STATISTICA* (вместо многооконного режима) 1020
Локальные и постоянные настройки (меню *Вид* и меню *Сервис*) 1021

Copyright © StatSoft, 1995

Общие настройки по умолчанию и параметры по умолчанию для графиков 1022
Поддержка нескольких различных конфигураций системы *STATISTICA* 1023
Пользовательские конфигурации для индивидуальных пользователей в сети 1023

ВЫБОР И ПРОСМОТР ПЕРЕМЕННЫХ И ЗНАЧЕНИЙ 1025
Выбор переменных для анализа 1025
Выбор переменных 1025
Быстрый способ выбрать одну переменную и выйти 1025
Отмена выбора ранее выбранных переменных 1025
Выбор переменных из одного списка 1026
Выбор переменных 1026
Выбрать все 1026
Подробно/Кратко 1026
Инф. 1026
Выбор переменных из двух (нескольких) списков 1027
Выбор переменных 1027
Выбрать все 1027
Подробно/Кратко 1027
Инф. 1027
Выбор одной переменной из списка 1027
Выбор переменной 1027
Подробно/Кратко 1027
Инф. 1028
Просмотр переменных из списка 1028
Подробно/Кратко 1028
Инф. 1028
Окно значений переменной 1028
Значения переменной 1028
Описательные статистики 1028
Выбор кодов 1029
Метод *Перетащить и отпустить* в системе *STATISTICA* 1029

ИНТЕРФЕЙС КЛАВИАТУРЫ 1031
Глобальные «горячие» клавиши 1031
Локальные ‘горячие’ клавиши 1032
Окно таблицы исходных данных 1032
Окно таблицы результатов 1032
Окно текста/вывода, окна программ на языках *STATISTICA BASIC* и *SCL* 1032
Графическое окно 1033
Окно редактора данных графика 1033
Окно *Менеджера мегафайлов* 1033

СОГЛАШЕНИЯ ПО ИСПОЛЬЗОВАНИЮ МЫШИ 1034
Кнопки микропрокрутки 1034
Левая и правая кнопки мыши 1034

StatSoft®

Copyright © StatSoft, 1995

Левая кнопка мыши ... 1034
Правая кнопка мыши ... 1035
Выбор элементов из списков с множественным выбором ... 1035
Использование клавиш SHIFT и CTRL в сочетании с мышью ... 1035
Другие применения ... 1036
Графические применения ... 1037
Остановка или прерывание текущей операции ... 1039

ДРУГИЕ ДИАЛОГОВЫЕ ОКНА И ПАРАМЕТРЫ ... 1041
Системное меню окна приложения ... 1041
Восстановить ... 1041
Переместить ... 1041
Размер ... 1041
Свернуть ... 1041
Развернуть ... 1041
Закрыть ... 1041
Переключиться в ... 1041
Повторное открытие файла, сохраненного в другом модуле ... 1041
Сохранение измененного файла перед открытием его в другом модуле ... 1042

***СЕРВЕР ФАЙЛОВ СИСТЕМЫ STATISTICA* ... 1045**
Пользовательская разработка новых модулей в системе *STATISTICA* ... 1045

ПРЕДМЕТНЫЙ УКАЗАТЕЛЬ ... 1047

Copyright © StatSoft, 1995

StatSoft®

Copyright © StatSoft, 1995

Глава 1:

ОСНОВНЫЕ СОГЛАШЕНИЯ

ИНТЕРФЕЙС ПОЛЬЗОВАТЕЛЯ СИСТЕМЫ *STATISTICA* (ОБЗОР)

Большая часть содержания этой главы вошла также в *Краткое руководство*, в котором основные характеристики и возможности системы *STATISTICA* изложены в форме вопросов-ответов. Данная глава представляет собой полное справочное руководство, в котором дается описание всех диалоговых окон.

Основные характеристики

Индивидуальный режим работы

В системе *STATISTICA* предусмотрено несколько способов управления системой. В следующих параграфах кратко описаны четыре основных пользовательских интерфейса: (1) интерактивный интерфейс, (2) макрокоманды, (3) командный язык *SCL* и (4) управление из других приложений Windows. При этом нужно иметь в виду, что:

- во многих отношениях эти варианты пользовательского интерфейса не исключают друг друга; поэтому их можно использовать в различных комбинациях в зависимости от конкретной ситуации и текущих потребностей;
- инструмент *Кнопки автозадач* (всплывающие или зафиксированные панели инструментов с кнопками, которые определяет сам пользователь, и которые могут соответствовать конкретным наборам действий, пользовательским процедурам или последовательностям клавиш) может использоваться для совмещения разных вариантов интерфейса, например, для быстрого доступа к пакетам заданий или наиболее часто используемым файлам;
- почти все характеристики этих интерфейсов определяются соответствующими настройками по умолчанию, и эти настройки можно менять по своему желанию (тем самым, изменяя внешний вид и действия программы); как правило, пользователю рекомендуется настроить систему так, чтобы полностью использовать возможности пакета *STATISTICA* с учетом своих вкусов и специфики решаемых задач (см. посвященный этому раздел *Настройка интерфейса пользователя*, стр. 1019).

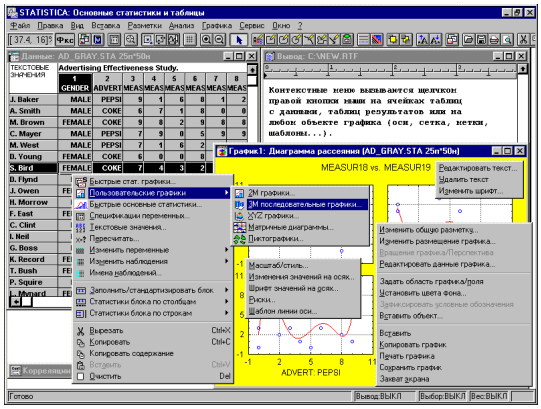
Различные способы доступа к одним и тем же возможностям; пользовательский стиль работы

Даже без предварительной настройки, конфигурация системы *STATISTICA* предлагаемая по умолчанию, дает пользователю возможность получать нужные ему результаты в рамках различных вариантов интерфейса. Благодаря принципу “альтернативного доступа”, последовательно и всесторонне реализованному в пользовательском интерфейсе пакета, система *STATISTICA* обладает возможностью поддерживать различные стили работы. Так, например, доступ к наиболее часто используемым средствам пакета *STATISTICA* осуществляется любым из следующих способов:

- через обычные выпадающие меню,
- через последовательности «горячих» клавиш,
- с помощью панелей инструментов (“линеек”) и активных полей строки состояния,

Copyright © StatSoft, 1995
StatSoft®

- через инструмент *Кнопки автозадач* (всплывающие или зафиксированные панели инструментов с кнопками, которые определяет сам пользователь, и которые могут соответствовать пользовательским процедурам или последовательностям клавиш),
- через контекстные меню, связанные с конкретными объектами (ячейками, графическими объектами, частями графических окон), которые вызываются нажатием правой кнопки мыши на данном объекте (см. далее контекстно-зависимые меню).

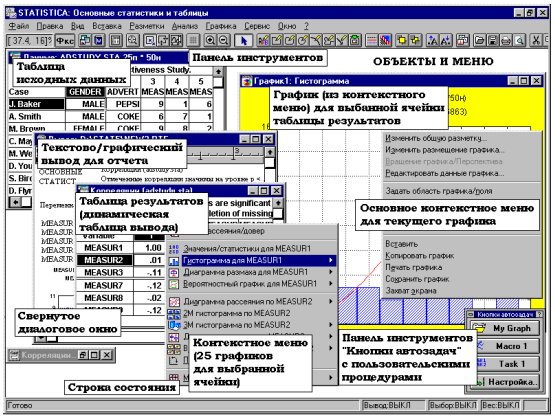


Прежде чем выбрать какой-то один определенный стиль, имеет смысл попробовать различные варианты пользовательского интерфейса системы *STATISTICA*.

1. Интерактивный пользовательский интерфейс

Введение

Основные компоненты интерактивного пользовательского интерфейса в системе *STATISTICA*. Хотя интерактивный пользовательский интерфейс - не единственный в системе *STATISTICA* (см. далее разделы 2, 3, 4 про альтернативные варианты интерфейса), он наиболее простой и, как правило, именно он используется чаще всего.



На этом рисунке представлены основные элементы окна приложения системы *STATISTICA* (как правило, все представленные здесь инструменты и окна не выводятся на экран одновременно). Большая часть показанных здесь инструментов и средств описаны в последующих разделах этого руководства (материал организован таким образом, что каждый раздел посвящен какой-то одной панели инструментов).

Модульная структура

Модули. Система *STATISTICA* состоит из набора модулей, в каждом из которых собрана тематически связная группа процедур. При переключении модулей можно либо оставлять открытым только одно окно приложения *STATISTICA*, либо все вызванные ранее модули, поскольку каждый из них может выполняться в отдельном окне (как самостоятельное приложение Windows, см. ниже).

При работе с модулями *STATISTICA* как самостоятельными приложениями в любой момент времени и в любом из них сохраняется прямой доступ к “общим” возможностям (таблицам данных, языкам *BASIC* и *SCL*, графическим процедурам).

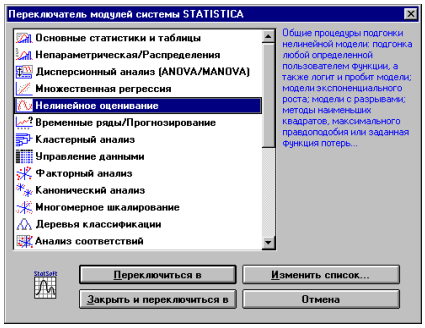
StatSoft®

Copyright © StatSoft, 1995

Архитектура пакета *STATISTICA* и реализованный в нем аппарат многозадачности позволяет не только без каких-либо усилий сравнивать результаты анализа, полученные разными процедурами, но и одновременно выполнять их в разных окнах приложения.

Быстро переключаться с одного модуля на другой можно:

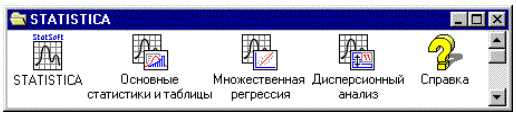
- щелкая мышью на их значках на рабочем столе,
- активируя соответствующее окно приложения (если оно уже было открыто), или
- выбирая их в диалоговом окне *Переключатель модулей*, причем эту операцию можно настроить так, чтобы было удобно обращаться к модулям, которые используются чаще всего.



Переключатель модулей. *Переключатель* работает так же, как и *Переключатель задач* системы Windows (который вызывается двойным щелчком на любом месте пустого пространства рабочего стола Windows). Переключатель *STATISTICA* вызывается аналогично - двойным щелчком где-нибудь в пустом месте окна приложения *STATISTICA*. Другие способы вызвать его - нажать первую (слева) кнопку на любой панели инструментов или выбрать пункт *Другие виды анализа* в меню *Анализ*.

В диалоговом окне *Переключатель модулей* новый модуль можно открыть в том же самом или в новом окне приложения (см. стр. 1020).

Значки модулей системы *STATISTICA*. При первичной инсталляции системы *STATISTICA* программа установки создает на рабочем столе группу приложений под названием *STATISTICA* и помещает туда значок окна *Переключатель модулей* (значок *STATISTICA* - первый в группе на следующем рисунке), модуля *Основные статистики и таблицы* и некоторых других программ (*Help*, *Setup*).



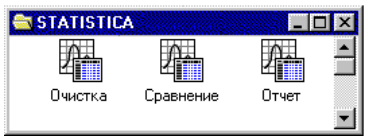
Как уже отмечалось, пользователю может показаться более удобным запускать модули, щелкая по их значкам на рабочем столе (вместо того чтобы вызывать окно *Переключатель модулей*); поэтому он, вероятно, захочет создать дополнительные значки для модулей помимо тех, которые будут автоматически созданы программой установки. Для того, чтобы создать еще одну пиктограмму в данной группе, следуйте стандартной процедуре Windows (выберите пункт *Новый (New)* в меню *Файл (File)* в окне *Диспетчер программ (Program Manager)* и создайте новый *программный элемент (Program Item)*).

Обратите внимание, что при создании нового *программного элемента (Program Item)* (пиктограммы) в группе автоматически будет создан значок соответствующего модуля и помечен сокращенным названием (шириной не

Copyright © StatSoft, 1995
StatSoft®

больше, чем значок). Если какие-то из полученных таким образом названий окажутся недостаточно информативны, их можно расширить или отредактировать в поле *Описание (Description)* в окне *Свойства программного элемента (Program Item Properties)* системы Windows.

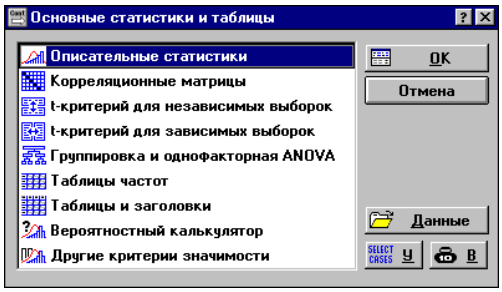
Пользовательская настройка значков. Вы можете настроить значки приложений таким образом, чтобы они ассоциировались с конкретными файлами данных (для этого нужно ввести название файла в поле *Командная строка (Command Line)* в окне *Свойства программного элемента (Program Item Properties)* системы Windows), и тогда при щелчке на этом значке будет открываться нужный файл данных.



Можно создать несколько значков для одного модуля как показано выше на рисунке, так, чтобы каждая ассоциировалась со своим файлом входных данных (они различаются по разным пользовательским именам значков в их окнах *Свойства программного элемента (Program Item Properties)*, в поле *Описание (Description)*).

"Сеанс" интерактивного анализа

Стартовая панель. При запуске модуля открывается его стартовая панель, например:

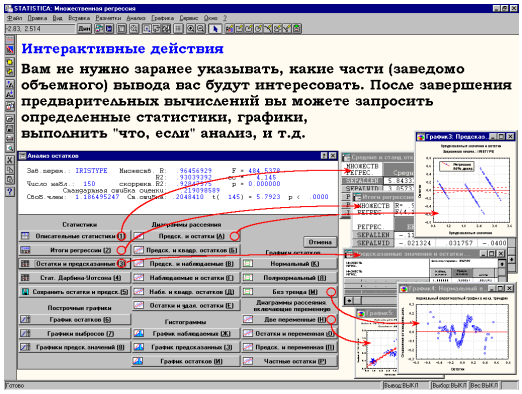


Эта панель содержит средства прямого доступа к тем возможностям данного модуля (стандартным образом доступным через меню), с которых пользователь чаще всего начинает работу. Эта панель не позволяет обратиться к тем пунктам меню, которые требуют совершения каких-то предварительных действий. (Наши наблюдения показали, что стартовая панель значительно сокращает число операций в типичном сеансе работы с системой *STATISTICA*, особенно в тех случаях, когда анализ складывается из многих действий.) Однако при желании пользователь может воспользоваться и обычными меню (выпадающие меню *Анализ* и *Файл* содержат все команды стартовой панели), поскольку стартовая панель - это немодальное диалоговое окно, которое сворачивается в значок при щелчке где-либо вне его (вывод этого диалогового окна можно запретить, см. стр. 1019).

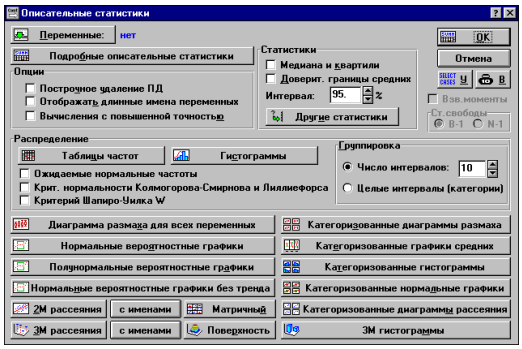
Панель инструментов *Кнопки автозадач*, дублирующая функции стартовой панели. В качестве альтернативного, и притом настраиваемого варианта стартовой панели может быть использована всплывающая панель *Кнопки автозадач* (см. главу 6). Если она не отключена специально, то появляется при запуске модуля; поэтому ее кнопки могут быть заданы таким образом (с помощью кнопки *Настройка…*), чтобы они обозначали действия, которые обычно выполняются сначала. Эти действия могут быть заданы через последовательность нажатий клавиш или с помощью команд языка *SCL* (*Командного языка системы STATISTICA*).

Окна "постановки задачи " и "задания формата вывода ". После того, как на стартовой панели будут выбраны метод анализа и (если требуется), файл данных, открывается окно постановки задачи, в котором можно выбрать подлежащие анализу переменные и другие параметры задачи.

StatSoft®
Copyright © StatSoft, 1995



В простых случаях (например, для описательных статистик, см. рисунок), окно постановки задачи служит одновременно и для задания формата вывода, и здесь можно задать нужный вид и формат выходной информации (например, конкретные таблицы результатов или графики).



Вывод (очередь таблиц результатов и графических окон). Последовательно выводимые таблицы результатов и графики появляются в отдельных окнах и образуют очереди фиксированной длины, которыми можно управлять (т.е. автоматически закрывать) по принципу "первым пришел - первым ушел".

По умолчанию длина обеих очередей устанавливается равной трем. Иначе говоря, при создании четвертой таблицы результатов (или графика) первая закрывается. Длину каждой из очередей можно изменять (временно - через меню *Окно* или постоянно - через меню *Сервис*).

Независимо от принятой длины очереди, можно *заблокировать* отдельные окна, запретив "удалять" их из очереди (для этого служат две команды *Диспетчер*...из меню *Окно*). Заблокированные окна не будут закрываться до тех пор, пока не будут разблокированы, закрыты вручную или пока не будет закончена работа с программой. Все таблицы результатов и графики управляются своей соответствующей очередью независимо от того, как они были получены (они могут строиться программой автоматически или самим пользователем непосредственно из таблиц исходных данных, таблиц результатов, меню и диалоговых окон).

Автоматическое создание отчетов (текстов и графиков) и автоматизация интерактивного анализа

В дополнение к альтернативным пользовательским интерфейсам, основанным на макрокомандах (см. далее), в системе *STATISTICA* реализованы различные возможности для автоматизации последовательностей операций при анализе и/или для автоматической подготовки результатов, при которой не требовалось бы отдельно посылать каждую порцию результатов (таблиц результатов, графиков) на печать или в окно отчета (или в файл).

Кнопки автозадач. Эта всплывающая или зафиксированная панель инструментов (см. стр. 1015) содержит определенные пользователем кнопки, соответствующие стандартным задачам, операциям или пользовательским процедурам (используйте кнопку *Настройка...* для назначения кнопкам новых процедур: программ на *Командном языке системы STATISTICA* (*SCL*), программ на языке *STATISTICA BASIC* или просто последовательностей нажатий клавиш). После этого простым нажатием кнопки можно будет выполнить задачу или

Copyright © StatSoft, 1995
StatSoft®

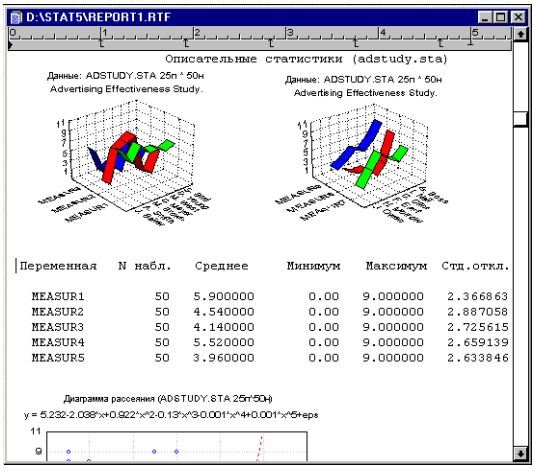
последовательность задач практически неограниченной сложности.

Внутренний режим пакетной обработки. Там, где это возможно, система *STATISTICA* предоставляет пользователю *внутренний режим пакетной обработки*. Этот режим используется чаще всего при выборе для анализа не одной переменной, а некоторого списка. После этого заданная схема действий автоматически повторяется для каждой переменной из выбранного списка (например, для всех перечисленных зависимых переменных может быть выполнена одна и та же схема вычислений множественной регрессии, разбиения, *t*-критерия или критерия Краскела-Уоллиса). В таких процедурах обычно бывают реализованы возможности *внутренней пакетной обработки и печати*, позволяющие выполнять заданную последовательность действий, не требуя от пользователя ввода какой-либо информации, а затем автоматически посылать результаты на печать, в *Окно текста/вывода* или в файл на диске.

Автоматические отчеты и автоматическая распечатка таблиц результатов. Независимо от того, организованы ли этапы анализа автоматически в режиме пакетной обработки (см. предыдущий пункт) или они интерактивно запрашиваются пользователем, может быть выбран режим *Автоотчет* (см. главу 2, описание меню *Сервис* или активного поля *Вывод* в строке состояния). Этот режим позволяет автоматически, без каких-либо действий пользователя распечатывать (или направлять в окно отчета или файл) содержание всех окон вывода, которые получаются в процессе анализа.

Режим, позволяющий автоматически порождать соответствующую запись для каждой таблицы результатов и/или графика, строящихся на экране, может оказаться полезным не только в тех случаях, когда необходимо иметь полную запись выводимой информации; например, при разведочном анализе данных иногда бывает полезно вернуться к более раннему шагу в последовательных этапах анализа. Для этого всю выходную информацию (таблицы результатов и графики), вместо того, чтобы полностью ее распечатывать, можно направить во временное *Окно текста/вывода*, и уже затем, в случае необходимости, ее можно сохранить, распечатать или скопировать в файл текстового редактора (см. главу 5).

Автоматическая печать графиков. Организация автоматической печати всех возникающих на экране графиков (см. активное поле *Вывод* в строке состояния; глава 2) особенно полезна как средство пакетной графической печати. Как правило, изготовление твердых копий графиков занимает довольно много времени. Поэтому имеет смысл воспользоваться этой возможностью для распечатки последовательности (“каскада”) графиков, получающихся в процессе анализа определенным методом (например, для зрительного представления конфигурации средних для связей высших порядков в дисперсионном анализе необходима длинная последовательность графиков, а для многомерных таблиц - каскад трехмерных гистограмм для двух переменных).



StatSoft
Copyright © StatSoft, 1995

Однако, гораздо быстрее будет направить генерируемую последовательность графиков непосредственно в *Окно текста/вывода* (см. главу 5). (Имейте в виду, что имеется также возможность пакетной печати для распечатки всех ранее сохраненных графиков и таблиц результатов; для этого нужно выбрать пункт *Печать файлов* в меню *Файл*, см. главу 3.)

"Автоматическое нажатие" кнопки *Далее* для ускорения пакетного вывода. Если необходимо, не просматривая экраны вывода, быстро получить твердые копии графиков и таблиц (или послать их в *Окно текста/вывода*), то выгоднее будет выбрать режим *Авт. выход из таблиц результатов и графиков* (см. главу 3) в окне *Параметры страницы/вывода*. При выборе этого режима система *STATISTICA* будет "сама" нажимать кнопку *Далее* на каждом графике или таблице результатов, позволяя тем самым распечатывать (или пересылать в *Окно текста/вывода*) большие последовательности таблиц результатов и графиков без необходимости нажимать кнопку *Далее* всякий раз, когда очередь документов "заполнится" (по умолчанию это будет происходить на каждом третьем графике или таблице результатов).

Окна документов в системе *STATISTICA*

Четыре основных типа *окон документов* в системе *STATISTICA* (MDI). Система *STATISTICA* следует стандартным соглашениям MDI (*Многодокументный пользовательский интерфейс*). Каждое окно вывода рассматривается как отдельный *документ*, и с содержанием такого *документа* можно работать с помощью ряда стандартных процедур, в частности, редактировать, сохранять, открывать. Система *STATISTICA* поддерживает работу с четырьмя основными типами документов:

- таблицы исходных данных
- таблицы результатов
- графики
- *Окна Текста/вывода* (содержат отчеты, состоящие из текстов и графиков)

(см. иллюстрацию в начале этого раздела).

Совместимость *документов*. Каждый из четырех основных видов окон документов системы *STATISTICA* работает со своим типом "данных". При этом, однако, все они совместимы между собой не только в смысле оптимизированной поддержки буфера обмена Windows (что дает возможность автоматически преобразовывать данные одного типа в другой), но и через различные методы "конвертирования" документов как целого. Например, любой текстовый или числовой документ (или его часть) может быть несколькими способами преобразован в график. Текст может быть включен в рисунок, а график преобразован в текстовое (т.е. числовое) представление. Эти свойства архитектуры пакета *STATISTICA,* вместе с поддержкой методов *DDE* (специальный механизм обмена данными в системе Windows) и полной совместимостью с клиент-серверной технологией *OLE* (включая вложенные составные графические документы) предоставляют пользователю неограниченные возможности для анализа данных, проверки гипотез и представления полученных результатов.

Панели инструментов, зависящие от типа активного окна документа. Каждый из основных видов окон документов системы *STATISTICA* (см. выше) работает со своим типом данных, и потому с ним связаны различные возможности настройки и управления. Эти различия находят свое отражение в структуре панелей инструментов, имеющихся у окон разных типов. Кнопки панелей инструментов для каждого из четырех основных типов окон описаны в следующих частях документации:

- таблицы исходных данных - глава 3,

Copyright © StatSoft, 1995
StatSoft®

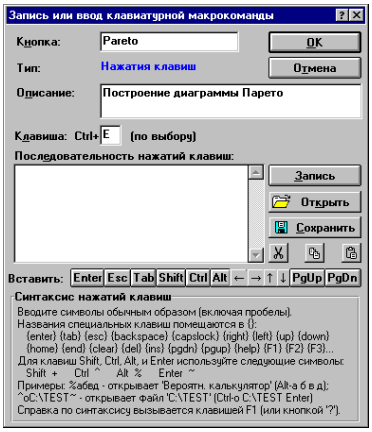
- таблицы результатов - глава 4,
- *Окна текста/вывода* - глава 5,
- графики - том II.

2. Макрокоманды

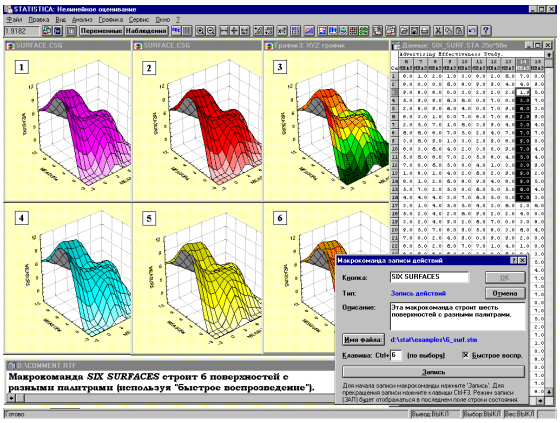
В пакете *STATISTICA* можно создавать макрокоманды двух видов: редактируемые последовательности нажатий клавиш и записи последовательностей действий (перемещений мыши и/или действий на клавиатуре).

Макрокоманды обоих видов можно создавать с помощью пункта *Макрокоманда* выпадающего меню *Сервис* или с помощью кнопки *Настройка* на всплывающей панели инструментов *Кнопки автозадач* (см. стр. 1015).

Редактируемые последовательности нажатий клавиш. Последовательность нажатий клавиш можно вводить (и редактировать) в текстовом формате и приписывать их кнопкам на всплывающей панели инструментов *Кнопки автозадач* (см. стр. 1015). Этот простой способ позволяет автоматизировать выполнение задач любой степени сложности (от таких простых, как выбор определенного набора переменных, до настройки сложных графиков, преобразования данных и длинных цепочек процедур). Уже существующие макрокоманды можно комбинировать друг с другом в специальном редакторе с кнопками *быстрого ввода*. В этом же редакторе можно открыть и редактировать записанные макрокоманды (см. следующий раздел).



Записи последовательностей действий. В дополнение к редактируемым последовательностям нажатий клавиш (см. предыдущий раздел) которые можно приписывать кнопкам на всплывающей панели инструментов *Кнопки автозадач* (см. главу 6), в пакете *STATISTICA* реализована внутренняя система записи макрокоманд,



реагирующая не только на действия клавиатуры, но и на движения мыши, которые, по желанию пользователя, можно воспроизводить с той же скоростью, с которой они были записаны, что позволяет создавать слайдовые презентации, учебные материалы и т.п.

StatSoft
Copyright © StatSoft, 1995

Доступ к средствам управления макросами осуществляется через пункт *Макрокоманда* из выпадающего меню *Сервис*, или нажатием клавиш CTRL+F3, или нажатием кнопки *Настройка…* на всплывающей панели инструментов *Кнопки автозадач* (см. главу 6).

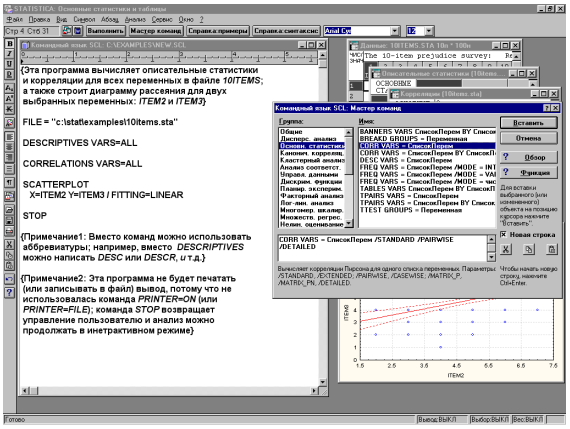
Приложения. Все эти возможности (доступные в любой момент работы с пакетом) могут служить весомой альтернативой или дополнением к обычному интерактивному пользовательскому интерфейсу, поскольку они позволяют автоматизировать рутинный процесс многократного выполнения одних и тех же, в том числе весьма сложных задач. Кроме того, макрокоманды расширяют возможности интерактивного пользовательского интерфейса; например, макрокоманда (которая запускается путем нажатия кнопки на панели *Кнопки автозадач* или одним нажатием клавиши) может содержать длинный список переменных, часто используемый график, вложенную операцию и т.д.

3. Командный язык системы *STATISTICA* (*SCL*)

Ввод и исполнение *SCL*-программ

Программа *STATISTICA* может работать в “истинном” пакетном режиме как система, управляемая командами, с помощью встроенного языка управления приложениями *SCL* (*STATISTICA Command Language*), доступного в любом модуле из меню *Анализ*. Для выполнения определенных действий можно ввести последовательность команд, записанную на языке, близком к английскому, а затем сколько угодно раз исполнять ее в пакетном режиме.

Возможен и другой способ действий - использовать для быстрого выбора нужных команд окно *Мастер команд* (см. иллюстрацию).



Для написания и отладки “пакетов” команд на языке *SCL* служит интегрированная среда. Эта среда включает совмещенный с окном *Мастер команд* текстовый редактор (см. иллюстрацию выше; описание кнопки *Мастер команд* на панели инструментов окна *Командный язык SCL*), справочную систему по синтаксису языка с примерами и интегрированные средства проверки правильности программ (доступны через меню *Сервис*).

Средства редактора для проверки правильности *SCL*-программ. При написании программ на *SCL* можно проверять не только формальную правильность команд, но и их соответствие содержанию файлов данных, которые предстоит обрабатывать (см. команды выпадающего меню *Сервис*). Например, непосредственно во время написания *SCL*-программы можно проверить, действительно ли переменная с таким именем или такое текстовое значение присутствуют в обрабатываемом наборе данных (подробнее см. интерактивную систему документации, которая вызывается клавишей F1 , кнопкой ? на панели инструментов или двойным щелчком на строке состояния в нижней части окна).

Прерывание исполнения *SCL*-программы. Исполнение *SCL*-программы может быть прервано щелчком мыши (любой ее

Copyright © StatSoft, 1995
StatSoft®

кнопкой) или клавишами ESC или CTRL+BREAK. При этом система *STATISTICA* запросит подтверждение на прерывание программы.

Пользовательские расширения языка *SCL*

Программы на языке *SCL* могут включать не только предопределенные команды и параметры для выполнения действий по статистической обработке, управлению и графическому выводу данных (см. кнопки *Справка: примеры* и *Справка: синтаксис* на панели инструментов), но и пользовательские “команды”, определенные с помощью инструмента *Назначить клавиши* (в соответствии с правилами, принятыми в MS Visual BASIC).

Например, программы, написанные на пользовательском расширении языка *SCL,* могут выполнять операции с буфером обмена (*Копировать*, *Вставить*), менять режимы вывода, принятые по умолчанию в различных процедурах, и т.д.

SCL-программы могут также включать в себя программы и процедуры, написанные на языке *STATISTICA BASIC* (языке системы *STATISTICA,* предназначенный для преобразования данных и графики и управления ими; язык доступен из любого модуля, см. *Электронное руководство*). Например, определенные пользователем графические или вычислительные процедуры на языке *STATISTICA BASIC* могут выполняться как часть пакета команд *SCL*.

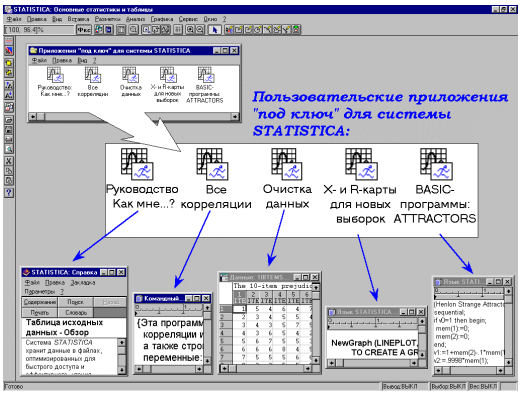
Пользовательский интерактивный интерфейс для *SCL*-программ

Несмотря на то, что в командном языке *SCL* не заложен (в непосредственном виде) специальный интерактивный пользовательский интерфейс, тем не менее, для обращения к диалоговым окнам и другим процедурам в процессе работы программы можно использовать программы на языке *STATISTICA BASIC* (вызываемые из *SCL*-программ), позволяющие выбирать переменные, файлы данных и т.д. (см. примеры в *Электронном руководстве*).

SCL-программы “под ключ”

***Модуль запуска* системы *STATISTICA*.** *Командный язык* содержит специальный *Модуль запуска (исполняемый модуль)*, позволяющий разрабатывать приложения “под ключ”, которые вызываются двойным щелчком по значку соответствующего “пользовательского приложения” на рабочем столе Windows.

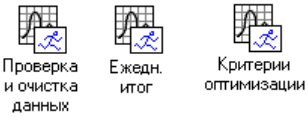
Эта возможность бывает полезна, когда одна и та же процедура или последовательность процедур анализа повторяются многократно, и дает возможность использовать *SCL*-программы даже тем, кто не знаком с соглашениями системы *STATISTICA*.



Чтобы создать такое приложение «под ключ», сначала нужно создать саму *SCL*-программу, которая должна при этом исполняться, и сохранить ее обычным образом (например, как файл *Program1.scl*). Затем с помощью окна *Диспетчер программ (Program Manager)* системы Windows создать значок для приложения *Модуль запуска STATISTICA* с именем *Sta_run.exe* (оно находится в папке *STATISTICA* на диске).

StatSoft
Copyright © StatSoft, 1995

В поле команд нужно задать имя *SCL*-программы, подлежащей исполнению (например, *d:\data\program1.scl*). Теперь, при щелчке мышью по этому значку программа (в данном случае *Program1.scl*) будет вызываться на исполнение. Описанным способом можно создать любое количество пользовательских приложений *STATISTICA*, а с помощью окна *Диспетчер программ (Program Manager)* дать им содержательные имена,

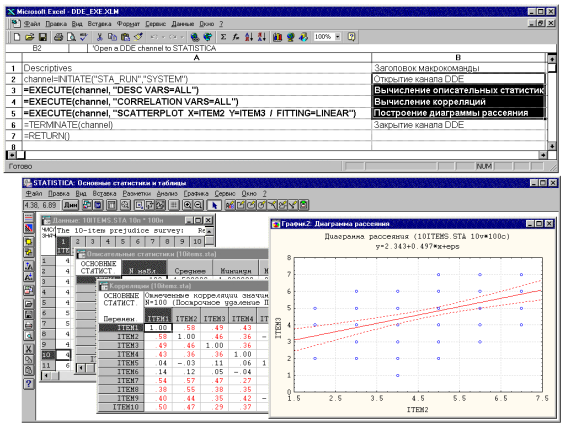


соответствующие тем задачам анализа данных, которые эти приложения выполняют.

Присваивание *SCL*-программ кнопкам на панели инструментов *Кнопки автозадач*. Определенные *SCL*-программы можно поставить в соответствие кнопкам на настраиваемой всплывающей панели инструментов *Кнопки автозадач* (см. ниже).

4. Работа с пакетом *STATISTICA* из среды других Windows-приложений

Поскольку система *STATISTICA* поддерживает стандартный интерфейс Microsoft Windows *DDE*/*API*, можно встраивать команды на языке *SCL* (*Командный язык STATISTICA*, см. предыдущий раздел) или целые *SCL*-программы в макрокоманды, созданные и выполняемые из других Windows-приложений (например, MS Excel и MS Word).



Свойства совместимости *Командного языка* системы *STATISTICA*, реализованная в системе *STATISTICA* поддержка динамического обмена данными (*DDE*) (см. главы 3 и 7) и методов *OLE* (клиент/сервер, вложенные составные графики, см. том II), а также совместимость со стандартными форматами файлов, буфера обмена и графическими форматами системы Windows позволяют достичь высокой степени интеграции между различными Windows-приложениями.

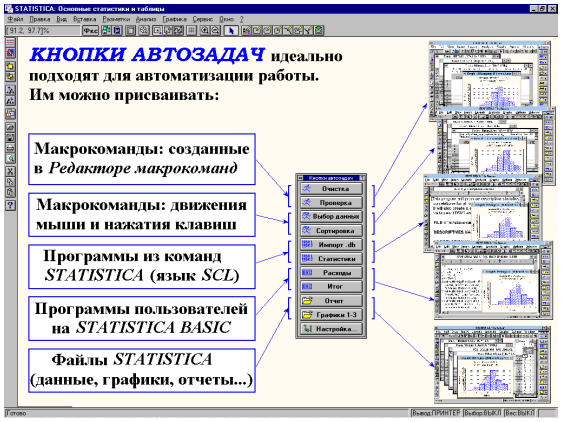
Например, на приведенной выше иллюстрации видно, что макрокоманда Excel может содержать команды на языке *SCL* , которые будут вызывать приложение *STATISTICA* из среды Excel, выполнять определенные операции по анализу данных, а затем преобразовывать нужные фрагменты вывода и/или графиков, выдаваемых пакетом *STATISTICA*, и вставлять их в рабочую книгу Excel.

Кнопки автозадач - всплывающая или зафиксированная панель инструментов

Кнопки автозадач - это всплывающая настраиваемая панель инструментов (включить/выключить ее можно клавишами CTRL+M).

Copyright © StatSoft, 1995

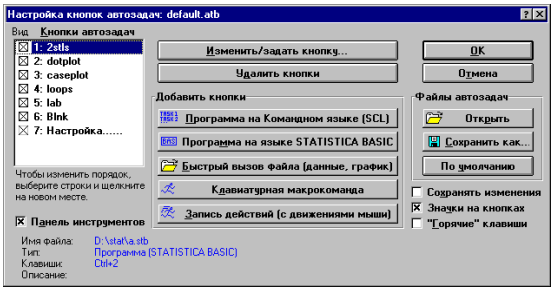
StatSoft®



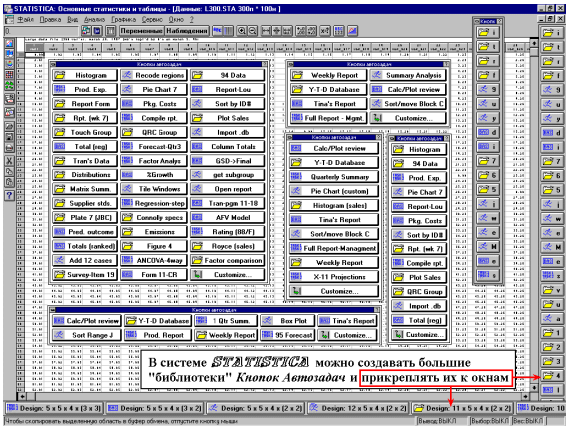
Кнопки на этой панели инструментов можно назначить/переопределить с помощью кнопки *Настройка…* (или нажимая на соответствующую кнопку при нажатой клавише CTRL). В диалоговом окне, которое при этом открывается, уже имеющимся и новым кнопкам можно присвоить:

- последовательности заданий или операций системы *STATISTICA*, заданных на *Командном языке STATISTICA* (*SCL*, см. стр. 1013);
- разработанные пользователем процедуры, написанные на языке *STATISTICA BASIC* (это могут быть процедуры вычислительного характера, преобразования данных, операции по управлению данными, пользовательские графические процедуры, а также процедуры, написанные на любом другом языке программирования и вызываемые из окна *STATISTICA BASIC* , где реализована поддержка внешнего динамического обмена данными, см. *Электронное руководство*);
- процедуры вызова файлов данных и любых вспомогательных файлов системы *STATISTICA* (графиков, отчетов, таблиц результатов), которыми приходится часто пользоваться (например, для справок) и к которым поэтому было бы удобно организовать прямой доступ, приписывая их *кнопкам автозадач*;
- редактируемые последовательности нажатий клавиш, соответствующие часто выполняемым процедурам, заданиям или настройкам. Такие редактируемые макрокоманды можно вводить в текстовом виде (следуя стандартным соглашениям *Sendkeys* фирмы Microsoft) и затем вносить в них изменения.

Настроенную панель инструментов *Кнопки автозадач* можно затем сохранить (в том же диалоговом окне *Настройка…*).

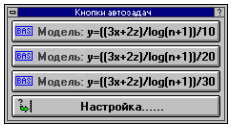


Панель инструментов может быть глобальной или же локальной – связанной с конкретным модулем или проектом (имя открытой в данный момент панели инструментов высвечивается в строке заголовка диалогового окна, см. выше), и может содержать большие библиотеки пользовательских заданий и процедур.



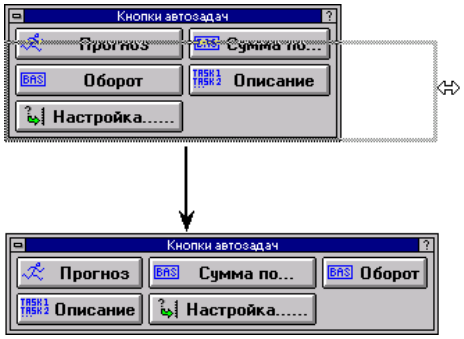
StatSoft
Copyright © StatSoft, 1995

Панель инструментов *Кнопки автозадач* можно использовать как удобный интерфейс для пользовательских расширений стандартных процедур,

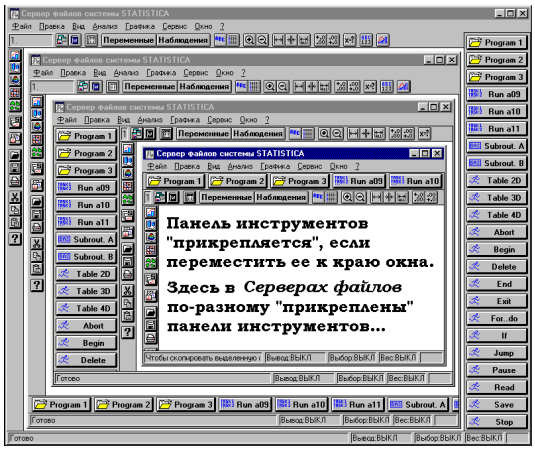


ее можно легко настроить так, чтобы она занимала очень мало места на экране.

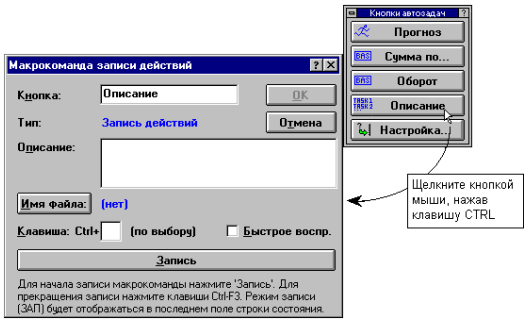
Размеры панелей инструментов можно менять с помощью мыши:



Панель можно зафиксировать, притащив ее к границе окна приложения *STATISTICA* (как показано на следующем рисунке).

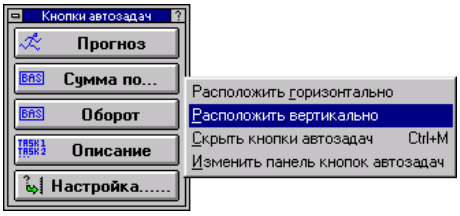


Как уже было отмечено, кнопки панели инструментов *Кнопки автозадач* можно настроить или переназначить в общем диалоговом окне *Настройка кнопок автозадач* (которое вызывается с помощью кнопки *Настройка...* на этой панели инструментов). Кроме этого, отдельные кнопки можно настроить и/или переназначить непосредственно в соответствующем окне настройки; для этого нужно щелкнуть мышью на этой кнопке при нажатой клавише CTRL:



При этом откроется окно настройки (данной конкретной кнопки).

Можно быстро переключаться между разными, предварительно сохраненными панелями инструментов *Кнопки автозадач*, выбирая последний пункт контекстного меню, которое появляется по щелчку правой кнопкой мыши где-либо на панели инструментов.



Более подробные сведения о панели инструментов *Кнопки автозадач* системы *STATISTICA* можно найти в главе 6.

Copyright © StatSoft, 1995
StatSoft®

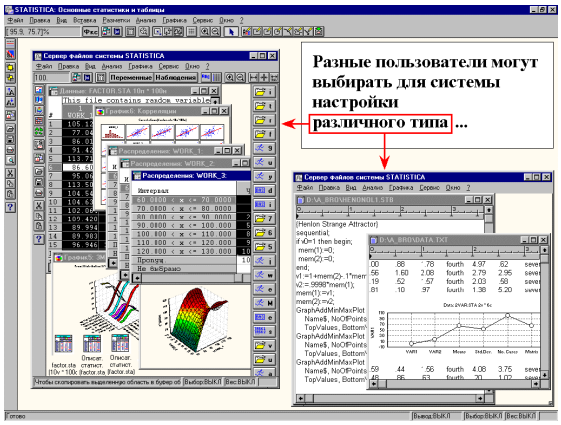
StatSoft®

Copyright © StatSoft, 1995

НАСТРОЙКА ОПЕРАЦИЙ И ВНЕШНЕГО ВИДА ОКНА ПРИЛОЖЕНИЯ *STATISTICA*

Настройка пользовательского интерфейса

Работа и внешний вид окна приложения системы *STATISTICA* могут быть настроены пользователем, а пользовательский интерфейс программы можно менять так, чтобы он стал более “продуманным” с точки зрения потребностей конкретного пользователя.

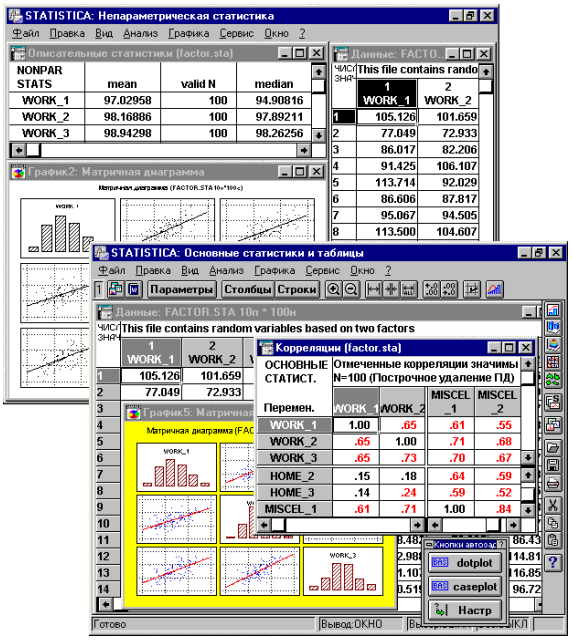


В зависимости от требований задачи и личных предпочтений относительно “режимов” работы (а также эстетических соображений) можно преобразовать окно приложения, например, одним из перечисленных ниже способов.

Убрать все значки, панели инструментов, строки состояния, длинные меню, всплывающие панели инструментов *Кнопки автозадач*, отключить Рабочую книгу, функции метода *Перетащить и отпустить*, динамические (автоматические) связи между графиками и данными, трехмерные эффекты в таблицах и диалоговых окнах и затребовать “голый” последовательный вывод с простыми черно-белыми таблицами результатов и графиками и приказать системе в каждый момент времени держать открытым не более одного простого окна вывода (см. левый экран на рисунке внизу).

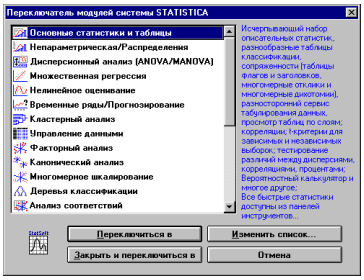
Возможен и другой вариант: создать сложные глобальные и локальные всплывающие панели инструментов *Кнопки автозадач*, использовать все преимущества всевозможных специальных инструментов и средств управления, пиктограммы, панели инструментов, макрокоманды (в частности, назначить некоторым клавишам специальные действия), Рабочие книги, метод *Перетащить и отпустить*, задать множественные динамические (автоматические) связи между графиками и данными и внутренние *OLE*-связи между графическими объектами, настроить цвета, специальные шрифты и затенения в окнах вывода, настроить стили и режимы вывода графиков, применяемые по умолчанию, увеличить максимальное число одновременно открытых окон вывода и *заблокировать* “дежурные” графики и таблицы результатов, - и в результате всего этого получить сложную многоуровневую среду для анализа данных, которая упростит анализ сложных совокупностей данных и позволит рассматривать и сопоставлять различные аспекты выходной информации (см. правый экран на рисунке внизу).

Copyright © StatSoft, 1995
StatSoft®

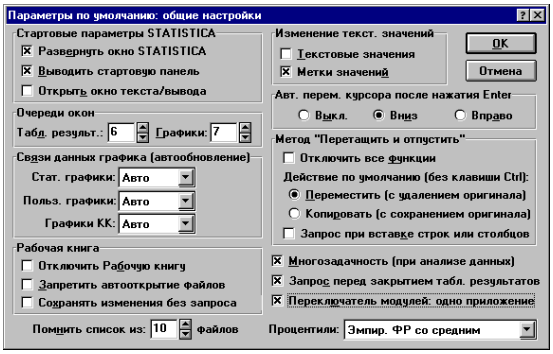


Работа в одном окне приложения системы *STATISTICA* (вместо многооконного режима)

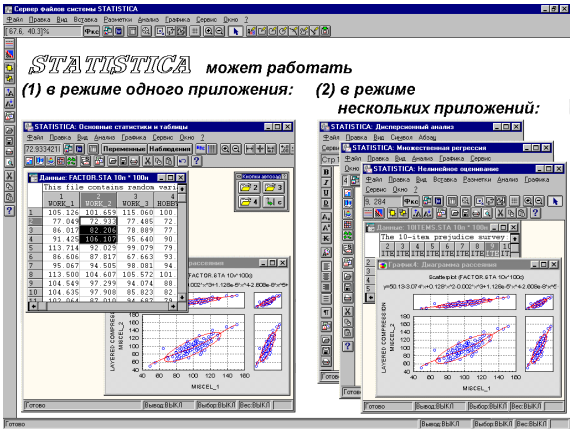
Один из вариантов глобальной системной настройки пакета *STATISTICA* позволяет пользователю задать режим, в котором по умолчанию будет работать программа: в одном окне приложения или же, как набор приложений (каждое в своем окне). Одним из непосредственных следствий этого выбора будет то, в каком режиме будет работать окно *Переключатель модулей* (см. главу 2): при двойном щелчке на названии модуля в этом окне данный модуль будет открываться либо вместо уже открытого, либо для него будет открываться новое окно приложения, а старое останется открытым.



Выбор того или другого режима работы производится в поле *Переключение модулей: режим одного приложения* в окне *Параметры по умолчанию: общие настройки* (доступ – через меню *Сервис*).



Если эта позиция отмечена, система *STATISTICA* будет работать в режиме одного приложения.



StatSoft®
Copyright © StatSoft, 1995

(1) Режим одного приложения. При выбранном режиме одного окна приложения переключение с одного модуля на другой при работе в системе *STATISTICA* будет происходить без открытия новых окон. Новый модуль всякий раз будет открываться в том же самом окне, заменяя предыдущий. Некоторые пользователи предпочтут такой "простой" режим, потому что весь анализ будет происходить в одном окне приложения, а количество активных программ на рабочем столе будет минимальным.

Примерно такого же эффекта можно достичь, нажимая кнопку *Закрыть и переключиться в...* в окне *Переключатель модулей*; при этом окно приложения текущего модуля закроется, но не будет заменено новым окном; вместо этого новый модуль откроет "следующее" окно приложения.

(2) Режим нескольких приложений. Основное преимущество режима нескольких приложений - возможность параллельно выполнять различные процедуры анализа (модули) в разных, одновременно открытых окнах приложения. При этом можно переключаться между модулями, не закрывая предыдущие, и использовать все преимущества работы с независимыми очередями таблиц результатов и графиков для окон приложений разных модулей. Этот режим имеет очевидные преимущества для большинства задач анализа данных и дает возможность использовать различные методы анализа (и сравнивать полученные результаты).

Локальные и постоянные настройки (меню *Вид* и меню *Сервис*)

Многие параметры общего вида программы можно настраивать и через выпадающие меню *Вид* и *Сервис*.

Меню *Вид*. Различие между этими двумя возможностями состоит в том, что изменения, сделанные в меню *Вид,* влияют только на текущий внешний вид программы (например, скрыть панель инструментов) или на активное окно документа (например, сменить используемый в таблице шрифт).

Меню *Сервис*. Многие из указанных возможностей изменения параметров доступны также в меню *Сервис,* однако, здесь они устанавливаются как постоянные параметры, используемые программой по умолчанию. При этом те параметры (из диалоговых окон меню *Сервис*), которые применимы только к окну документа определенного типа (например, графику или таблице исходных данных), не меняют уже готовый документ; вместо этого они сохраняются как программные параметры по умолчанию и вступают в силу только при создании *следующего* (то есть *нового*) объекта этого типа.

Так, например, при установке параметров по умолчанию для внешнего вида таблицы исходных данных в меню *Сервис* их можно будут увидеть только при создании нового файла (с помощью команд *Создать данные* или *Импорт*). Эти настройки, однако, не окажут никакого влияния на файлы, открываемые с диска, потому что эти файлы будут появляться в том виде, в каком они были сохранены (настроить уже существующие объекты можно с помощью команд меню *Вид*). Как правило, настройки по умолчанию для таблиц результатов и графиков имеют более непосредственный и заметный эффект, поскольку программа создает *новые*

Copyright © StatSoft, 1995

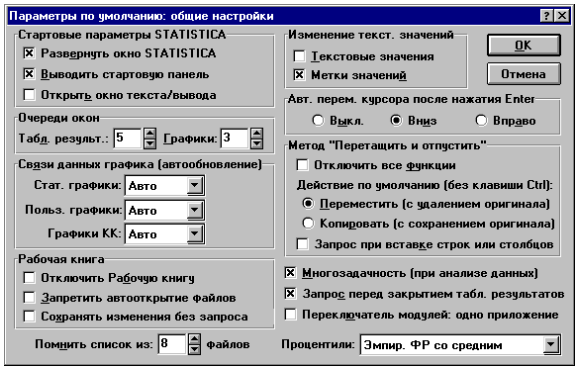
таблицы результатов и графики прямо во время сеанса работы.

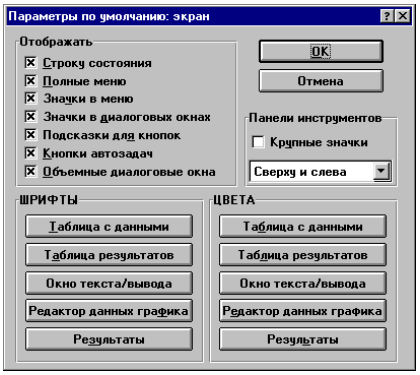
Общие настройки по умолчанию и параметры по умолчанию для графиков

Настройка общих параметров системы. Настройку общих параметров системы можно изменить в любой момент работы с программой. Эти параметры определяют некоторые общие характеристики системы:

- общие аспекты поведения программы (максимизация окна *STATISTICA* при запуске, Рабочие книги, инструмент *Перетащить и отпустить*, автоматические связи между графиками и данными, многозадачный режим и т.д.),
- режим вывода (например, автоматическая распечатка таблиц результатов или графиков, форматы отчетов, буферизация и т.д.),
- общий вид окна приложения (значки, панели инструментов и т.д.),
- вид окон документов (цвета, шрифты).

Каждый из этих параметров можно настроить в соответствующем окне, доступ к которому осуществляется через меню *Сервис*. Два примера таких окон показаны ниже.

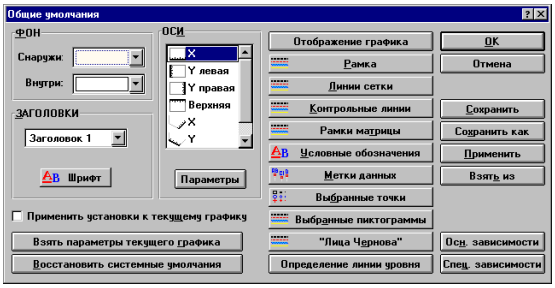




Все эти и другие общие параметры могут быть настроены независимо от типа окна документа, которое активно в данный момент (например, таблица результатов или график).

Настройка параметров по умолчанию для графиков. Графические средства системы *STATISTICA* также позволяют установить постоянные параметры по умолчанию; эти параметры (а всего их более 500) влияют практически на все аспекты представления графиков.

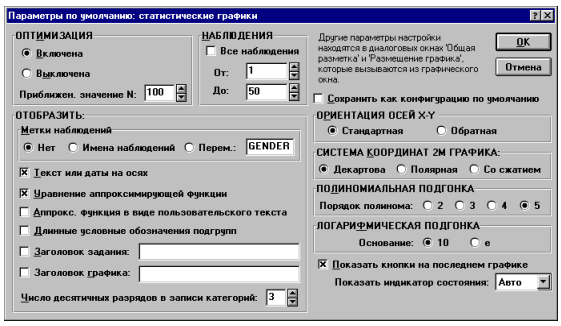
Поскольку многие из этих методов настройки по умолчанию являются специфическими для конкретных типов графиков и имеется возможность быстро скопировать все настройки из или в *активное* графическое окно, то окна, в которых производится настройка различных параметров по умолчанию для графиков, доступны через выпадающее меню *Сервис* только при активном графическом окне.



StatSoft®

Copyright © StatSoft, 1995

Единственным исключением из правила является окно, в котором задаются основные соглашения по доступу к данным из *статистических графиков* (пункт *Статистические графики*),



которое всегда доступно в этом меню, потому что пользователю может потребоваться изменить эти параметры до начала построения графиков.

Описание этих диалоговых окон дано в томе II.

Поддержка нескольких различных конфигураций системы *STATISTICA*

До внесения специальных изменений система *STATISTICA* будет хранить все текущие настройки и параметры по умолчанию, включая определения всплывающих панелей инструментов *Кнопки автозадач* в той папке, из которой вызывается программа (большинство настроек хранится в файле *statist.ini*, дополнительную информацию об этом файле см. в *Электронном руководстве*).

То обстоятельство, что сведения о конфигурации системы хранятся в той же папке, из которой вызывается программа *STATISTICA*, позволяет иметь в своем распоряжении различные варианты конфигурации программы для разных проектов или видов работ. Например, можно вызывать программу из разных папок на диске, каждая из которых содержит определенный связный набор документов, и для каждой из этих папок система *STATISTICA* может быть сконфигурирована со своими настройками вывода, параметрами графиков по умолчанию и т.д. Или же можно сделать несколько значков *STATISTICA* в разных группах приложений на рабочем столе Windows (каждая из которых соответствует определенному проекту или виду работ) и задать для них различные значения в поле *Рабочая директория (Working Directory)* (с помощью диалогового окна системы Windows *Свойства программного элемента (Program Item Properties)*).

Пользовательские конфигурации для индивидуальных пользователей в сети

Те же принципы (см. предыдущий раздел) применимы к сетевому варианту системы *STATISTICA*. В сети программа инсталлируется один раз (на диске сервера), но каждый пользователь может конфигурировать ее по-своему, потому что параметры конфигурации определяются содержимым (локального) диска, с которого программа запускается. Имейте в виду, что для правильной инсталляции программы на удаленном диске (на сетевом сервере) необходимо выбрать режим *Сетевая инсталляция (Network Installation)* в программе *Установка пакета STATISTICA (STATISTICA Setup)*; после этого нужно запустить программу *reinst.exe* с каждой рабочей станции (чтобы сконфигурировать рабочие станции).

Сетевая версия системы *STATISTICA* необходима для надежной работы системы в случае, когда предполагается одновременная работа нескольких пользователей.

Copyright © StatSoft, 1995
StatSoft®

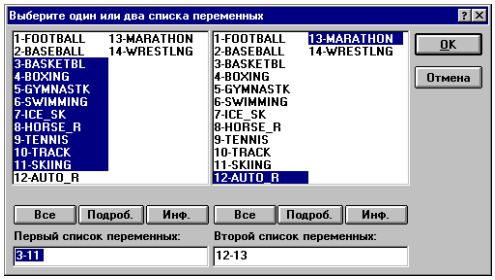
StatSoft®

Copyright © StatSoft, 1995

ВЫБОР И ПРОСМОТР ПЕРЕМЕННЫХ И ЗНАЧЕНИЙ

Выбор переменных для анализа

Когда в системе *STATISTICA* требуется выбрать одну или несколько переменных для анализа, будет открыто окно выбора переменных.



В зависимости от того, какие методы анализа будут применяться, будет открыто одно из трех видов окон (см. ниже): (1) окно выбора одного списка переменных, (2) окно выбора нескольких списков переменных и (3) окно выбора одной переменной.

Выбор переменных

Переменные можно выбирать или выделяя их в общем списке, или вводя их имена в поле ввода в этот же окне. Как правило (когда это возможно) при выделении блока данных в какой-нибудь таблице и первом открытии окна *Выберите переменную из списка* или *Выберите переменные из двух (нескольких) списков*, в этом окне будут предложены (т.е. “заранее выбраны”) переменные из выделенного блока. Границы блока выделенных переменных высвечиваются в поле ввода в нижней части окна. Для того чтобы задать другие номера переменных, просто впечатайте их поверх высвеченных (предварительно удалять их необязательно).

Выбор с помощью клавиш SHIFT и CTRL. При выборе переменных путем выделения их в списке можно выбрать несколько переменных, идущих подряд, держа нажатой левую кнопку мыши и проводя ее указателем по тем переменных, которые нужно выделить. Есть и другой способ выбрать несколько переменных, идущих подряд: нужно щелкнуть на имени первой из них, а затем, при нажатой клавише SHIFT, щелкнуть на последней. В результате окажутся выбранными не только первая и последняя переменные, но и все, расположенные в списке между ними. Можно выбирать и несколько переменных, не обязательно идущих в списке подряд: для этого нужно щелкнуть на каждой из них, держа нажатой клавишу CTRL.

Быстрый способ выбрать одну переменную и выйти

Если требуется выбрать только одну переменную из одного списка переменных, достаточно дважды щелкнуть на ее имени - переменная будет выбрана и диалоговое окно закроется. Двойной щелчок на имени переменной в окне выбора из нескольких списков не влияет на ранее выбранные переменные; поэтому, если в одном из списков надо выбрать ровно одну переменную, нужно сначала выбрать все переменные во всех остальных списках, а затем дважды щелкнуть на имени оставшейся переменной в последнем списке - в результате будут выбраны все отмеченные переменные, и окно закроется.

Отмена выбора ранее выбранных переменных

Отменить выбор ранее выбранных переменных (идущих подряд или нет) можно, держа нажатой клавишу CTRL и щелкая мышью на именах

Copyright © StatSoft, 1995

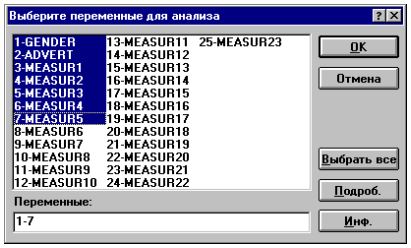
StatSoft®

переменных, выбор которых нужно отменить. Другой способ сделать это - удалить их номера в поле ввода.

Замечание. Информацию о том, как проводится анализ или строится график, и каким образом при этом используются выбранные переменные, можно найти в соответствующей главе, посвященной тому модулю, из которого вызывается окно выбора переменных. Подробное описание каждого типа окна выбора переменных дается ниже.

Выбор переменных из одного списка

"Это окно открывается, когда пользователю требуется выбрать переменные из одного списка переменных.



Если список переменных слишком велик и не умещается в окне, то он будет иметь полосу прокрутки.

Выбор переменных

Для выбора переменных в этом окне нужно или выделить их имена, или ввести их номера в поле ввода, а затем нажать кнопку *OK* (более подробно о выборе и отмене выбора переменных см. выше).

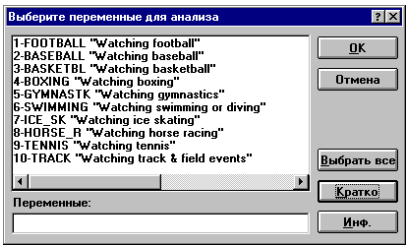
Выбрать все

При нажатии на эту кнопку будут автоматически выбраны все переменные в данном списке.

Другой способ выбрать весь список переменных - ввести символ * (звездочка) в поле ввода.

Подробно/Кратко

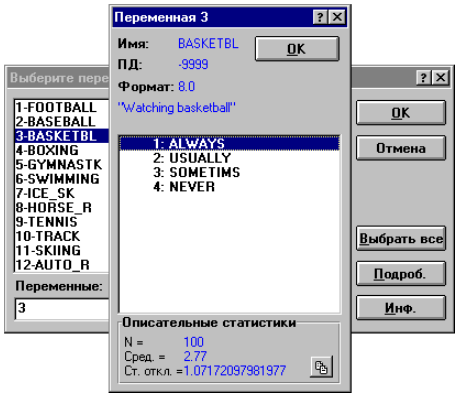
При нажатии на эту кнопку будут отображаться длинные имена всех переменных.



Повторное нажатие на эту кнопку удалит из списка длинные имена.

Информация

При нажатии на эту кнопку открывается окно значений переменной *Переменная...* (для первой из выбранных переменных, см. далее), в котором можно просмотреть упорядоченный список ее значений (числовых и текстовых эквивалентов).

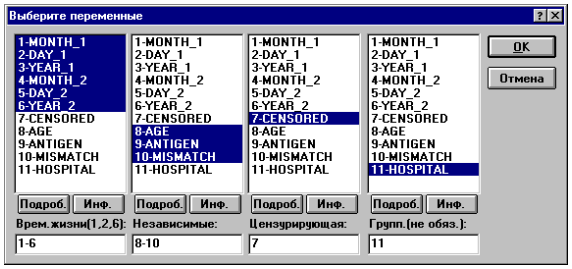


В нижней части окна показаны описательные статистики для данной переменной (их можно скопировать в буфер обмена кнопкой К*опировать*).

StatSoft®
Copyright © StatSoft, 1995

Выбор переменных из двух (нескольких) списков

Для некоторых методов анализа данных требуется выбирать переменные не из одного, а из нескольких списков (например, один список - для независимых переменных, другой - для зависимых).



В таком случае будет открыто окно, содержащее два (или более) списка, из которых можно выбирать переменные.

Выбор переменных

Для выбора переменных в каждом списке нужно или выделить их имена, или ввести их номера в поле ввода соответствующего списка, а затем нажать кнопку *OK* (более подробно о выборе и отмене выбора переменных см. выше).

Выбрать все

При нажатии на эту кнопку, относящуюся к нужному списку, будут автоматически выбраны все переменные из этого списка. Другой способ выбрать все переменные из какого-то списка - ввести символ * (звездочка) в поле ввода, относящееся к этому списку.

Подробно/Кратко

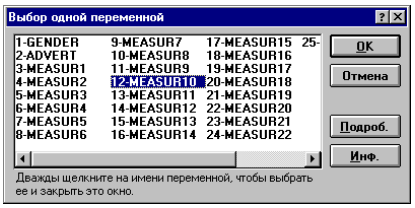
При нажатии на эту кнопку будут отображаться длинные имена всех переменных (см. об этом выше в разделе, где описан выбор из одного списка). Повторное нажатие на эту кнопку удалит из списка длинные имена.

Информация

При нажатии на эту кнопку открывается окно значений переменной *Переменная...* (для первой из выбранных переменных, см. далее), в котором можно просмотреть упорядоченный список ее значений (числовых и текстовых эквивалентов) - см. об этом выше в описании *одного списка*.

Выбор одной переменной из списка

Это окно открывается, когда пользователю требуется выбрать одну переменную из списка переменных.



Если список переменных слишком велик и не умещается в окне, то он будет иметь полосу прокрутки.

Выбор переменной

Для выбора переменной в этом окне нужно или дважды щелкнуть на ее имени, или выделить ее (щелкнув один раз на ее имени), а затем нажать кнопку *OK*.

Подробно/Кратко

При нажатии на эту кнопку будут отображаться длинные имена (если таковые имеются) всех переменных списка (см. об этом выше в разделе об *одном списке*). Повторное нажатие на эту кнопку удалит из списка длинные имена.

Copyright © StatSoft, 1995

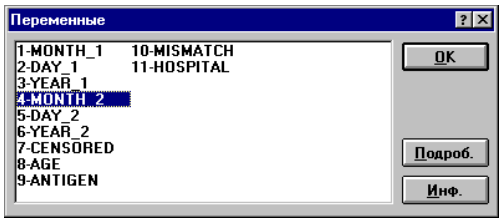
StatSoft®

Информация

При нажатии на эту кнопку открывается окно значений переменной *Переменная…*, в котором можно просмотреть упорядоченный список ее значений (числовых и текстовых эквивалентов) - см. об этом выше в разделе об *одном списке*.

Просмотр переменных из списка

Это окно открывается, когда пользователю требуется просмотреть имена переменных и другие их атрибуты.



Если список переменных слишком велик и не умещается в окне, то он будет иметь полосу прокрутки.

Подробно/Кратко

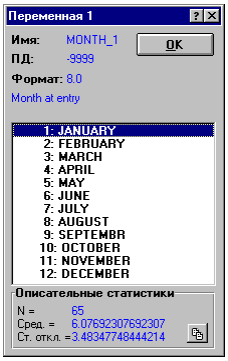
При нажатии на эту кнопку будут отображаться длинные имена (если таковые имеются) всех переменных списка (см. об этом выше в разделе об *одном списке*). Повторное нажатие на эту кнопку удалит из списка длинные имена.

Информация

При нажатии на эту кнопку открывается окно значений переменной *Переменная…*, в котором можно просмотреть упорядоченный список ее значений (числовых и текстовых эквивалентов) - см. ниже.

Окно значений переменной

В этом окне (доступ к нему осуществляется кнопкой *Инф.* в окнах выбора переменных, см. выше, или через пункт контекстного меню *Быстрые статистические графики – Значения для…*, открывающегося правой кнопкой мыши) можно просмотреть значения переменной, ее имя, пропущенные значения, ее формат и длинное имя (если оно есть), а также некоторые описательные статистики для этой переменной (см. рис.).



Значения переменной

Значения переменной отображаются в виде упорядоченного списка с прокруткой. Если переменная имеет текстовые значения, то они отображаются вслед за их числовыми эквивалентами.

Описательные статистики

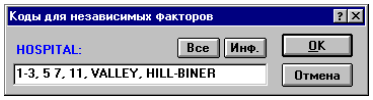
Описательные статистики для переменной включают объем выборки (*N*), среднее и стандартное отклонение. Эти статистики вместе с именем переменной можно скопировать в буфер обмена (с помощью расположенной здесь же кнопки *Копировать*) для того, чтобы вставить их затем в графическое или текстовое окно или использовать в каком-то другом приложении (например, Microsoft Word для Windows).

StatSoft®

Copyright © StatSoft, 1995

Выбор кодов

Диалоговое окно *Коды* открывается, когда нужно выбрать коды для группирующей переменной (специальные значения, которые используются для определения принадлежности наблюдений (случаев) к той или иной группе). В этом окне отображается имя группирующей переменной и поле ввода с подсказкой (например, 1, *Муж (Male)*, "#52"), куда следует ввести коды.



Можно ввести в это поле либо конкретные коды, либо символ * (звездочка) - и тогда будут приняты все коды, доступные для данной переменной. Текстовые коды в этом окне можно для экономии времени печатать в нижнем регистре, так как система *STATISTICA* автоматически переведет их в верхний регистр (например, *smith* будет преобразовано в *SMITH*). Система *STATISTICA* различает верхний и нижний регистр в кодах, поэтому необходимо следить за тем, чтобы те коды, в которых требуется сохранить буквы нижнего регистра, были написаны в одинарных или двойных кавычках (например, 'Smith', "Pepsi", "StatSoft", см. далее).

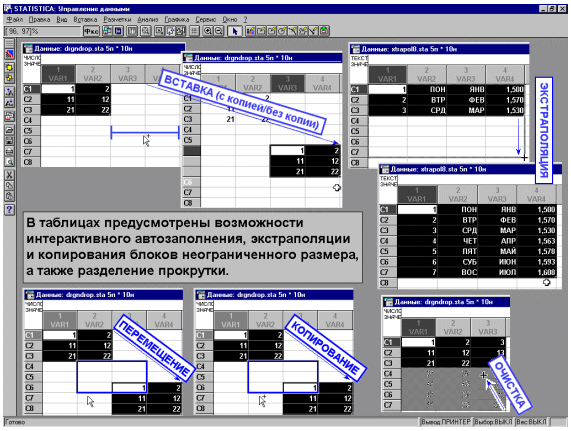
При вводе кодов действуют следующие основные соглашения.

- Если код состоит только из заглавных букв и цифр (например, *ITEM9*, *MALE*) и начинается не с цифры, то в поле ввода он будет отображаться без кавычек.
- Коды, введенные с использованием букв верхнего *и* нижнего регистров или *только* нижнего регистра (например, '*Male*', "*test1*") будут отображаться в этом поле ввода в кавычках (одинарных или двойных), чтобы предотвратить переформатирование из одного регистра в другой.
- Коды, начинающиеся с цифры или символа, не являющегося буквой (например, "=*shift3*", '*39lbs*', "*49 °C*", "*15-Apr*") отображаются в кавычках (одинарных или двойных).
- Даты отображаются в кавычках (одинарных или двойных) в том случае, если для формата переменной установлен тип *Дата* (см. раздел *Изменить (текущие) спецификации переменных*, глава 3).

Быстрый способ. Если в этом окне оставить все поля пустыми (чистыми) и нажать *OK*, система *STATISTICA* автоматически определит и будет использовать все имеющиеся коды (для заданных группирующих переменных). Тот же результат будет получен, если не задавать явным образом никаких кодов и, не входя в это окно, нажать кнопку *OK* в предыдущем окне (спецификаций).

Метод *Перетащить и отпустить* в системе *STATISTICA*

Как видно из следующего рисунка, система *STATISTICA* поддерживает полный набор стандартных операций метода *Перетащить и отпустить* с таблицами исходных данных (типа Excel).



Copyright © StatSoft, 1995

StatSoft®

В таблицах результатов также доступны все операции метода *Перетащить и отпустить*, за исключением вставки.

Подробные сведения о методе *Перетащить и отпустить* в системе *STATISTICA* см. в главе 3.

StatSoft®

Copyright © StatSoft, 1995

ИНТЕРФЕЙС КЛАВИАТУРЫ

Для доступа к различным процедурам пакета *STATISTICA* можно использовать «горячие» клавиши. «Горячие» клавиши могут быть глобальными (доступными из любого окна) или локальными (для отдельных окон или *кнопок автозадач*). «Горячую» клавишу можно присвоить макрокоманде с помощью панели инструментов *Кнопки автозадач* (см. главу 6). Далее приводится полный список глобальных и локальных «горячих» клавиш.

Глобальные «горячие» клавиши

Следующие ‘горячие’ клавиши доступны в любом окне документа.

Ввод/вывод файлов:
- Открыть файл того же типа — CTRL+O
- Сохранить — CTRL+S
- Сохранить как — F12

Открыть другой:
- Новые данные — CTRL+N
- Файл данных — CTRL+F12
- Графический файл — SHIFT+F3
- Таблицу результатов — CTRL+F11

Панели инструментов:
- Кнопки Автозадач — CTRL+M

Печать:
- Печать — CTRL+P, F4
- Параметры страницы/вывода — SHIFT+F4

Правка:
- Отмена — CTRL+Z или ALT+BACKSPACE
- Выделить все (см. *Примечание*) — CTRL+A
- Отменить выделение — DEL

Буфер обмена:
- Вырезать — CTRL+X
- Копировать — CTRL+C
- Вставить — CTRL+V
- Захват экрана — ALT+F3

Анализ:
- Стартовая панель — CTRL+T
- Продолжить анализ — CTRL+R
- Условия выбора наблюдений — F8
- Весовая переменная — F7

Макрокоманды:
- Запись — CTRL+F3
- Выполнение — CTRL+KEY

Окна:
- Упорядочить каскадом — HIFT+F6
- Упорядочить слева направо — ALT+F6
- Упорядочить сверху вниз — ALT+SHIFT+F6
- Справка — F1
- Закрыть — CTRL+F4
- Закрыть все незаблокированные окна, кроме данных — CTRL+L
- Выход — ALT+F4
- Переключиться в — CTRL+ESC

Примечание. Все содержимое графического окна постоянно является “выделенным” в том смысле, что при нажатии CTRL+C в буфер обмена будет скопирован весь график. Клавишами CTRL+A (глобальная «горячая» клавиша “выделить все”), последовательно выбираются пользовательские объекты; нажатие SHIFT+CTRL+A перемещает назад по списку объектов.

Copyright © StatSoft, 1995

StatSoft®

Локальные 'горячие' клавиши

Окно таблицы исходных данных

В дополнение к общим (глобальным) клавишам, доступным из каждого окна, в окне таблицы исходных данных работают следующие сочетания клавиш.

Преобразовать:

Преобразовать в таблицу результатов	F11

Пересчитать:

Пересчитать все формулы	F9

Правка:

Отмена	CTRL+Z или ALT+BACKSPACE
Изменить ячейку	F2
Изменить текущие спецификации	CTRL+F2
Расширить блок	SHIFT-курсор [или: поместить указатель в угол блока, который нужно выделить, затем переместить его в противоположный (по диагонали) угол и нажать кнопку мыши при нажатой клавише SHIFT].

Настройка:

Просмотреть шрифты	CTRL+F9
Настроить цвета	SHIFT+F9

Графика:

Быстрые статистические графики	F3

Окно таблицы результатов

Следующие комбинации клавиш (локальные 'горячие' клавиши) являются специфическими для окна таблицы результатов.

Правка:

Изменить ячейку	F2
Спецификации столбцов	CTRL+F2
Имя строки	ALT+F2
Расширить блок	SHIFT-курсор [или: поместить указатель в угол блока, который нужно выделить, затем переместить его в противоположный (по диагонали) угол и нажать кнопку мыши при нажатой клавише SHIFT].

Настройка:

Просмотреть шрифты	F9
Настроить цвета	SHIFT+F9

Графики:

Быстрые статистические графики	F3

Окно текста/вывода, окна программ на языках *STATISTICA BASIC* и *SCL*

Следующие комбинации клавиш (локальные 'горячие' клавиши) являются специфическими для *Окна текста/вывода* (см. главу 5), окна языка *STATISTICA BASIC* и окна *Командного языка (SCL)*.

Правка:

Отмена	CTRL+Z или ALT+BACKSPACE
Найти	CTRL+F
Найти далее	CTRL+ALT+F
Заменить	CTRL+H
Перейти	F5
Вставить рисунок из файла	F3
Размер объекта	SHIFT+ F 3
Цвета	SHIFT+F9
Вставить новую страницу	CTRL+ENTER

Символы:

Обычный	CTRL+0 (ноль)

StatSoft®

Copyright © StatSoft, 1995

Полужирный	CTRL+B
Курсив	CTRL+I
Подчеркивание	CTRL+U
Двойное подчеркивание	CTRL+D
Шрифт	F9

Мастер (для *SCL* и *STATISTICA BASIC*): ALT+I

Графическое окно

Следующие комбинации клавиш (локальные 'горячие' клавиши) являются специфическими для графического окна.

Правка:

Отмена	CTRL+Z
Выделить все	CTRL+A (см. ниже)

Свойства объекта:

Свойства объекта	ALT+ENTER

Выравнивание объекта:

Направляющая сетка	CTRL+G
Прикрепить к сетке	держать нажатой TAB

Выделить все. Все содержимое графического окна постоянно является "выделенным" в том смысле, что при нажатии CTRL+C в буфер обмена будет скопирован весь график. Клавишами CTRL+A (глобальная 'горячая' клавиша "выделить все"), последовательно выбираются пользовательские объекты; нажатие SHIFT+CTRL+A перемещает назад по списку объектов.

Вращение текста с помощью клавиш курсора. Выделите добавленный на график текст, и с помощью клавиш PAGE DOWN или PAGE UP вращайте выделенный текстовый объект в плоскости графика, соответственно по или против часовой стрелки с шагом 5°.

Для вращения с шагом 1° нажимайте клавиши PAGE DOWN и PAGE UP при нажатой клавише CTRL.

Перемещение и изменение размеров объектов с помощью клавиш курсора. Выделите объект, установите на него курсор, для изменения его размеров или перемещения при помощи клавиш курсора на клавиатуре перетащите тень графического объекта на новое место.

Для медленного перемещения (с шагом в один пиксель) держите нажатой клавишу CTRL. Нажатие клавиши ENTER заканчивает выполнение операции.

Окно редактора данных графика

Следующие комбинации клавиш (локальные 'горячие' клавиши) являются специфическими для окна *Редактор данных графика* (см. том II).

Правка:

Изменить ячейку	F2
Изменить текущие спецификации	CTRL+F2

Настройка:

Окно просмотра шрифтов	F9
Настроить цвета	SHIFT+F9

Окно *Менеджера мегафайлов*

Следующие комбинации клавиш (локальные 'горячие' клавиши) являются специфическими для окна *Менеджера мегафайлов* (см. том III руководства).

Правка:

Изменить ячейку	F2

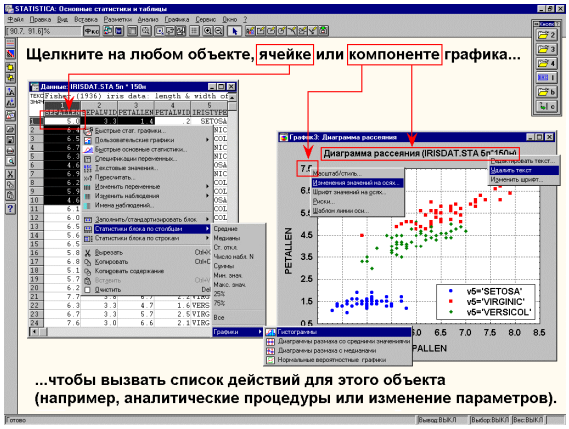
Настройка:

Просмотреть шрифты	F9
Настроить цвета	SHIFT+F9

Copyright © StatSoft, 1995

СОГЛАШЕНИЯ ПО ИСПОЛЬЗОВАНИЮ МЫШИ

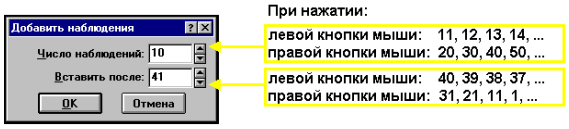
Помимо возможностей, стандартных для всех Windows-приложений, система *STATISTICA* обладает рядом специальных средств, быстрый доступ к которым осуществляется с помощью мыши.



Далее дается обзор функций и соглашений, относящихся к мыши.

Кнопки микропрокрутки

Во всех *полях ввода* числовых значений имеются *кнопки микропрокрутки* (например, 10). На следующем примере показано, как можно управлять этими *кнопками микропрокрутки* с помощью левой и правой кнопок мыши.



Наши исследования показали, что благодаря реализации этого инструмента в системе *STATISTICA* значительно возросла эффективность пользовательского интерфейса.

В отличие от простых *кнопок микропрокрутки*, имеющихся в некоторых других приложениях, все *кнопки микропрокрутки* системы *STATISTICA* автоматически настраиваются на текущее содержимое редактируемого поля и меняют значение последней значащей цифры.

С помощью правой кнопки мыши можно менять предпоследнюю цифру (например, при нажатии на *кнопку микропрокрутки* левой кнопкой мыши, значение *0.15* превратится в *0.16*, затем в *0.17*, *0.18*, и т.д.; если же нажимать на правую кнопку, то *0.15* изменится на *0.25*, затем на *.35*, *.45*, и т.д.).

Левая и правая кнопки мыши

Левая кнопка мыши

Левая кнопка мыши используется для выбора (например, режима в окне диалога), выделения (например, блока данных в таблице), и перетаскивания (например, блока таблицы или объекта на графике; см. далее). Двойной щелчок на объекте левой кнопкой мыши напрямую вызывает наиболее часто используемые диалоговые окна системы *STATISTICA*.

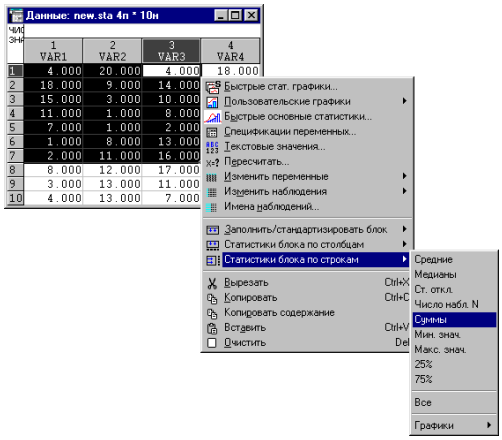
Например, по двойному щелчку левой кнопкой мыши на ячейке, заголовке, имени столбца и т.п. в таблице исходных данных или таблице результатов, открывается диалоговое окно, в котором можно редактировать соответствующие атрибуты; двойной щелчок левой кнопкой на графическом объекте, названии, маркере точки и т.д. раскроет диалоговое окно, в котором можно редактировать соответствующую часть графика. Таким образом, при обращении к часто используемым методам настройки экономится один шаг. Кроме того, двойной щелчок на фоновой поверхности окна системы *STATISTICA*

StatSoft
Copyright © StatSoft, 1995

вызывает диалоговое окно *Переключатель модулей*.

Правая кнопка мыши

При щелчке на каком-либо объекте (например, ячейке таблицы) правой кнопкой мыши, на экране возникает динамическое контекстное меню.



Через эти контекстные меню осуществляется быстрый доступ к некоторым наиболее часто используемым операциям всех окон системы *STATISTICA* (например, графических окон, окон таблиц исходных данных, *Окон текста/вывода*, и т.д.).

Так, например, при щелчке правой кнопкой на ячейке таблицы открывается контекстное меню с командами для графических и статистических процедур, операций с буфером обмена и пр. (см. рис. выше), с помощью которого можно быстро выполнить требуемую операцию. Щелчок правой кнопкой на фоновой поверхности окна приложения *STATISTICA* вызывает диалоговое окно *Переключатель модулей*.

Выбор элементов из списков с множественным выбором

Выбрать элементы из списка с возможностью множественного выбора можно одним из перечисленных ниже способов.

- Щелкнуть на элементе, чтобы выбрать (выделить) его. Нажать кнопку OK в окне, чтобы завершить выбор.
- Дважды щелкнуть на элементе, чтобы выбрать его и завершить выбор (закрыть окно).
- Для выбора нескольких элементов, идущих подряд: (1) держа нажатой левую кнопку, провести указателем по тем элементам, которые нужно выбрать, или (2) выделить первый элемент, затем при нажатой клавише SHIFT щелкнуть на последнем из выбираемых элементов.
- Для выбора нескольких элементов, идущих вразброс: держа нажатой клавишу CTRL , щелкнуть на каждом из них.

Использование клавиш SHIFT и CTRL в сочетании с мышью

В некоторых случаях работы с мышью, результат действий меняется, если при этом держать нажатой клавишу CTRL или SHIFT. Например:

- Чтобы быстро выделить блок ячеек в таблице исходных данных или таблице результатов, можно щелкнуть на ячейке, соответствующей началу блока, затем (возможно, предварительно воспользовавшись линейками прокрутки) при нажатой клавише SHIFT щелкнуть на ячейке, соответствующей концу блока (SHIFT-соглашение). В результате будут

Copyright © StatSoft, 1995

StatSoft®

выделены все данные в промежутке между этими двумя ячейками.

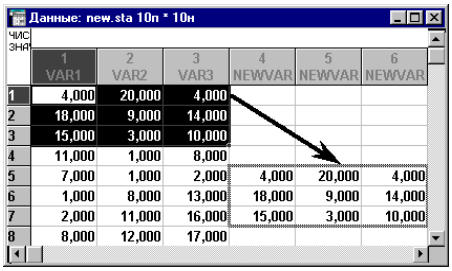
- Если при нажатой клавише CTRL щелкнуть на *кнопке автозадач* (см. стр. 1015 и главу 6), откроется диалоговое окно редактирования содержимого кнопки.
- Если при нажатой клавише CTRL нажать кнопку *Увеличить шрифт* или кнопку *Уменьшить шрифт* графической панели инструментов, то база отображения будет меняться более мелкими шагами (4% вместо 16% по умолчанию) (см. главу 2).
- Если при нажатой клавише CTRL нажать кнопку на панели инструментов таблицы исходных данных, таблица переключится в режим отображения меток значений (см. главу 2).
- Держите нажатой клавишу SHIFT для того, чтобы нарисовать мышью квадрат, прямоугольник со скругленными углами, круг или сектор в графическом окне (см. главу 2).

Более специальные возможности описаны далее.

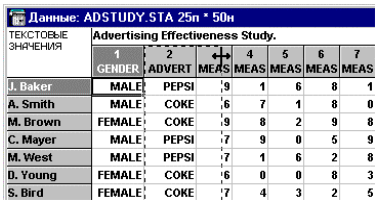
Другие применения

Помимо уже перечисленных возможностей, с помощью мыши можно проделывать описанные ниже действия.

- Метод *Перетащить и отпустить*. С помощью мыши можно быстро перемещать, копировать, удалять, вставлять, размножать и т.д. блок ячеек таблицы исходных данных или таблицы результатов (см. 3).

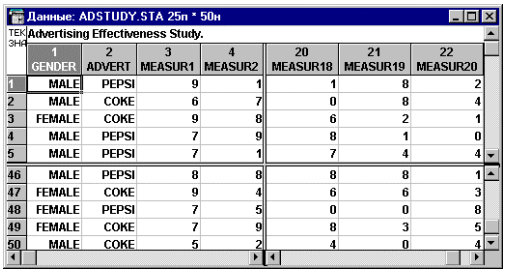


- Увеличение или уменьшение ширины столбца. Установите нужную ширину столбца в таблице исходных данных или таблице результатов, перемещая мышью правую границу столбца на новое место (см. главу 3).



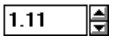
Новая ширина столбца показывается пунктиром (см. рис.).

- Разделение прокрутки (в таблице исходных данных или таблице результатов). Чтобы разделить окно таблицы исходных данных или таблицы результатов (т.е. разделить прокрутку, см. главу 3), перетащите мышью маркер разделения (черный прямоугольник в верхней части вертикальной линейки прокрутки или в левой части горизонтальной линейки прокрутки).



StatSoft®
Copyright © StatSoft, 1995

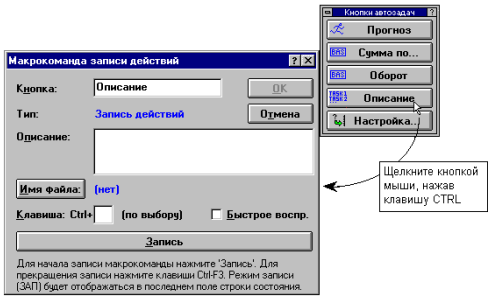
- Скорость прокрутки переменных. Управление скоростью прокрутки (на одну строку при небольшом движении курсора мыши и на одну страницу при сильном движении) при расширении блока за видимые границы таблицы исходных данных или таблицы результатов (см. главу 3).

- *Кнопки микропрокрутки* 

позволяют плавно изменять значение в числовом поле ввода - в последнем знаке (нажимать на кнопку левой кнопкой мыши, например, 1.11, 1.12, 1.13, ...) или в 10 раз быстрее (правой кнопкой, например, 1.11, 1.21, 1.31, ...; см. стр. 1034).

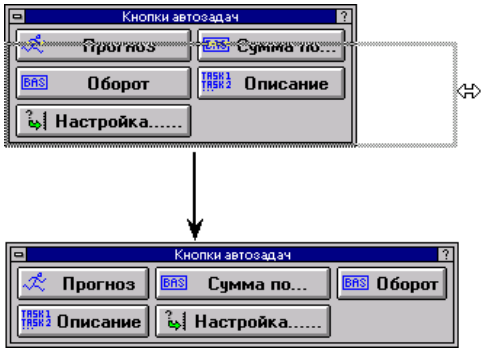
- Конфигурация панели инструментов. Двойной щелчок в свободном пространстве между кнопками панели инструментов переключает варианты конфигурации панели инструментов (всего их четыре: 1 строка, 2 строки, справа вверху, слева вверху; см. главу 2).

- *Кнопки автозадач.* Если при нажатой клавише CTRL нажать какую-нибудь *кнопку автозадач*, то



откроется диалоговое окно, в котором можно отредактировать содержимое этой *кнопки автозадач* (см. главу 6).

- Панель инструментов *Кнопки автозадач*. Двойной щелчок левой кнопкой мыши в свободном пространстве между кнопками панели инструментов *Кнопки автозадач* будет менять ее размер (другой способ изменить размеры этой панели инструментов - передвинуть ее границу мышью).



Щелчок правой кнопкой мыши открывает контекстное меню. Можно “прикрепить” панель инструментов, поместив ее в ту область окна приложения *STATISTICA* (над панелью инструментов, под строкой состояния или к любому краю окна), где она изменит свою конфигурацию, и будет оставаться на этом месте, пока ее не переместят в какую-то другую часть окна (см. главу 6).

- Переупорядочивание элементов списка. Элементы списка (например, в Рабочей книге, на панели инструментов *Кнопки автозадач*, в окне *Мастер автоматического размещения нескольких графиков* и т.д.) можно переупорядочить, выделив один или несколько из них (подряд или нет), и затем поставив указатель мыши в нужную позицию (при этом его вид изменится на ⇳). Щелчок левой кнопкой переместит выделенный элемент (или элементы) в указанное место списка.

Графические применения

Следующие способы использования мыши применимы в графическом окне системы

Copyright © StatSoft, 1995
StatSoft®

STATISTICA. При этом координаты мыши (или угол поворота, см. ниже) отображаются в окошке *Поле просмотра* графической панели инструментов.

Более подробную информацию по этим способам работы см. в томе II.

- Методы *OLE*. Связывание и внедрение внешних файлов документов в документы *STATISTICA* выполняется перетаскиванием их непосредственно из окна *Диспетчер файлов (File Manager)* (через окна приложений) в графическое окно *STATISTICA*.
- *Закрашивание*. Выделение точек на графике с помощью курсора в виде кисти.
- *Увеличение* и *уменьшение*. Соответственно увеличивает или уменьшает выделенную область в графическом окне (курсор в виде лупы).
- Средства рисования позволяют добавлять в графическое окно прямоугольники, овалы или круги, ломаные линии, произвольные рисунки от руки, стрелки и т.д. Для того, чтобы нарисовать квадрат, прямоугольник со скругленными углами, круг, сектор и т.п., держите нажатой клавишу SHIFT и рисуйте мышью.
- Изменение размеров и перемещение. Изменение размеров выделенных графических объектов с помощью передвижения маркера границы (⟷) или перемещение объекта в целом (✣). О точной настройке см. далее в параграфе *Управление мышью с помощью клавиатуры в графических окнах*.
- Редактирование ломаных линий. Изменить размеры отдельных кусков ломаной можно, двигая мышью масштабирующие квадратики всего объекта или его прямолинейных сегментов. О точной настройке см. далее параграф *Управление мышью с помощью клавиатуры в графических окнах*.
- Вращение текста. Пользовательский текст можно интерактивно вращать, выделив его в графическом окне и затем передвигая один из его "рычагов" (маленьких черных квадратов) в нужную сторону [при этом угол поворота (в градусах) отображается в окошке *Поле просмотра* графической панели инструментов]. О точной настройке (поворот на один градус) см. далее в параграфе *Управление мышью с помощью клавиатуры в графических окнах*.
- *Прикрепить к сетке* . Выравнивает рисованные объекты по направляющей сетке из невидимых линий, покрывающих поле рисунка.
- *База отображения графика*. Нажатие кнопки или на графической панели инструментов меняет значение параметра *База отображения графика* соответственно на ±16%. Если при этом держать нажатой клавишу CTRL, шаг изменения уменьшится (до 4%).
- *Управление мышью с помощью клавиатуры в графических окнах*: Действия мыши в операциях перемещения и изменения размеров объекта можно эмулировать с помощью клавиатуры: нужно выделить объект, поместив на него указатель мыши, а затем с помощью клавиш курсора переместить его или изменить его размеры. Если при этом держать нажатой клавишу CTRL, движение будет медленным (по одному пикселю). Для завершения действия нажмите клавишу ENTER.

Курсор мыши меняет свой вид и превращается в условное изображение того инструмента, которые используется в данной ситуации. Нажатием клавиши ESC мышь возвращается в обычный режим.

StatSoft®

Copyright © StatSoft, 1995

Кроме того, с помощью мыши (левой кнопкой) можно прервать текущую операцию (перерисовку графика, выполнение *SCL*-программы и т.д.; более подробную информацию см. далее).

Остановка или прерывание текущей операции

Иногда бывает необходимо прервать текущую операцию или процедуру анализа данных. Это можно сделать одним из перечисленных ниже способов.

Анализ. Для прерывания текущей процедуры анализа данных нажмите кнопку *Отмена* в поле *Индикатор состояния*.

Перерисовка графика. Перерисовку графика можно прервать, щелкнув мышью (в любом месте) или нажав любую клавишу.

Работа с кистью (на графике). Чтобы выключить режим закрашивания достаточно нажать кнопку *Режим выделения (указатель)* на панели инструментов или клавишу ESC.

Печать. Печать таблицы исходных данных, таблицы результатов, графика, текстового вывода и т.д. можно прервать кнопкой *Отмена* в окне *Печать*.

Программа на языке *STATISTICA BASIC*. Для того, чтобы прервать исполнение программы на языке *STATISTICA BASIC* , нажмите клавишу ESC или кнопку *Отмена* в поле *Индикатор состояния*.

Программа на языке *SCL*. Исполнение программы на языке *SCL* можно в любой момент прервать щелчком мыши или клавишами ESC или CTRL+BREAK.

Программа на языке *MM*. Для того, чтобы прервать исполнение программы на языке *MM*, нажмите клавишу ESC или кнопку *Отмена* в поле *Индикатор состояния*.

Макрокоманда. Чтобы остановить запись макрокоманды, нажмите CTRL+F3. Чтобы остановить выполнение макрокоманды, щелкните кнопкой мыши или нажмите клавишу ESC или CTRL+BREAK.

Copyright © StatSoft, 1995

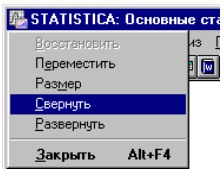
StatSoft®

Copyright © StatSoft, 1995

ДРУГИЕ ДИАЛОГОВЫЕ ОКНА И ПАРАМЕТРЫ

Системное меню окна приложения

Описанные далее команды *системного меню* позволяют проводить различные операции с окном приложения. Для того чтобы открыть его, нажмите кнопку в левом верхнем углу окна *STATISTICA* или на минимизированном значке *STATISTICA*.



Восстановить

Эта команда восстанавливает окно в его предыдущих размерах после того, как оно было свернуто или развернуто.

Переместить

Данная команда позволяет переместить окно в другое место с помощью клавиатуры.

Размер

Эта команда позволяет изменить размер окна с помощью клавиатуры.

Свернуть

Данная команда сворачивает окно в значок.

Развернуть

Эта команда расширяет окно до максимального размера.

Закрыть

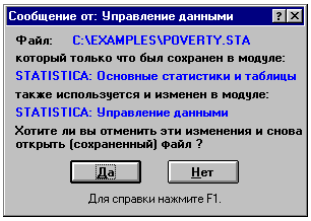
Данная команда закрывает работающий модуль системы *STATISTICA*.

Переключиться в

Эта команда открывает окно *Список задач (Task List)*, в котором можно переключиться с одного работающего приложения на другое, а также упорядочить их окна и значки.

Повторное открытие файла, сохраненного в другом модуле

Это диалоговое окно открывается в том случае, если при одновременной работе с двумя (или более) модулями системы *STATISTICA* попытаться в одном из модулей сохранить файл, измененный одновременно в нескольких модулях.



Пример:

1. При работе с файлом *A.sta* в модуле *Основные статистики и таблицы* в этот файл внесены изменения.
2. Одновременно в файл *A.sta* вносятся изменения при работе в модуле *Управление данными*.
3. Задается команда *Сохранить* из модуля *Управление данными*.

В этом случае, сохранив файл в модуле *Управление данными*, программа спросит, нужно ли открыть уже новый (сохраненный) вариант

Copyright © StatSoft, 1995

StatSoft®

файла *A.sta* в модуле *Основные статистики и таблицы*.

При ответе *Да* файл будет открыт повторно и в нем будут отражены изменения, зафиксированные операцией *Сохранить*, которая была выполнена в модуле *Управление данными*. При ответе *Нет* в окне останется вариант файла из модуля *Основные статистики и таблицы*.

Похожая ситуация возникает, когда в одном модуле выполняется операция *Сохранить как* и в качестве имени файла указывается имя файла, открытого в данный момент в другом модуле и пока еще не сохраненного.

Пример:

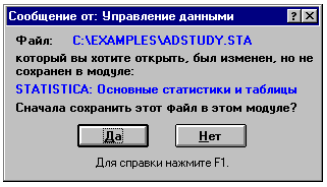
1. Проводится работа с файлом *A.sta* в модуле *Основные статистики и таблицы* и в этот файл внесены изменения.
2. Проводится работа с файлом *B.sta* в модуле *Управление данными*.
3. Задана команда *Сохранить* из модуля *Управление данными* и в качестве нового имени для сохраняемого файла указано *A.sta*.

В этом случае программа *STATISTICA* прежде всего предупредит о том, что файл *A.sta* уже существует и будет заменен. После этого, сохранив файл в модуле *Управление данными*, программа спросит, нужно ли открыть уже новый (сохраненный) вариант файла *A.sta* в модуле *Основные статистики и таблицы*.

При ответе *Да* файл будет открыт повторно, и в нем будут отражены изменения, зафиксированные операцией *Сохранить как*, которая была выполнена в модуле *Управление данными*. При ответе *Нет* в окне останется вариант файла из модуля *Основные статистики и таблицы*.

Сохранение измененного файла перед открытием его в другом модуле

Это диалоговое окно открывается при одновременной работе с двумя (или более) модулями *STATISTICA*, если попытаться в одном из них открыть файл, который в данный момент используется и был изменен (но не был сохранен) в другом модуле.



Пример:

1. Проводится работа с файлом *A.sta* в модуле *Основные статистики и таблицы* и в этот файл внесены изменения.
2. Выбрана команда *Открыть* из модуля *Управление данными* и в качестве имени открываемого файла указано *A.sta*.

В этом случае программа прежде всего спросит, нужно ли перед открытием этого файла в модуле *Управление данными* сохранить в нем изменения, которые были сделаны при работе в модуле *Основные статистики и таблицы*.

При ответе *Да* модуль *Основные статистики и таблицы* автоматически сохранит этот файл, и уже этот новый вариант будет открыт в модуле *Управление данными*. При ответе *Нет* в модуле *Управление данными* будет открыт исходный вариант файла *A.sta*.

Похожая ситуация возникает и при работе с файлом в одном из модулей. Предположим, в этот файл внесены изменения, но он не сохранен, в то время как предпринята попытка запустить другой модуль системы *STATISTICA*.

StatSoft®

Copyright © StatSoft, 1995

Пример:

1. При работе с файлом *A.sta* в модуле *Основные статистики и таблицы* в этот файл вносятся изменения.
2. Вызывается окно *Переключатель модулей системы STATISTICA* и в нем указывается модуль *Управление данными*.

При запуске модуль *Управление данными* попытается открыть файл *A.sta* в качестве файла данных. Обнаружив, что файл *A.sta* был изменен, но не сохранен в модуле *Основные статистики и таблицы*, программа предпримет те же действия, что и в предыдущем примере.

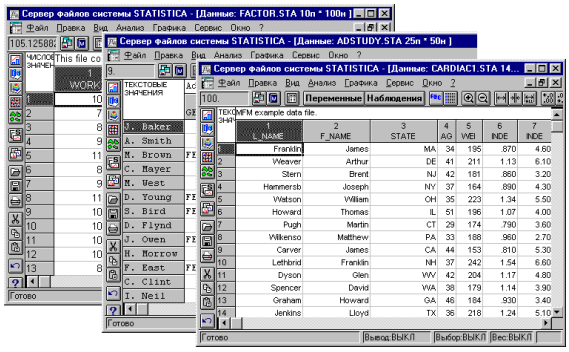
Copyright © StatSoft, 1995

StatSoft®

Copyright © StatSoft, 1995

СЕРВЕР ФАЙЛОВ СИСТЕМЫ STATISTICA

Сервер файлов STATISTICA - это небольшой, быстро загружаемый модуль пакета *STATISTICA*,



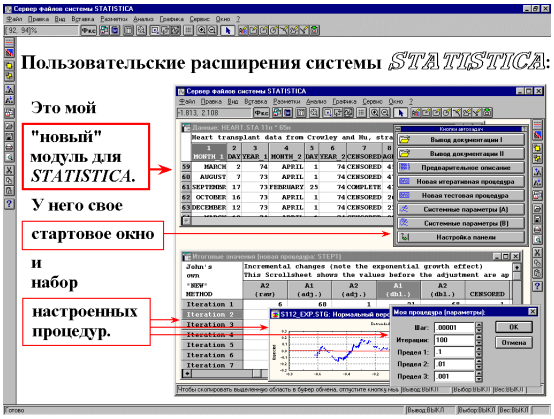
обеспечивающий возможности доступа и модификации всех файлов системы *STATISTICA*, включая файлы данных (расширение имени файла **.sta*), файлы таблиц результатов (расширение **.scr*), графические файлы (расширение **.stg*), файлы языка *SCL* (расширение **.scl*) и файлы языка *STATISTICA BASIC* (расширение **.stb*). Чтобы открыть файл системы *STATISTICA* в окне *Сервер файлов* системы *STATISTICA*, достаточно дважды щелкнуть на его имени в окне *Диспетчер файлов (File Manager)* Windows.

Кроме того, *Сервер файлов* обеспечивает поддержку возможностей *OLE* для файлов системы *STATISTICA* в других приложениях (если в этот момент не работает никакой другой модуль *STATISTICA*).

Можно одновременно открыть столько экземпляров окна *Сервера файлов*, сколько позволяет операционная система, и одновременно просматривать и редактировать большое количество файлов.

Пользовательская разработка новых модулей в системе *STATISTICA*

Сервер файлов может служить основой для разработки пользовательских "специализированных модулей *STATISTICA* " или "приложений *STATISTICA*", написанных на языке *STATISTICA BASIC* или на сочетании языков *SCL* (*Командный язык STATISTICA*) и *STATISTICA BASIC*.



Панель инструментов *Кнопки автозадач* (см. главу 6) можно использовать как удобную стартовую панель для такого нового модуля. Каждой кнопке на этой панели инструментов может быть назначена некоторая определенная пользователем вычислительная или графическая процедура или процедура управления данными.

Copyright © StatSoft, 1995

StatSoft®

StatSoft®

Copyright © StatSoft, 1995

ПРЕДМЕТНЫЙ УКАЗАТЕЛЬ

А

Автоматизация задач, 1009
Автоматическое нажатие кнопки *Далее*, 1011

В

Вид (меню), 1021
Внутренний режим пакетной обработки, 1010
Вращение текста с помощью клавиш курсора, 1033
Выбор
  одной переменной из списка, 1027
  одной переменной с выходом, 1025
  переменных для анализа, 1025
  переменных из нескольких списков, 1027
  переменных из одного списка, 1026
Выбор кодов, 1029
Выбор элементов из списка с помощью мыши, 1035

Г

Глобальные *горячие* клавиши, 1031
Графики
  выделить все, 1031, 1033
Графическое окно
  горячие клавиши, 1033

Д

Диспетчер программ системы Windows, 1014
Другие виды анализа в меню *Анализ*, 1007

З

Значки модулей STATISTICA, 1007

И

Изменение размеров объектов с помощью клавиш курсора, 1033
Индивидуальные пользователи в сети, 1023
Индивидуальный режим работы, 1005
Интерактивный пользовательский интерфейс, 1006
Интерфейс клавиатуры, 1031
  глобальные *горячие* клавиши, 1031
  графическое окно, 1033
Исполняемый модуль системы STATISTICA, 1014

К

Клавиатурный интерфейс
  Менеджер мегафайлов, 1033
  окно таблицы исходных данных, 1032
  окно таблицы результатов, 1032
Клавиатурный интерфейс
  локальные *горячие* клавиши, 1032, 1033
  окно редактора данных графика, 1033
  окно текста/вывода, 1032
Клавиша CTRL вместе с мышью, 1035
Клавиша SHIFT вместе с мышью, 1035
Кнопка переключателя модулей *Закрыть и переключиться в*, 1021
Кнопки автозадач, 1005, 1008, 1009, 1015, 1023
Коды
  выбор, 1029
  текстовые значения, 1029
Командный язык (SCL), 1005, 1013

Л

Левая кнопка мыши, 1034

М

Макрокоманды, 1012
Менеджер мегафайлов
  горячие клавиши, 1033
Метод *Перетащить и отпустить*
  обзор, 1029
Методы OLE, 1011
Микропрокрутка, 1034
Многодокументный пользовательский интерфейс (MDI), 1011
Модуль запуска, 1014
Модульная структура системы STATISTICA, 1006

Н

Настройка пользовательского интерфейса, 1019
Несколько окон приложений для модулей, 1021

О

Общие параметры системы, 1022
Окна документов в системе STATISTICA (MDI), 1011
Окна постановки задачи и задания формата вывода, 1008
Окно значений переменной, 1028
Окно приложения, 1020
Окно редактора данных графика
  горячие клавиши, 1033
Окно *Сервер файлов* системы *STATISTICA*, 1045
Окно таблицы исходных данных
  горячие клавиши, 1032
Окно таблицы результатов
  горячие клавиши, 1032
Окно текста/вывода
  горячие клавиши, 1032
Операции с блоками (*Перетащить и отпустить*), 1029
Основные компоненты интерактивного пользовательского интерфейса системы STATISTICA, 1006
Очереди, 1009

П

Пакетная обработка, 1010
Панели инструментов, зависящие от типа активного окна документа, 1011
Параметры по умолчанию
  общие настройки, 1020
Параметры по умолчанию для графиков, 1022
Переключатель модулей, 1007

Copyright © StatSoft, 1995

StatSoft®

Переключение модулей:режим
одного приложения, установка, 1020
Переменные, выбор
выбрать одну переменную и выйти, 1025
из нескольких списков, 1027
из одного списка, 1026
одна переменная из списка переменных, 1027
отмена выбора ранее выбранных переменных, 1025
просмотреть переменные из списка, 1028
Перемещение объектов с помощью клавиш курсора, 1033
Повторно загрузить файл, сохраненный в другом модуле, 1041
Пользовательская настройка значков (в окне диспетчера программ), 1008
Пользовательские расширения языка SCL, 1014
Правая кнопка мыши, 1035
Прерывание и остановка текущей операции, 1039
Приложение, системное меню окна, 1041
Программа установки, 1023
Программы на языке SCL, 1013
пользовательский интерактивный интерфейс, 1014
Программы на языке SCL »под ключ», 1014
Просмотр
значения переменной, 1028
Просмотреть переменные из списка, 1028

Р

Режим одного приложения, 1021
Режим приложения, 1020

С

Свойства программного элемента, диалоговое окно, 1008
Сервер файлов, 1045
Сервис (меню), 1021
Сетевая версия системы STATISTICA, 1023
Сетевая инсталляция, режим, 1023
Системное меню окна приложения, 1041
Совместимость документов, 1011
Соглашение о SHIFT-щелчке, 1035
Соглашения по использованию мыши, 1034
выбор элементов из списка, 1035
графические применения, 1037
использование клавиш SHIFT и CTRL, 1035
кнопки автозадач, 1037
конфигурация панели инструментов, 1037
левая кнопка мыши, 1034
метод *Перетащить и отпустить*, 1036
микропрокрутка, 1037
панель инструментов *Кнопки автозадач*, 1037
переупорядочивание элементов списка, 1037
правая кнопка мыши, 1035
прерывание и остановка текущей операции, 1039
разделение прокрутки, 1036
скорость прокрутки переменных, 1037
ширина столбца, 1036
Сохранить
измененный файл перед его загрузкой в другой модуль, 1042
Стандартный интерфейс DDE/API, 1015
Стартовая панель, 1008

У

Управление пакетом STATISTICA, 1015

Я

Язык SCL, 1013

StatSoft®

Copyright © StatSoft, 1995

Глава 2:
ПАНЕЛИ ИНСТРУМЕНТОВ

Оглавление

Вводный обзор 1053
Панель инструментов таблицы исходных данных 1054
Панель инструментов таблицы результатов (прокручиваемого окна вывода) 1071
Панель инструментов *Окна текста/вывода* 1081
Графическая панель инструментов 1086
Строка состояния 1110
Предметный указатель 1119

Подробное оглавление

ВВОДНЫЙ ОБЗОР 1053

ПАНЕЛЬ ИНСТРУМЕНТОВ ТАБЛИЦЫ ИСХОДНЫХ ДАННЫХ 1054

0. Поле просмотра 1054

1. Переключатель модулей 1054

2. Управление Рабочей книгой 1055

3. Спецификации всех переменных (таблица) 1056

4. Глобальные операции с переменными 1056

5. Глобальные операции с наблюдениями 1057

6. Отображение числовых/текстовых значений (для отображения меток значений нужно нажать клавишу CTRL) 1058

7. Отображение имен наблюдений (переключение) 1058

8. Увеличение 1058

9. Уменьшение 1059

10. Увеличить ширину столбца 1059

11. Уменьшить ширину столбца 1059

12. Изменение глобальной ширины столбцов 1059

13. Добавить десятичный разряд 1059

14. Удалить десятичный разряд 1059

15. Пересчитать 1059

16. Диспетчер текстовых значений 1060

17. Быстрые основные статистики 1060

18-23. Кнопки вызова графиков таблицы исходных данных 1062

Copyright © StatSoft, 1995

StatSoft®

18. 2M пользовательские графики (таблица исходных данных) 1062
19. 3M пользовательские последовательные графики (таблица исходных данных) 1063
20. Пользовательские XYZ графики (таблица исходных данных) 1063
21. Пользовательские матричные графики (таблица исходных данных) 1064
22. Пользовательские пиктографики (таблица исходных данных) 1065
23. Быстрые статистические графики (таблица исходных данных) 1065
24. Галерея графиков системы *STATISTICA* 1066
25. Открыть файл данных 1067
26. Сохранить файл данных 1067
27. Печать файла данных (или выделенного блока) 1067
28. Вырезать (блок) 1067
29. Копировать (блок) 1068
30. Вставить (Блок) 1069
31. Отмена 1069
32. Справка 1069

ПАНЕЛЬ ИНСТРУМЕНТОВ ТАБЛИЦЫ РЕЗУЛЬТАТОВ (ПРОКРУЧИВАЕМОГО ОКНА ВЫВОДА) 1071

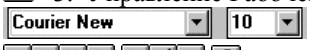
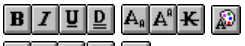
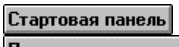
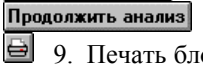
0. Поле просмотра 1071
1. Переключатель модулей 1071
2. Управление Рабочей книгой 1071
3. Увеличение 1072
4. Уменьшение 1072
5. Увеличить ширину столбца 1072
6. Уменьшить ширину столбца 1072
7. Изменить ширину столбцов и полей 1072
8. Добавить десятичный разряд 1073
9. Удалить десятичный разряд 1073
10. Маркировать ячейки 1073
11. Быстрые основные статистики 1074
12-17. Кнопки вызова графиков таблицы результатов 1074
12. 2M пользовательские графики (таблица результатов) 1074
13. 3M пользовательские последовательные графики (таблица результатов) 1075
14. Пользовательские XYZ графики (таблица результатов) 1075
15. Пользовательские матричные графики (таблица результатов) 1076
16. Пользовательские пиктографики (таблица результатов) 1076
17. Быстрые статистические графики (таблица результатов) 1077
18. Галерея графиков системы *STATISTICA* 1078

StatSoft®

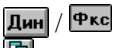
Copyright © StatSoft, 1995

19. Открытие таблицы результатов 1078
20. Сохранение таблицы результатов 1079
21. Печать таблицы результатов (или выделенного блока) 1079
22. Вырезать (блок) 1080
23. Копировать (блок) 1080
24. Вставить (Блок) 1080
25. Справка 1080

ПАНЕЛЬ ИНСТРУМЕНТОВ ОКНА ТЕКСТА/ВЫВОДА 1081

1. Поле просмотра 1082
2. Переключатель модулей 1082
3. Управление Рабочей книгой 1082
 4. Имя и размер шрифта 1082
 5. Форматирование символов 1082
6. Форматирование абзацев 1083
 7. Стартовая панель 1083
 8. Продолжить анализ 1083
9. Печать блока 1084
 10. Галерея графиков системы *STATISTICA* 1084
11. Открытие файла отчета 1084
 12. Сохранение файла отчета 1085
13. Вырезать блок текста и/или график 1085
 14. Копировать блок текста и/или график 1085
15. Вставить текст и/или график 1085
 16. Отменить 1085
17. Справка 1085

ГРАФИЧЕСКАЯ ПАНЕЛЬ ИНСТРУМЕНТОВ 1086

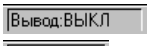
0. Поле просмотра 1086
 1. Динамический /Фиксированный режимы (переключатель) 1086
2. Переключатель модулей 1087
3. Управление Рабочей книгой 1087
4. Редактор данных графика 1087
5. Кисть 1088
6. Фиксировать пропорции графика 1090
7. Изменить пропорции графика 1091
8. Подобрать область графика и поля 1091
9. Прикрепить к сетке 1092
10. Увеличение 1093
11. Уменьшение 1094

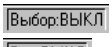
Copyright © StatSoft, 1995

12. Режим выделения (указатель) 1094
13. Редактор текста графика 1095
14. Вставка прямоугольника 1097
15. Вставка скругленного прямоугольника 1097
16. Вставка эллипса 1097
17. Вставка дуги 1097
18. Вставка ломаной линии, произвольное рисование 1098
19. Вставка стрелок или отрезков 1099
20. Вставка объектов (средства *OLE*) 1100
21. Шаблон линии 1102
22. Шаблон заполнения 1102
23. Вынести на передний план 1102
24. Поместить на задний план 1103
25. Увеличить все шрифты на 16% (при нажатой клавише CTRL - на 4%) 1103
26. Уменьшить все шрифты на 16% (при нажатой клавише CTRL – на 4%) 1104
27. Галерея графиков системы *STATISTICA* 1104
28. Открытие графического файла 1105
29. Сохранение графического файла 1105
30. Печать графика 1105
31. Предварительный просмотр 1106
32. Вырезать графический объект 1106
33. Копировать график, графический объект или текст 1106
34. Вставить график, графический объект или текст 1106
35. Справка 1107
36. Вращение графика 1108
37. Динамическое расслоение 1108

СТРОКА СОСТОЯНИЯ (В НИЖНЕЙ ЧАСТИ ОКНА СИСТЕМЫ *STATISTICA*) 1110

0. Индикатор состояния с таймером 1110
Остановка анализа 1110
Многозадачность 1110
Поля контроля строки состояния 1111

1. Поле сообщений 1111


2. Контроль вывода 1111


3. Условия выбора наблюдений 1115

4. Вес наблюдений 1117

5. Макрокоманды (текущее состояние) 1117

ПРЕДМЕТНЫЙ УКАЗАТЕЛЬ 1119

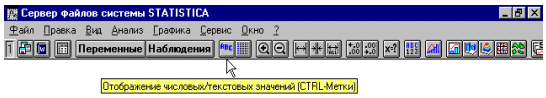
StatSoft®

Copyright © StatSoft, 1995

Глава 2:
ПАНЕЛИ ИНСТРУМЕНТОВ

ВВОДНЫЙ ОБЗОР

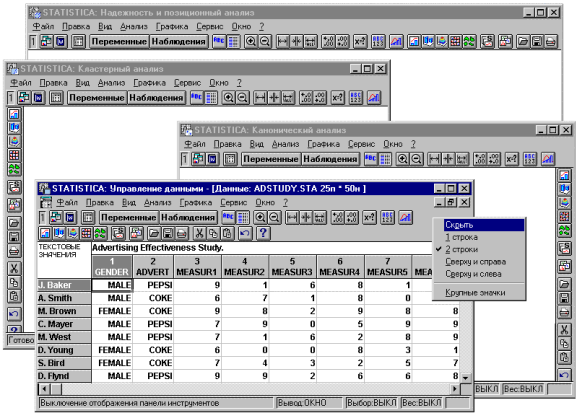
Этот раздел содержит краткое описание функций, которые вызываются с помощью кнопок панели инструментов. Когда курсор указывает на кнопку, ее название отображается в *подсказке* .



Полное описание каждой из этих функций содержится в *Электронном руководстве*. Его можно открыть, нажав клавишу F1, дважды щелкнув на строке состояния в нижней части окна системы *STATISTICA*, или нажав кнопку *Справка* ? панели инструментов. Обратите внимание, что большинство функций, доступных из панели инструментов, вызываются также и другими способами. Например, из выпадающих меню или из контекстных меню, которые появляются при нажатии правой кнопки мыши или комбинации клавиш.

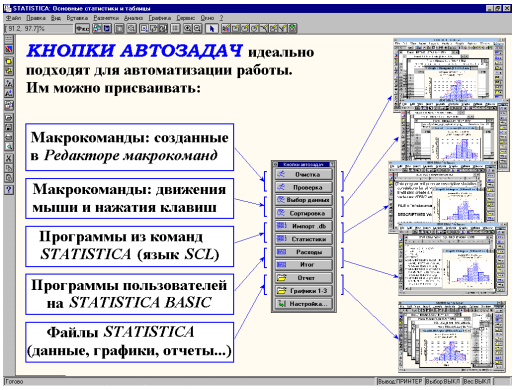
Изменение панелей инструментов. Все панели инструментов системы *STATISTICA* могут быть изменены с помощью команды *Панели инструментов* выпадающего меню *Сервис* или контекстного меню, вызываемого нажатием правой кнопки мыши на панели инструментов.

Например, некоторые панели инструментов содержат больше кнопок, чем может уместиться на стандартном VGA (640x480) экране; поэтому их можно разделить на блоки и изменить их расположение.



Дважды щелкая на панели инструментов за пределами кнопок, можно переключаться между четырьмя конфигурациями.

В дополнение к панелям инструментов таблицы исходных данных, таблицы результатов, *Окна текста/вывода* и графиков, система *STATISTICA* предоставляет *Кнопки автозадач* - всплывающую (или закрепленную) пользовательскую панель инструментов, которая позволяет быстро вызывать большинство средств автоматизации задач системы *STATISTICA*.



Подробную информацию о панели *Кнопки автозадач* см. в главе 6.

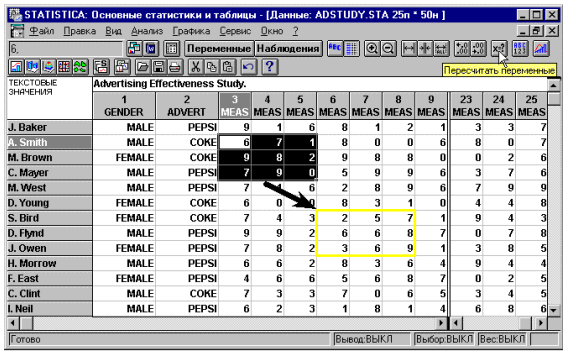
Copyright © StatSoft, 1995

StatSoft

ПАНЕЛЬ ИНСТРУМЕНТОВ ТАБЛИЦЫ ИСХОДНЫХ ДАННЫХ

Окно таблицы исходных данных позволяет изменять вводимые данные и предоставляет большой набор операций по управлению базами данных, преобразованию и перекодированию данных.

Таблица исходных данных системы *STATISTICA* также поддерживает разделение отображения и/или разделение и изменение скорости прокручивания, и расширенный набор стандартных операций с блоками методом *Перетащить и отпустить* (стиль программы MS Excel) и операций по расширению блока.



Некоторые другие специализированные возможности управления и преобразования данных доступны в модуле *Управление данными*.

Панель инструментов таблицы исходных данных предоставляет быстрый доступ к часто используемым функциям управления данными и графикам, которые строятся по исходным данным. Все функции, доступные через приведенные ниже кнопки панели инструментов, также могут быть вызваны другими методами (см. выше).

0. Поле просмотра

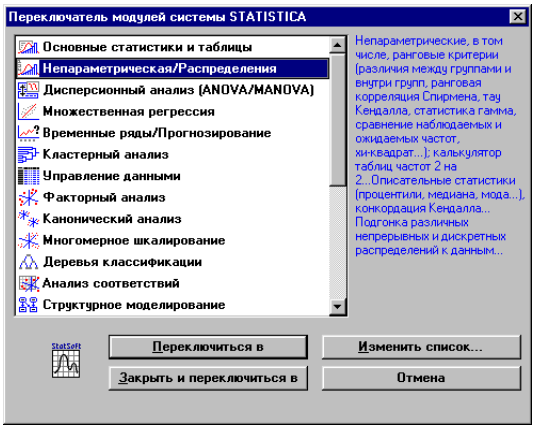
В окне *Поле просмотра* значение выделенной ячейки отображается с большим количеством десятичных символов, чем умещается в соответствующем столбце таблицы исходных данных. Ширина окна *Поле просмотра* может быть изменена (переключена) щелчком мыши на этом поле

3.12 3.123456 3.12345678901237

или выбором команды *Изменить поле просмотра* из выпадающего меню *Вид*.

1. Переключатель модулей

Первая кнопка любой панели инструментов вызывает окно *Переключатель модулей* системы *STATISTICA*. Он предоставляет возможность быстрого доступа ко всем “модулям” (т.е. группам аналитических процедур), доступным в используемой версии системы. *Переключатель* работает так же, как и *Переключатель задач* Windows (вызываемый двойным щелчком на Рабочем столе Windows), и может быть открыт двойным щелчком в любом месте фона окна системы *STATISTICA*.



StatSoft®
Copyright © StatSoft, 1995

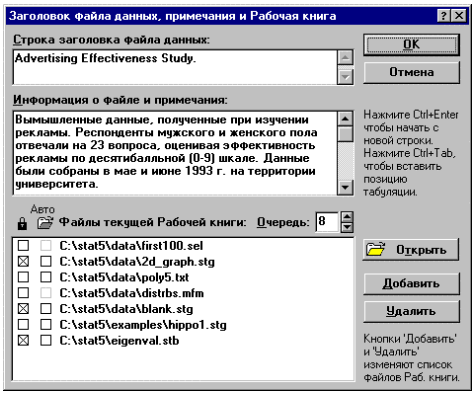
В зависимости от текущей конфигурации *переключатель* может открывать новые модули в этом же или в новых окнах.

Обратите внимание, что хотя список модулей в окне *Переключатель модулей* можно прокручивать, удобнее переместить часто используемые модули в начало списка, чтобы прокручивания не требовалось. Для изменения порядка модулей в списке, нужно нажать кнопку *Изменить список...* (в окне *Переключатель модулей*).

2. Управление Рабочей книгой

Открывает диалоговое окно *Заголовок файла данных, примечания и Рабочая книга* (это вторая кнопка на любой панели инструментов). Это диалоговое окно позволяет ввести одну строку *Заголовка файла,* которая отображается в верхней части таблицы исходных данных, подробную информацию о файле или замечания о содержании набора данных, их источнике и т.д., а также управлять всеми дополнительными файлами (например, графиками, отчетами, программами), которые используются с текущим набором данных.

Рабочие книги (обзор). Файлы данных системы *STATISTICA* могут рассматриваться как "Рабочие книги" файлов, поскольку они содержат (автоматически сохраняют) информацию обо всех дополнительных файлах (например, графиках, отчетах, программах), которые используются совместно с текущим набором данных. Список этих файлов можно посмотреть, нажав кнопку *Рабочая книга* панели инструментов или дважды щелкнув на области названия (заголовка) таблицы исходных данных.

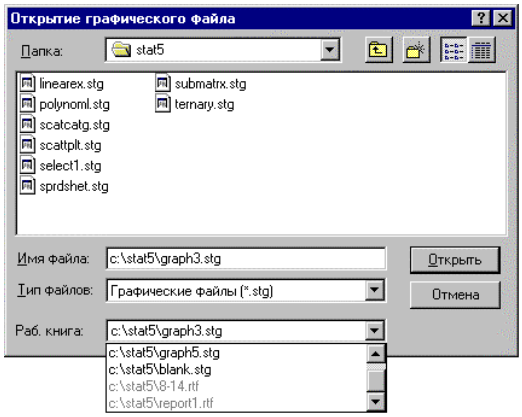


Некоторые файлы из списка Рабочей книги можно автоматически открывать при открытии файла данных. Для этого нужно поставить пометку рядом с именем файла в поле под значком *Автоматически открывать* (Авто).

Очередь файлов Рабочей книги. Список недавно открытых или созданных дополнительных файлов хранится как "очередь", в которой старые файлы удаляются из списка по мере того, как сохраняются новые графики, отчеты и программы. Файлы удаляются из очереди на основе правила "первый пришел, первый ушел". Длина этой очереди сохраняется вместе с набором данных и может изменяться пользователем в показанном выше диалоговом окне *Заголовок файла данных, примечания и Рабочая книга* (по умолчанию она равна 32 файлам). Вне зависимости от длины очереди, отдельные файлы могут быть "заблокированы" (т.е. "удалены" из очереди), если поставить пометку рядом с именем файла под значком *Заблокировать*.

Список дополнительных файлов (Рабочая книга), связанных с текущим набором данных также доступен из каждого диалогового окна *Открытие/Сохранение* системы *STATISTICA* (см. нижнюю часть показанного ниже диалогового окна).

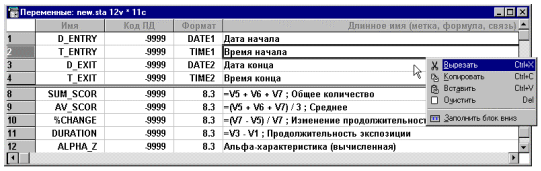
Copyright © StatSoft, 1995



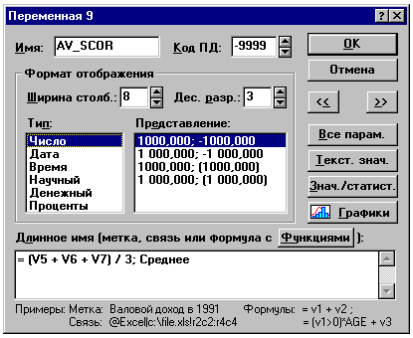
Соответствующие текущему диалоговому окну *Открытие/Сохранение* имена файлов не будут затенены в этом списке, и их можно выбрать двойным щелчком мыши.

3. Спецификации всех переменных (таблица)

Открывает объединенное окно *Переменные: (Спецификации переменных)*. В этом окне можно просмотреть и изменить таблицу всех спецификаций переменных (имена, форматы, значения пропущенных данных, длинные метки, формулы или связи и другие параметры, определяющие отдельные переменные).



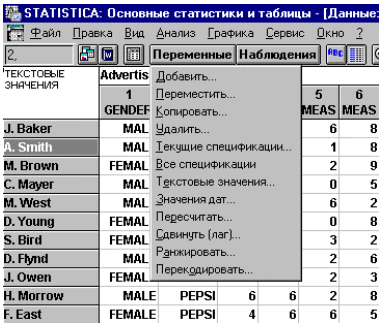
Обратите внимание, что подробные спецификации для каждой переменной в отдельности можно вызывать, дважды щелкнув на имени переменой в таблице исходных данных, в результате чего откроется диалоговое окно спецификаций выбранной переменной.



В этом диалоговом окне спецификаций отдельной переменной можно также снова перейти в таблице спецификаций всех переменных, нажав кнопку *Все парам.*

Переменные 4. Глобальные операции с переменными

Вызывает меню для глобального изменения структуры *переменных*, включающее следующие пункты: *Добавить*, *Переместить*, *Копировать*, *Удалить*, *Текущие спецификации*, *Все спецификации* (см. выше кнопку), *Текстовые значения*, *Значения дат*, *Пересчитать*, *Сдвинуть (лаг)*, *Ранжировать*, *Перекодировать*.



***Операции с буфером обмена и глобальные операции* с переменными.** В отличие от операций вырезания, копирования и вставки блоков данных (или содержания целых столбцов) через буфер обмена, глобальные операции влияют не только на содержание столбцов данных, но и на сами столбцы (там, где

StatSoft®
Copyright © StatSoft, 1995

это возможно). Например, операция *Удалить* удаляет не только содержание выделенного диапазона столбцов, но также и сами столбцы.

Обратите внимание на разницу между этими глобальными операциями с переменными (рассматриваемыми как логические элементы файлов данных системы *STATISTICA*) и всеми операциями с таблицей исходных данных, осуществляемыми через буфер обмена, которые действуют точно так же, как и в стандартных электронных таблицах (например, MS Excel).

Например, глобальные операции копирования, удаления или перемещения переменных, вызываемые из этого меню, не зависят от текущего положения курсора или блока (выделения), выделенные имена переменных выводятся в соответствующих диалоговых окнах (предлагая тем самым быстрый метод выделения переменных для операции). Операции всегда будут производиться для всех наблюдений выделенных переменных, независимо от того, выделены ли для соответствующих переменных все наблюдения или только их часть.

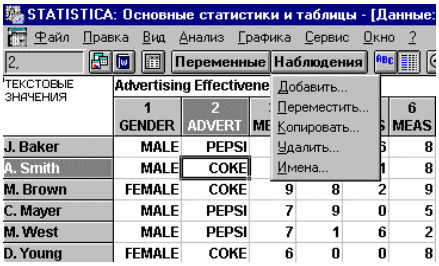
С другой стороны, в случае использования операций с буфером обмена, будет скопирована только часть выделенных данных и (следуя общим соглашениям об электронных таблицах) вставлена с текущего положения курсора.

Поэтому, даже если (a) выделить и скопировать всю переменную, (b) выделить другой (весь) столбец и затем (c) попытаться вставить содержание буфера обмена на новое место (“заменяя” предыдущие значения), то операция пройдет так, как задумано, только в том случае, если курсор был помещен в верхней части нового столбца. Если курсор был где-то в середине, то данные будут вставлены вниз с этого места.

Операции метода *Перетащить и отпустить*. Обратите внимание, что все операции метода *Перетащить и отпустить с вставкой* (с нажатой клавишей SHIFT) изменяют размер файла данных; то же верно и для простого перетаскивания блоков за границы текущего файла данных. Однако перетаскивание (без *вставки*) внутри границ текущего файла данных дает такие же результаты, как и использование буфера обмена (только сам буфер обмена при этом не используется).

Наблюдения 5. Глобальные операции с наблюдениями

Вызывает меню для глобального изменения структуры *наблюдений*, которое включает следующие пункты: *Добавить*, *Переместить*, *Копировать*, *Удалить*, *Имена*.



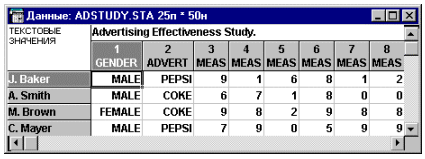
***Операции с буфером обмена и глобальные операции* с наблюдениями.** В отличие от операций вырезания, копирования и вставки блоков данных (или содержания целых строк, то есть наблюдений) через буфер обмена, глобальные операции влияют не только на содержание строк данных, но и на сами строки.

Например, операция *Удалить* удаляет не только содержание выделенного диапазона строк, но также и строки (подробное описание и примеры различий между глобальными операциями и операциями через буфер обмена, а также операциями методом *Перетащить и отпустить*, см. в предыдущем разделе, посвященном описанию кнопки Переменные).

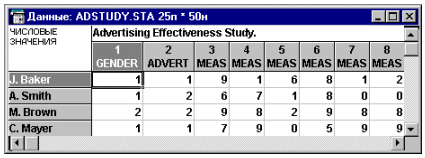
Copyright © StatSoft, 1995

6. Отображение числовых/текстовых значений (для отображения меток значений нужно нажать клавишу CTRL)

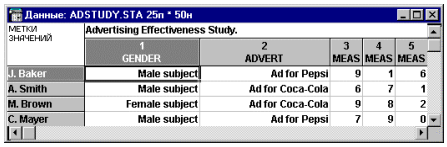
Переключает отображение текстовых значений в таблице исходных данных



на отображение их числовых эквивалентов



или отображение меток значений (для этого нужно, удерживая нажатой клавишу CTRL, нажать на кнопку).

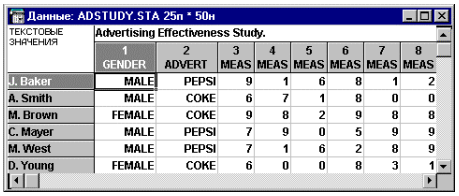


Система *STATISTICA* поддерживает “двойную запись” значений, когда каждое из них может иметь одновременно числовой (например, *1*) и текстовый (например, *MALE*) эквивалент (см. ниже описание окна *Диспетчер текстовых значений*, кнопка).

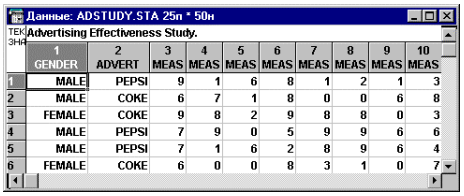
Метки значений. Обратите внимание, что каждое значение может также иметь длинную метку (комментарий), которая автоматически используется в отчетах (в зависимости от текущих параметров отчета). Эти длинные метки (комментарии) можно отображать и изменять в таблице исходных данных (для их отображения нужно, удерживая нажатой клавишу CTRL, нажать кнопку) или ими можно управлять и изменять их в окне *Диспетчер текстовых значений*.

7. Отображение имен наблюдений (переключение)

Позволяет переключать отображение имен наблюдений в таблице исходных данных



на отображение их номеров.

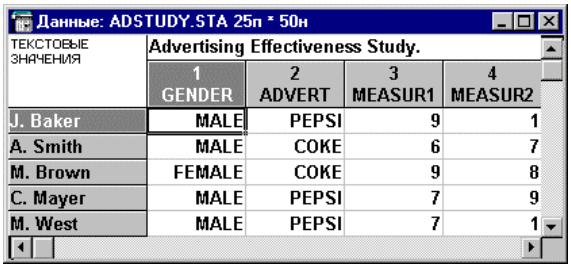


Чтобы вводить или изменять (например, копировать из переменной, варьировать ширину и пр.) текущие имена наблюдений, нужно дважды щелкнуть на любом имени наблюдения таблицы исходных данных.

Если текущий набор данных не содержит имен наблюдений, можно создать их, дважды щелкнув на столбце номеров наблюдений или изменив мышью размер этого столбца.

8. Увеличение

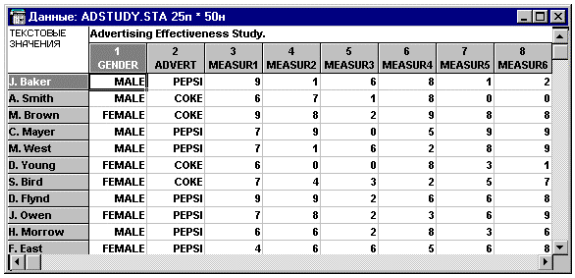
Увеличивает таблицу (увеличивает размер шрифта и пропорционально увеличивает все части таблицы исходных данных).



StatSoft®
Copyright © StatSoft, 1995

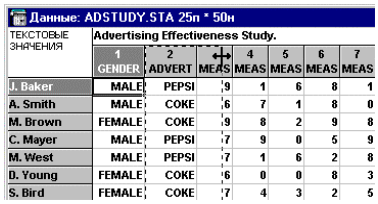
9. Уменьшение

Уменьшает таблицу (уменьшает размер шрифта и пропорционально уменьшает все части таблицы исходных данных).



10. Увеличить ширину столбца

Увеличивает ширину текущего столбца (т.е., столбца, в котором находится курсор). Такой же результат можно получить, перетаскивая мышью правую границу заголовка (имени) столбца.



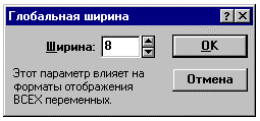
Новая ширина столбца показывается пунктирной линией (как показано выше на рисунке).

11. Уменьшить ширину столбца

Уменьшает ширину текущего столбца (т.е., столбца, в котором находится курсор). Такой же результат можно получить, перетаскивая мышью правую границу заголовка (имени) столбца (см. предыдущий раздел).

12. Изменение глобальной ширины столбцов

Вызывает диалоговое окно для задания ширины всех столбцов таблицы исходных данных.



Этот параметр определяет формат отображения всех переменных текущего файла данных.

13. Добавить десятичный разряд

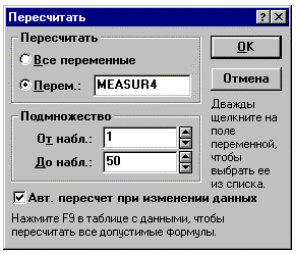
Добавляет один десятичный разряд к формату отображения значений текущей переменной (например, после нажатия этой кнопки значение *12,23* будет отображаться как *12,234*).

14. Удалить десятичный разряд

Удаляет один десятичный разряд из формата отображения значений текущей переменной (например, после нажатия этой кнопки значение *12,234* будет отображаться как *12,23*).

15. Пересчитать

Пересчитывает текущую переменную или, по выбору, все заданные формулами переменные в текущем файле данных (чтобы не открывать это диалоговое окно и пересчитать все правильно введенные формулы, можно нажать клавишу F9).



Эти формулы можно вводить как часть спецификаций переменой. Чтобы ввести или изменить спецификации, нужно дважды щелкнуть на имени переменной в таблице исходных данных или нажать кнопку *Спецификации переменой* панели инструментов (см. выше описание кнопки).

Copyright © StatSoft, 1995

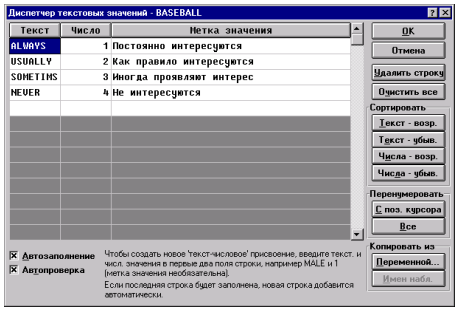
StatSoft

Обратите внимание, что если нужно перекодировать значения (а не производить арифметические операции), необходимо использовать отдельную функцию, вызываемую кнопкой **Переменные**.

16. Диспетчер текстовых значений

Открывает диалоговое окно *Диспетчер текстовых значений* с текстовыми значениями и длинными метками значений (комментариями) для текущей переменной.

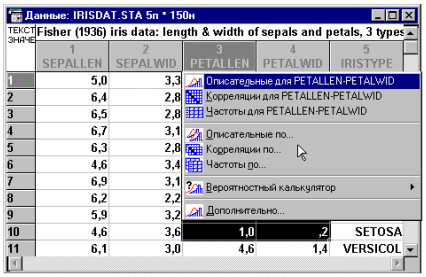
Окно *Диспетчер текстовых значений* можно использовать для изменения и управления (например, для сортировки, копирования между переменными, заполнения, перегруппировки), а также для выполнения операции присваивания между текстовыми и числовыми значениями и их метками.



Это окно также можно вызвать, нажав кнопку *Текст. знач.* в диалоговом окне спецификаций отдельной переменной (см. стр. 1056), или командой *Текущие спецификации* из меню, вызываемом кнопкой **Переменные** панели инструментов.

17. Быстрые основные статистики

Эти процедуры включают набор основных статистик, которые можно вычислить по длинным спискам переменных (например, матрицы корреляции для всех переменных в наборе данных), и все типы анализов, которые можно проводить *по группам*, для каждого значения (т.е., кода) выбранной группирующей переменной.



Меню *Быстрые основные статистики* также доступно из контекстного меню таблицы исходных данных и из выпадающего меню *Анализ* любого модуля системы *STATISTICA* (а также из таблиц результатов, см. стр. 1074). Меню *Быстрые основные статистики* можно вызвать в любой момент анализа данных (например, для получения дополнительной информации при просмотре вывода любого модуля системы *STATISTICA*).

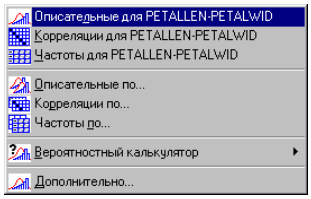
Подробное описание меню *Быстрые основные статистики* см. в главе 10.

"Быстрый" выбор переменных в меню *Быстрые основные статистики.* Одним из основных преимуществ меню *Быстрые основные статистики* является то, что здесь, как правило, нет необходимости в выборе переменных для анализа с помощью стандартных диалоговых окон выбора переменных. Вместо этого переменные автоматически выбираются из выделенного блока таблицы исходных данных или таблицы результатов. Система *STATISTICA* запрашивает выбор переменных только в том случае, если в блоке недостаточно переменных или нужна группирующая переменная.

Во всех таблицах исходных данных (и тех таблицах результатов, которые включают только один список переменных, а не два списка или группирующую переменную и список)

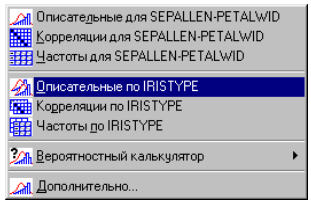
StatSoft®
Copyright © StatSoft, 1995

контекстное меню будет выглядеть следующим образом:



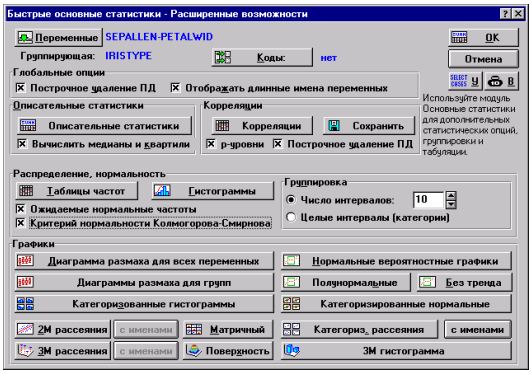
Если выбрать любой анализ, включающий группирующую переменную (*по…*), система *STATISTICA* откроет стандартное диалоговое окно *Выберите группирующую переменную*, в котором можно выбрать переменную, по которой будет производиться анализ.

Для таблиц результатов матричного формата или формата, в котором положение курсора может указывать на две переменные, а не на одну, из меню *Быстрые основные статистики* можно выбрать вариант анализа *по группам* (при этом система *STATISTICA* не будет выводить запрос выбора "*по…*" переменной для анализа).



Если положение курсора не указывает ни на одну переменную, то любая команда из меню *Быстрые основные статистики* выведет запрос на выбор переменной из списка.
Дополнительные статистики и графики вызываются с помощью пункта *Дополнительно…* из этого меню.

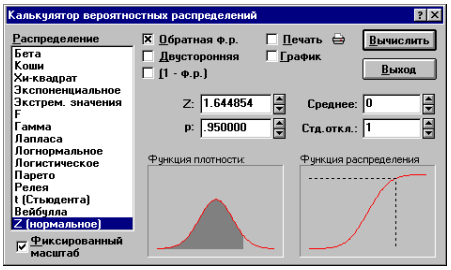
Параметры в этом диалоговом окне отменяют настройки выбора переменных по умолчанию (основанные на выделенном блоке данных).



При вычислениях с помощью команд меню *Быстрые основные статистики* система *STATISTICA* принимает во внимание текущие условия выбора наблюдений и взвешивания для анализируемых переменных (эти условия можно также изменить в диалоговом окне *Быстрые основные статистики – Расширенные возможности*).

См. также меню *Блоковые статистики* (глава 3), которые вычисляют описательные статистики и создают статистические графики по данным выделенного блока.

Вероятностный калькулятор. Эта команда открывает *Калькулятор вероятностных распределений* с интерактивными графиками функций распределения и плотностей, которые автоматически обновляются при изменении параметров.



Этот достаточно гибкий инструмент можно использовать для исследования форм множества теоретических распределений. Кроме того, он

Copyright © StatSoft, 1995

предоставляет процедуры для вычисления вероятностей для различных тестовых статистик, а также критических значений (например, критическое значение для распределения *хи-квадрат*, которое значимо при уровне *p=0,05* для определенного числа степеней свободы *df*).

18-23. Кнопки вызова графиков таблицы исходных данных

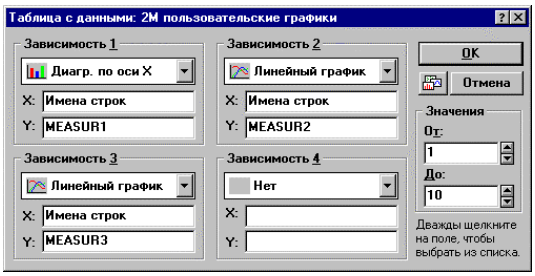
С помощью любой из этих шести кнопок можно построить графики для таблицы исходных данных. Аналогичные кнопки присутствуют и на панели инструментов таблиц результатов(см. стр. 1071). Небольшие различия между кнопками для двух типов таблиц описаны в следующих разделах.

Однострочный заголовок файла данных, расположенный в поле заголовков таблицы исходных данных, переносится и на *пользовательский* (но не на *статистический*) график (этот текст можно удалить или отредактировать).

Динамическая связь с данными. При изменении данных в таблице (или при динамическом обмене данными (*DDE)* все графики, построенные с помощью этих кнопок, автоматически обновляются). Для любого из графиков можно изменить режим связи с данными с автоматического на ручной (для этого используйте команду *Связи данных и графика* из меню *Графика*). Для всех графиков способ обновления можно изменить или отменить (из меню *Сервис*). При сохранении каждый из этих графиков автоматически добавляется к списку Рабочей книги. В диалоговом окне *Рабочая книга* (см. стр. 1055) можно задать режим автоматического открытия данного графического файла вместе с его файлом данных.

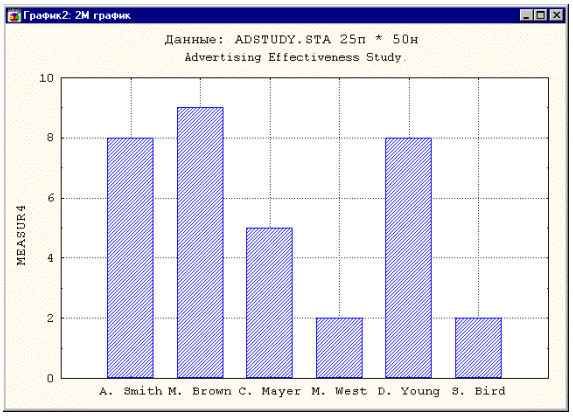
18. 2M пользовательские графики (таблица исходных данных)

Открывает диалоговое окно *2M пользовательские графики.* Здесь задаются параметры для построения различных двумерных графиков по любой комбинации наблюдений и/или переменных (или их групп) текущей таблицы исходных данных.



Для задания графика можно использовать не только наблюдения и переменные, но и их названия. Поэтому здесь возможно построение не только обычных (научных) графиков на плоскости , но и простых последовательных ("деловых") графиков.

По умолчанию в качестве меток осей (для обычных графиков) или последовательных точек (на последовательных графиках) используются имена либо наблюдений, либо переменных в зависимости от типа графика.



StatSoft

Copyright © StatSoft, 1995

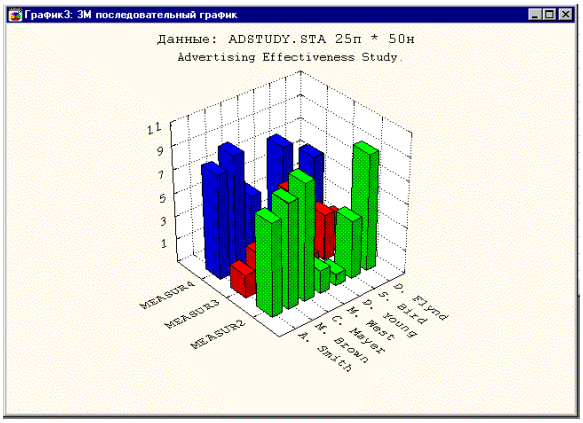
Значения параметров, задаваемые программой при открытии диалогового окна, определяются положением курсора или выделением блока данных в текущей таблице. По умолчанию будет построен график для выделенного блока данных.

Информацию об автоматической связи с файлом данных можно найти на стр. 1062.

19. 3M пользовательские последовательные графики (таблица исходных данных)

Открывает диалоговое окно *3M пользовательские последовательные графики*. Здесь задаются параметры для построения различных трехмерных последовательных графиков по любой комбинации наблюдений и/или переменных (или их групп) текущей таблицы исходных данных.

В отличие от *XYZ графиков* (см. описание следующей кнопки), где тройки значений *X*, *Y* и *Z* интерпретируются как координаты точек в трехмерном пространстве, на последовательных графиках в трехмерном виде (например, в виде столбцов, блоков или лент) представлена простая последовательность значений выбранного ряда данных.



В качестве меток последовательностей на оси *Y* отображаются имена наблюдений или переменных (в зависимости от способа построения) (например, *Measur2*, *Measur3*, *Measur4*).

На оси *X* показаны последовательные номера значений в последовательности данных (или имена наблюдений). Сами значения ("высоты" точек) откладываются по оси *Z*. Значения параметров, задаваемые программой при открытии диалогового окна, определяются положением курсора или выделением блока данных в текущей таблице исходных данных. По умолчанию будет построен график для выделенного блока.

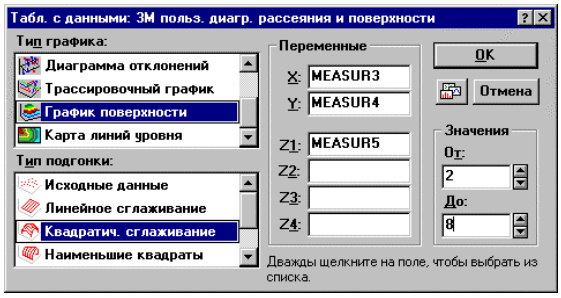
Информацию об автоматической связи с файлом данных можно найти на стр. 1062.

20. Пользовательские XYZ графики (таблица исходных данных)

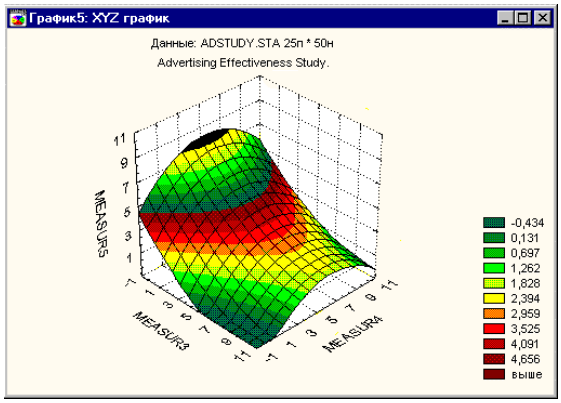
Открывает диалоговое окно *3M пользовательские диаграммы рассеяния и поверхности*. Здесь задаются параметры для построения различных трехмерных графиков рассеяния и поверхностей по любой комбинации наблюдений и/или переменных (или их групп) текущей таблицы исходных данных.

Copyright © StatSoft, 1995

StatSoft



В отличие от *3М последовательных графиков* (см. предыдущую кнопку), где просто представлена последовательность данных таблицы, на *XYZ графиках* , триплеты значений *X*, *Y* и *Z* (или несколько значений *Z*) интерпретируются как координаты точек в трехмерном пространстве.

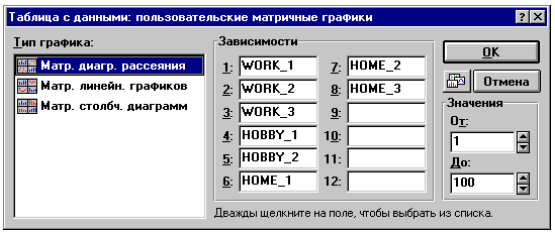


Для обозначения осей *X* и *Y* могут быть использованы не только наблюдения и переменные, но и их имена. Поэтому имена наблюдений и переменных в зависимости от выбранного типа графика могут обозначать как сами оси, так и последовательные точки на них. На оси *Z* откладываются значения наблюдений или переменных. Значения параметров, задаваемые программой при открытии диалогового окна, определяются положением курсора или выделением блока данных в текущей таблице исходных данных. По умолчанию будет построен график для выделенного блока.

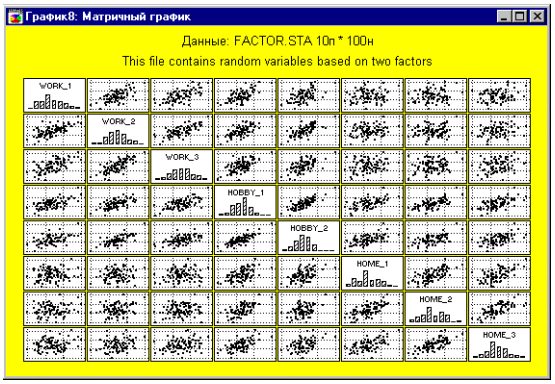
Информацию об автоматической связи с файлом данных можно найти на стр. 1062.

21. Пользовательские матричные графики (таблица исходных данных)

Вызывает диалоговое окно *Пользовательские матричные графики.* Здесь задаются параметры для построения различных матричных графиков по любой комбинации наблюдений и/или переменных (или их групп) текущей таблицы исходных данных.



На каждом из графиков матрицы выбранные данные (наблюдения или переменные таблицы исходных данных) рассматриваются как "переменные" и строится их зависимость друг от друга. Вид этой зависимости определяется выбранным типом графика.

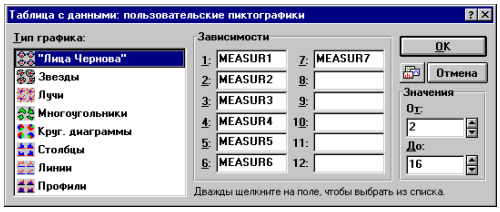


StatSoft®
Copyright © StatSoft, 1995

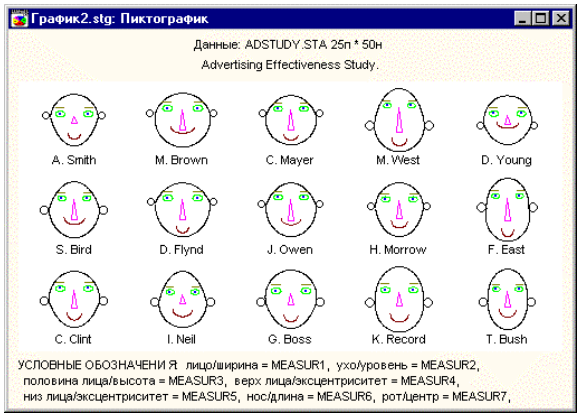
На таких графиках можно представить корреляционные связи не только между столбцами (переменными), но и между строками таблицы (т.е. наблюдениями).

22. Пользовательские пиктографики (таблица исходных данных)

Эта кнопка вызывает диалоговое окно *Пользовательские пиктографики.* Здесь задаются параметры для построения различных пиктографиков по любой комбинации наблюдений и/или переменных (или их групп) текущей таблицы исходных данных.



Последовательность пиктограмм на этих графиках представляет конфигурацию значений наблюдений или переменных, выбранных для построения графика. Пиктограммы обозначены именами соответствующих наблюдений или переменных.

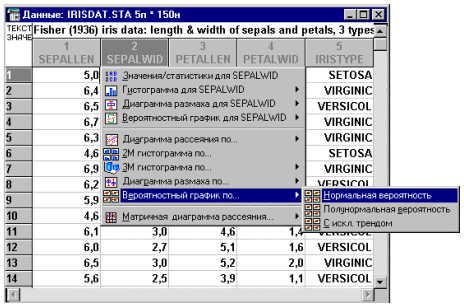


Значения параметров, задаваемые программой при открытии диалогового окна, определяются положением курсора или выделением блока данных в текущей таблице. По умолчанию будет построен график для выделенного блока.

Информацию об автоматической связи с файлом данных можно найти на стр. 1062.

23. Быстрые статистические графики (таблица исходных данных)

Вызывает список *Быстрые статистические графики* для текущей переменной (т.е. той, на которую указывает курсор в таблице). В отличие от *пользовательских графиков* (см. пять предыдущих кнопок), которые дают возможность наглядно представлять численные результаты с помощью определенных пользователем типов графиков, *быстрые статистические графики* - это заранее заданные типы графического представления статистических данных. Они включают стандартные методы графического представления исходных данных (например, на графиках рассеяния, гистограммах или на графиках центральных тенденций), а также стандартные графические аналитические методы (например, построение категоризованных вероятностных зависимостей, зависимостей с исключенным трендом или графиков доверительных интервалов линий регрессии).

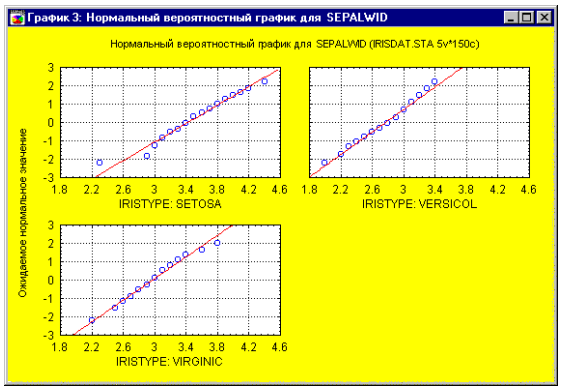


Если для выбранного типа графика требуется больше одной переменной, то появится диалоговое окно выбора переменных, чтобы

Copyright © StatSoft, 1995

StatSoft

можно было задать недостающие параметры. Например, на показанном ниже категоризованном графике выбранная переменная построена при различных значениях некоторой другой переменной.



При построении этих графиков учитываются текущие условия выбора наблюдений и их веса (см. описание полей *Выбор* и *Вес* в строке состояния окна системы *STATISTICA*).

Меню *Быстрые статистические графики* можно вызвать и из окна большинства таблиц результатов. Но при построении графиков двух переменных (например, категоризованных) по данным этих таблиц программа делает выбор переменных автоматически, чтобы получить значимое представление результатов.

Сравнение *быстрых статистических* и *статистических графиков*. *Быстрые статистические графики* легко вызываются с панели инструментов или контекстных меню таблиц исходных данных и таблиц результатов. Они включают наиболее часто используемые типы *статистических графиков*. Сами *статистические графики* доступны только из *Галереи графиков* (см. описание следующей кнопки) и из выпадающего меню *Графика*. Меню *Быстрые статистические графики* не предоставляет такой широкий спектр возможностей, как меню *Статистические графики*, но значительно облегчает и ускоряет процедуру построения графика. В отличие от *статистических графиков быстрые статистические графики*:

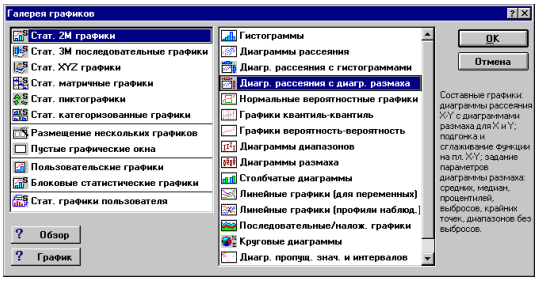
- вызываются из контекстных меню или с панели инструментов (обычно они не требуют обращения к выпадающим меню или диалоговым окнам),
- не требуют от пользователя выбора переменных (этот выбор определяется текущим положением курсора в таблице), и
- не требуют промежуточной настройки параметров (формат соответствующих графиков определяется по умолчанию).

Об автоматической связи с файлом данных можно прочитать на стр. 1062.

Блоковые статистические графики. Эти графики можно вызывать только из контекстных меню. Они представляют собой статистическое обобщение данных по строкам или столбцам выделенного блока таблицы.

24. Галерея графиков системы *STATISTICA*

Открывает диалоговое окно *Галерея графиков* системы *STATISTICA*.



Эта кнопка присутствует в диалоговом окне каждого типа графиков и позволяет быстро и легко вызывать любые *статистические и пользовательские графики*, *пустые графические окна* и *статистические графики пользователя*. Для этого нужно выделить название нужного

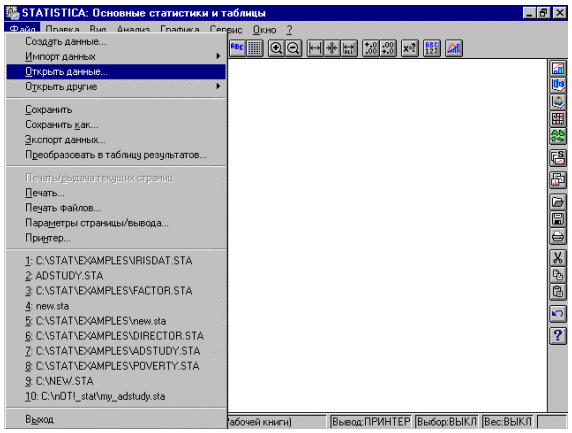
StatSoft®
Copyright © StatSoft, 1995

типа графика и дважды щелкнуть на нем (или нажать кнопку *OK*).

25. Открыть файл данных

Открывает окно выбора файлов, в котором можно выбрать файл данных (расширение **.sta*), а затем открывает выбранный файл в окне таблицы исходных данных (заменяя текущее окно). Файлы данных системы *STATISTICA* можно также открыть непосредственно из окна *Диспетчер файлов* двойным щелчком на имени файла.

Обратите внимание, что к выпадающему меню *Файл* автоматически добавляется список недавно использованных файлов, как показано ниже.



Длину этого списка можно изменить в диалоговом окне *Параметры по умолчанию: общие настройки*, вызываемом из выпадающего меню *Сервис*.

26. Сохранить файл данных

Сохраняет текущий файл данных под тем же самым или другим именем (по умолчанию расширение имени файла **.sta*). Если перезаписывается уже существующий файл, то по умолчанию, предыдущий файл переименовывается в резервный (с расширением **.bak* вместо **.sta*).

27. Печать файла данных (или выделенного блока)

Печатает данные из текущей таблицы исходных данных. Это кратчайший способ отправить текущий файл данных в канал вывода: принтер, файла или *Окно текста/вывода* (в зависимости от текущих параметров вывода, см. поле *Вывод* строки состояния в нижней части окна программы, стр. 1111).

В отличие от команды *Печать* (или клавиш CTRL+P) выпадающего меню *Файл*, эта кнопка не открывает диалоговое окно *Печать данных* (содержащее опции настройки формата и содержания списка данных). Вместо этого в канал вывода будет отправлен весь файл (или выделенный блок, если он есть). Формат списка данных будет соответствовать текущим настройкам *Печати* (доступным из выпадающего меню *Файл*). Информацию о стилях отчета, форматировании, изменении полей, печати сетки и других параметрах см. в разделе о строке состояния (поле *Вывод*, стр. 1111).

28. Вырезать (блок)

Вырезает (удаляет) содержание выделенных ячеек и копирует их в буфер обмена. Очищенные ячейки заменяются значениями пропущенных данных соответствующих переменных.

***Операции с буфером обмена и глобальные операции* с переменными или наблюдениями.** Обратите внимание, что в отличие от *глобальных* операций с диапазонами наблюдений или переменных, где они рассматриваются как логические элементы (см. кнопки **Переменные** и **Наблюдения**, стр. 1056 и стр. 1057 соответственно), операции с буфером

Copyright © StatSoft, 1995

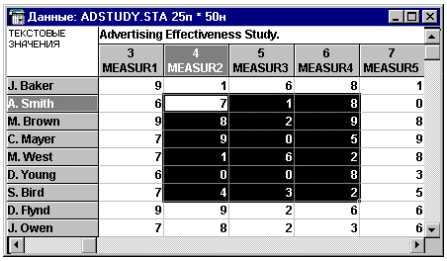
обмена следуют стандартным соглашениям об электронных таблицах и применяются только к содержанию выделенных ячеек. Они полностью зависят от текущего (выделенного) блока и положения курсора.

Например, обратите внимание, что *глобальная* операция удаления или перемещения переменных (см. кнопку **Переменные**, стр. 1056) воздействует не только на содержание соответствующих столбцов, но и на сами столбцы, поэтому структура файла данных при этом изменяется. Эти операции всегда производятся со всей переменной независимо от того, выделены ли для выбранных переменных все наблюдения или только их часть. Однако в случае использования операций с буфером обмена, только часть выделенных данных будет скопирована и (следуя общим соглашениям об электронных таблицах) вставлена с текущего положения курсора.

Например, необходимо осуществить следующую операцию перемещения переменной: (a) выделить и вырезать всю переменную, (b) выделить другой (весь) столбец и затем (c) вставить содержимое буфера обмена на новое место ("заменяя" предыдущие значения). Эта операция пройдет так, как задумано, только в том случае, если курсор был помещен в верху нового столбца. В противном случае данные будут вставлены вниз, начиная с того места, где был расположен курсор.

29. Копировать (блок)

Копирует выделенную часть таблицы исходных данных в буфер обмена (чтобы выделить всю таблицу исходных данных, нужно щелкнуть в ее верхний левый угол). Эта опция не только копирует содержание выделенного блока ячеек, но и имена (или номера) наблюдений и переменных, связанных с выделенными ячейками (чтобы скопировать только содержание выделенного блока ячеек без соответствующих имен наблюдений и переменных, нужно воспользоваться командой *Копировать содержание* выпадающего меню *Правка*).



Копировать:

	MEASUR2	MEASUR3	MEASUR4
A. Smith	7	1	8
M. Brown	8	2	9
C. Mayer	9	0	5
M. West	1	6	2
D. Young	0	0	8
S. Bird	4	3	2

Копировать содержание:

7	1	8
8	2	9
9	0	5
1	6	2
0	0	8
4	3	2

Обратите внимание, что блоки, скопированные из таблиц исходных данных системы *STATISTICA,* разделены символами табуляции. Поэтому при вставке в документ в текстовом редакторе они также будут разделены символами табуляции (совместимыми с таблицами текстовых редакторов Windows, например, MS Word).

Техническое замечание. Для совместимости с другими приложениями система *STATISTICA* автоматически создает различные форматы буфера обмена. Например, создаются форматы электронных таблиц, которые используются при вставке содержания буфера обмена (скопированного из системы *STATISTICA*) в электронные таблицы в MS Excel или Lotus для Windows.

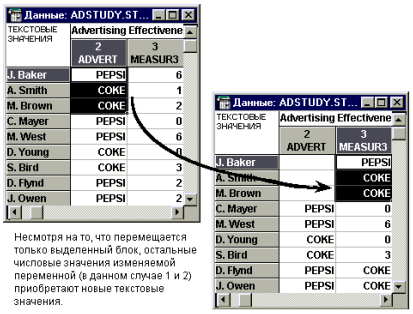
StatSoft®

Copyright © StatSoft, 1995

30. Вставить (Блок)

Вставляет текущее содержание буфера обмена в таблицу исходных данных, начиная с текущего положения курсора. См. выше в разделе, посвященном операции *Вырезать* (кнопка), описание различий между *глобальными* операциями над наблюдениями и переменными (использующими их как логические элементы, см. кнопки **Переменные** и **Наблюдения**, стр. 1056 и 1057 соответственно) и операции с буфером обмена, которые следуют стандартным соглашениям об использовании буфера обмена в электронных таблицах. Например, при копировании (через буфер обмена) всей переменной на новое место, курсор нужно поместить в верхней части конечного столбца, поскольку данные всегда вставляются вниз от текущего положения курсора (даже если выделен весь конечный столбец).

Операции с буфером обмена в таблице исходных данных. При копировании или перемещении блока в таблице исходных данных (например, с помощью метода *Перетащить и отпустить*) режим отображения таблицы будет определять, какие именно значения будут скопированы. Если при копировании блока в буфер обмена в таблице отображаются числовые значения (или метки значений), то будут скопированы только числовые значения (или метки значений). Если при копировании блока в буфер обмена в таблице отображаются текстовые значения, то в буфер обмена будут скопированы не только текстовые значения, но и соответствующие числовые значения и метки значений (если они есть). Это может привести к тому, что текстовые значения будут присвоены числовым значениям, у которых раньше не было текстовых эквивалентов (см. ниже).



Даты (и время). В файлах данных системы *STATISTICA* форматы отображения значений являются свойствами переменных, а не отдельных ячеек. Поэтому при вставке в систему *STATISTICA* блока данных, который содержит форматированные значения дат (или дату/время) (например, из MS Excel), даты будут отображаться в юлианском формате (т.е. как целые числа, например, *34092* вместо *Май 3, 1993*), если только для соответствующей переменной не установлен формат *Дата*.

31. Отмена

Эта кнопка позволяет отменить последнюю команду или действие в таблице исходных данных (изменение, перемещение или копирование блоков, заполнение произвольными значениями, перекодирование или ранжирование переменных, и т.д.). Таблицы исходных данных системы *STATISTICA* поддерживают многоуровневую отмену (с 16 буферами); поэтому, выбирая эту опцию последовательно (до 16 раз), можно отменять несколько действий.

Комбинация клавиш: CTRL+Z (или ALT+BKSP)

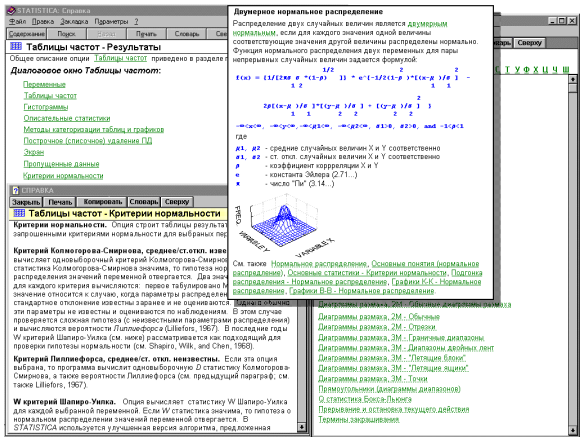
32. Справка

Последняя кнопка на всех панелях инструментов вызывает *Электронное руководство* системы *STATISTICA*, которое содержит исчерпывающую электронную документацию на все функции программы и все контекстно-зависимые опции. Оно вызывается также при нажатии клавиши F1

Copyright © StatSoft, 1995

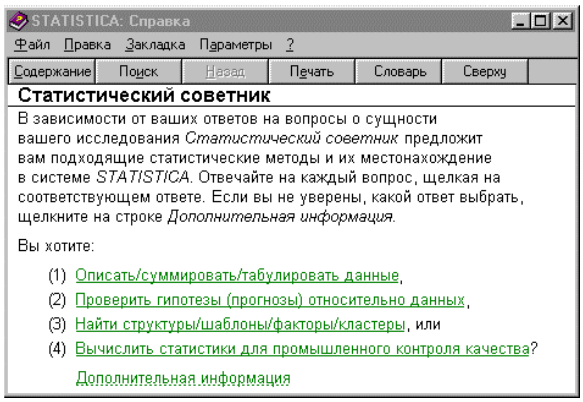
StatSoft

или кнопки ? на панели инструментов или в заголовке любого диалогового окна (прилагается более 10 мегабайт документации в упакованном виде). Гипертекстовая организация (и различные функции, позволяющие настраивать справочную систему) позволяет использовать *Электронное руководство* быстрее и эффективнее, чем документацию, поставляемую в печатном виде.



Справка также вызывается двойным щелчком на области справочных сообщений строки состояния в нижней части окна системы *STATISTICA* (в строке состояния также отображаются краткое описание опций выпадающего меню или кнопок панели инструментов, которое появляется при выделении опции или нажатии кнопки). Чтобы получить подробную информацию, нужно нажать клавишу справки (F1), когда выделена или выбрана соответствующая опция.

Статистический советник. Функция *Статистический советник* встроена в справочную систему. Если выбрать пункт *Советник* из выпадающего меню *?(Справка)*, программа выведет несколько вопросов о природе исследуемой проблемы и типе данных, а затем предложит наиболее подходящие статистические процедуры (и подскажет, где их искать в системе *STATISTICA*).

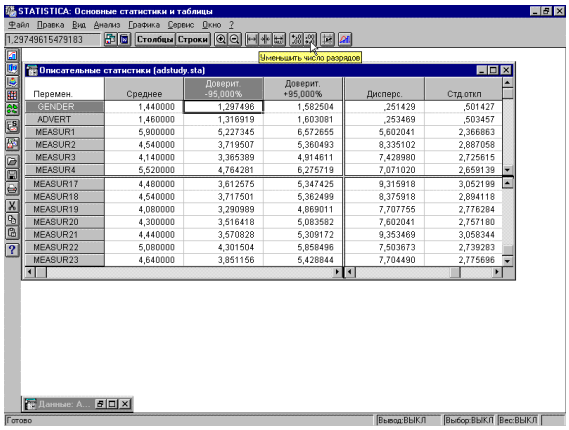


StatSoft®

Copyright © StatSoft, 1995

ПАНЕЛЬ ИНСТРУМЕНТОВ ТАБЛИЦЫ РЕЗУЛЬТАТОВ (ПРОКРУЧИВАЕМОГО ОКНА ВЫВОДА)

Таблицы результатов являются электронными таблицами, используемыми в системе *STATISTICA* для отображения числового и текстового вывода. Таблицы результатов могут содержать вывод размером от одной строки до нескольких мегабайт и предоставляют различные средства упрощения просмотра результатов, их визуализации на заранее заданных и заданных пользователем графиках, а также преобразования их в отчеты, которые можно использовать для презентаций. Таблицы результатов поддерживают операции метода *Перетащить и отпустить*, экстраполяции значений блока, а также разделение и изменение скорости прокрутки, которое показано ниже.



Панель инструментов таблицы результатов предоставляет быстрый доступ к часто используемым графикам и функциям управления выводом данных таблиц результатов.

Все приведенные ниже функции контроля, вызываемые через панель инструментов, также вызываются через меню и нажатием комбинаций клавиш (поэтому их можно использовать в клавиатурных макрокомандах). Состав панелей инструментов можно изменять (см. стр. 1053).

0. Поле просмотра

В окне *Поле просмотра* значение из выделенной ячейки отображается с большим количеством десятичных символов, чем умещается в соответствующем столбце.

Ширина окна *Поле просмотра* может быть изменена (переключена) щелчком мыши на этом поле или выбором команды *Изменить поле просмотра* из выпадающего меню *Вид*.

1. Переключатель модулей

Вызывает окно *Переключатель модулей* системы *STATISTICA* (эта кнопка есть во всех панелях инструментов). Он предоставляет возможность быстрого доступа ко всем “модулям” (т.е. группам аналитических процедур), доступным в используемой версии системы *STATISTICA*. В зависимости от текущей конфигурации *переключатель* может открывать новые модули в этом же или в новых окнах.

2. Управление Рабочей книгой

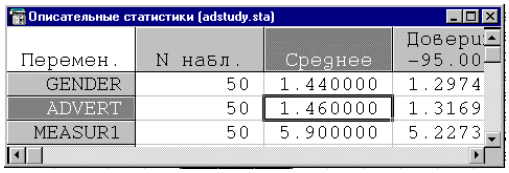
Открывает диалоговое окно *Заголовок файла данных, примечания и Рабочая книга* (это вторая кнопка на любой панели инструментов). Это диалоговое окно позволяет ввести одну строку в поле *Заголовок файла,* которая отображается в верхней части таблицы результатов, подробную информацию о файле или замечания о содержании набора данных, их источнике и т.д., а также управлять всеми дополнительными файлами (например, графиками, отчетами, программами), которые используются с текущим набором данных. Подробную информацию о Рабочих книгах см. на стр. 1055.

Copyright © StatSoft, 1995

StatSoft

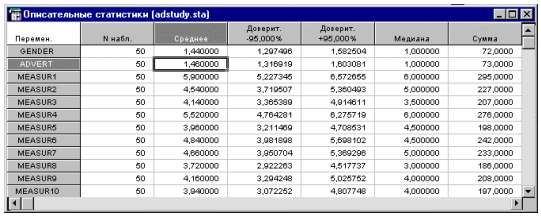
3. Увеличение

Увеличивает таблицу (пропорционально увеличивает все элементы и размер шрифта таблицы результатов).



4. Уменьшение

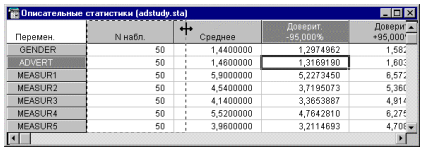
Уменьшает таблицу (пропорционально уменьшает все элементы и размер шрифта таблицы результатов).



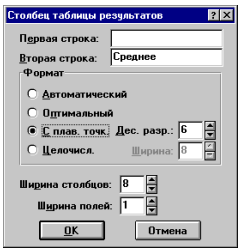
Эта опция может оказаться полезной при просмотре больших таблиц, в которых выделены данные, удовлетворяющие конкретным критериям (например, значимые коэффициенты корреляции) (см. ниже кнопку).

5. Увеличить ширину столбца

Увеличивает ширину текущего столбца (постепенно). Такой же результат можно получить, перетаскивая мышью любую из границ столбцов таблицы результатов.



Обратите внимание, что ширину столбца, также как и ширину полей, формат отображения значений и имена столбцов можно изменить в диалоговом окне *Столбец таблицы результатов*. Это диалоговое окно вызывается двойным щелчком на имени соответствующего столбца таблицы результатов (аналогично двойному щелчку на имени переменной в таблице исходных данных).



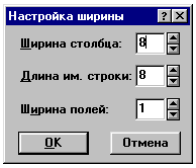
Например, показанное выше диалоговое окно открывается после двойного щелчка на столбце *Среднее*, выделенном в таблице результатов, приведенной в предыдущем разделе.

6. Уменьшить ширину столбца

Уменьшает ширину текущего столбца (постепенно, см. предыдущую кнопку). Такой же результат можно получить, перетаскивая мышью любую из границ столбцов таблицы результатов.

7. Изменить ширину столбцов и полей

Изменяет на указанное число ширину столбцов, полей, а также имени строки в (первом) столбце таблицы результатов.



Поля, которые можно изменять в этом диалоговом окне, отделяют активное поле отображения от линий сетки.

StatSoft®
Copyright © StatSoft, 1995

8. Добавить десятичный разряд

Добавляет один десятичный разряд к формату отображения значений текущей переменной (например, после нажатия этой кнопки значение *12,23* будет отображаться как *12,234*).

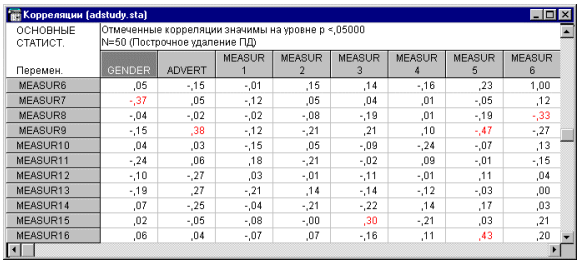
9. Удалить десятичный разряд

Удаляет один десятичный разряд из формата отображения значений текущей переменной (например, после нажатия этой кнопки значение *12,234* будет отображаться как *12,23*).

10. Маркировать ячейки

Маркирует (или снимает маркировку) отдельные значения или блок значений (если он выделен) в таблице результатов. Маркированные значения будут выделяться цветом, отличающимся от цвета остальных данных. По умолчанию маркированные значения отображаются красным, но этот параметр можно изменить с помощью опции выпадающего меню *Вид*, или опции *Экран* выпадающего меню *Сервис*. При печати маркированные ячейки будут обозначаться символом * (звездочка) рядом со значением. Другие варианты выделения помеченных ячеек в печатаемых или сохраняемых отчетах можно выбрать в диалоговом окне *Печать таблицы результатов*, где предлагаются различные виды печати таблиц результатов.

Ячейки, маркированные системой *STATISTICA*. Многие процедуры системы *STATISTICA* автоматически маркируют определенные ячейки или блоки в таблицах результатов, чтобы “выделить” некоторые результаты (например, необычайно высокие частоты в таблице частот, статистически значимые коэффициенты корреляции в матрице корреляций, или статистически важные эффекты в таблице эффектов ANOVA). Например, на следующей таблице результатов корреляции между *Gender* и *Measur7*, и между *Advert* и *Measur9* помечены другим цветом (указывая значимость при $p<0,05$):


Корреляции (adstudy.sta)

ОСНОВНЫЕ СТАТИСТ. — Отмеченные корреляции значимы на уровне p <,05000; N=50 (Построчное удаление ПД)

Перемен.	GENDER	ADVERT	MEASUR 1	MEASUR 2	MEASUR 3	MEASUR 4	MEASUR 5	MEASUR 6
MEASUR6	,05	-,15	-,01	,15	,14	-,16	,23	1,00
MEASUR7	-,37	,05	-,12	,05	,04	,01	-,05	,12
MEASUR8	-,04	-,02	-,02	-,08	-,19	,01	-,19	-,33
MEASUR9	-,15	,38	-,12	-,21	,21	,10	-,47	-,27
MEASUR10	,04	,03	-,15	,05	-,09	-,24	-,07	,13
MEASUR11	-,24	,06	,18	-,21	-,02	,09	-,01	-,15
MEASUR12	-,10	-,27	,03	-,01	-,11	-,01	,11	,04
MEASUR13	-,19	,27	-,21	,14	-,14	-,12	-,03	,00
MEASUR14	,07	-,25	-,04	-,21	-,22	,14	,17	,03
MEASUR15	,02	-,05	-,08	-,00	,30	-,21	,03	,21
MEASUR16	,06	,04	-,07	,07	-,16	,11	,43	,20

Такие таблицы результатов обычно предоставляют возможность изменить критерий, использованный для выбора выделяемых элементов (например, *p*-уровень для коэффициентов корреляции), с помощью кнопки **Параметры** панели инструментов таблицы результатов, в которой отображаются маркированные ячейки.

Выделенные значения в окне *Редактор данных графика*. *Редактор данных графика* - это специальная таблица результатов, связанная с каждым графиком, которая содержит набор значений, представленных на графике. В этом окне подсвечены (т.е. отображены другим цветом) те значения, которые *выделены* на графике с помощью процедуры *закрашивания* (см. описание кнопки на стр. 1088).

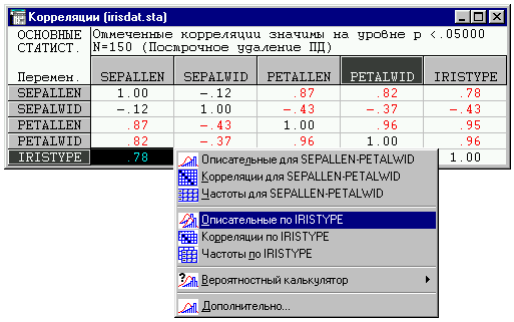
Обратите внимание, что термин *подсвеченный* имеет специальное значение при использовании в контексте закрашивания точек данных на графике. Здесь *подсвеченные* точки данных (на графике) - это те точки, которые выбраны с помощью мыши, но еще не *маркированы*, *помечены* или *выключены* с помощью соответствующей *операции* закрашивания (например, добавление меток или маркировка).

Copyright © StatSoft, 1995

StatSoft

11. Быстрые основные статистики

Как и в таблицах исходных данных (см. стр. 1060), эта кнопка панели инструментов вызывает динамическое меню *Быстрые основные статистики* для переменных текущей таблицы результатов.



Информацию о выборе переменных из меню *Быстрые основные статистики* см. на стр. 1060.

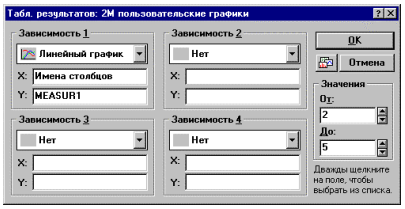
12-17. Кнопки вызова графиков таблицы результатов

С помощью любой из шести описанных далее кнопок можно построить графики по данным таблицы результатов. Аналогичные кнопки присутствуют и на панели инструментов таблиц исходных данных (см. описание предыдущей панели инструментов). Небольшие различия между кнопками для двух типов таблиц описаны в следующих разделах.

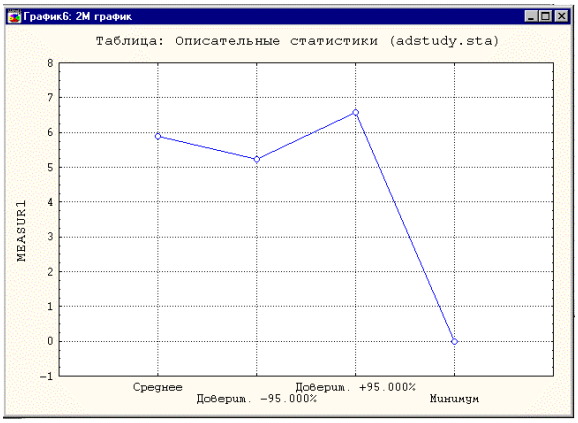
Текст из поля заголовков таблицы результатов переносится на *пользовательский* (но не на *статистический*) график (этот заголовок можно удалить или отредактировать).

12. 2M пользовательские графики (таблица результатов)

Открывает диалоговое окно *2M пользовательские графики.* Здесь задаются параметры для построения различных двумерных графиков по любой комбинации строк и/или столбцов (или их групп) текущей таблицы



Для задания параметров графика можно использовать не только столбцы и строки, но и их названия. Поэтому здесь возможно построение не только обычных (научных) графиков на плоскости, но и простых последовательных ("деловых") графиков. По умолчанию в качестве меток осей (для обычных графиков) или последовательных точек (на последовательных графиках) в зависимости от типа графика используются либо имена строк, либо столбцов.

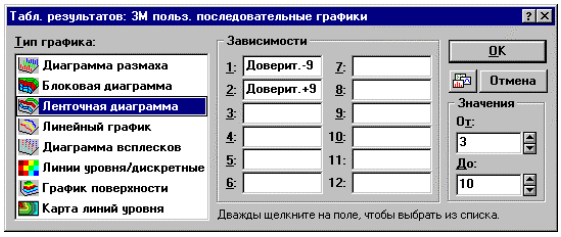


Значения параметров, задаваемые программой при открытии диалогового окна, определяются положением курсора или выделением блока данных в текущей таблице. По умолчанию будет построен график для выделенного блока данных.

StatSoft®
Copyright © StatSoft, 1995

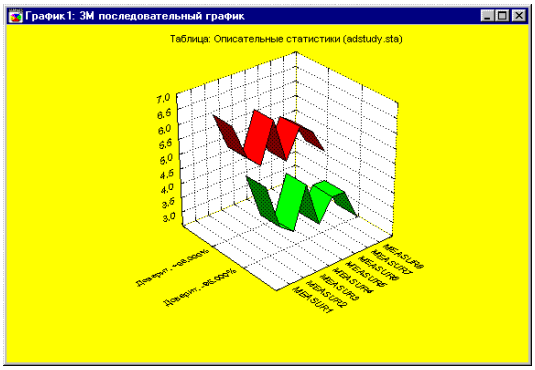
13. 3М пользовательские последовательные графики (таблица результатов)

Открывает диалоговое окно *3М пользовательские последовательные графики*. Здесь задаются параметры для построения различных трехмерных последовательных графиков по любой комбинации строк и/или столбцов (или их групп) текущей таблицы результатов.



В отличие от *XYZ графиков* (см. описание следующей кнопки), где тройки значений *X*, *Y* и *Z* интерпретируются как координаты точек в трехмерном пространстве, на последовательных графиках в трехмерном виде (например, в виде столбцов, блоков или лент) представлена простая последовательность значений выбранного ряда данных.

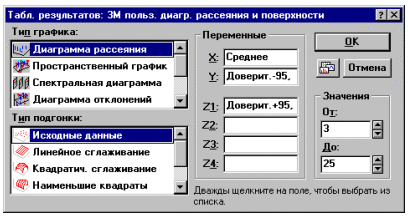
В качестве меток последовательностей на оси *Y* отображаются имена строк или столбцов таблицы результатов (в зависимости от способа построения). На оси *X* показаны последовательные номера значений в последовательности данных (или названия строк). Сами значения ("высоты" точек) откладываются по оси *Z*.



Значения параметров, задаваемые программой при открытии диалогового окна, определяются положением курсора или выделением блока данных в текущей таблице исходных данных. По умолчанию будет построен график для выделенного блока.

14. Пользовательские XYZ графики (таблица результатов)

Открывает диалоговое окно *3М пользовательские диаграммы рассеяния и поверхности*. Здесь задаются параметры для построения различных трехмерных графиков рассеяния и поверхностей по любой комбинации строк и/или столбцов (или их групп) текущей таблицы результатов.

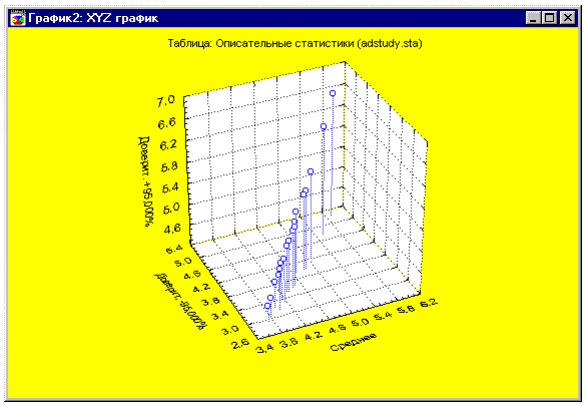


В отличие от *3М последовательных графиков* (см. описание предыдущей кнопки), где просто представлена последовательность данных таблицы, на *XYZ графиках* , триплеты значений *X*, *Y* и *Z* (или несколько значений *Z*) интерпретируются как координаты точек в

Copyright © StatSoft, 1995

StatSoft

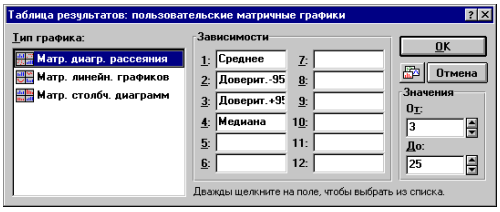
трехмерном пространстве. Для обозначения осей *X* и *Y* могут быть использованы не только строки и столбцы, но и их имена. Поэтому названия строк и столбцов в зависимости от выбранного типа графика могут обозначать как сами оси, так и последовательные точки на них. На оси *Z* откладываются значения из строк или столбцов.



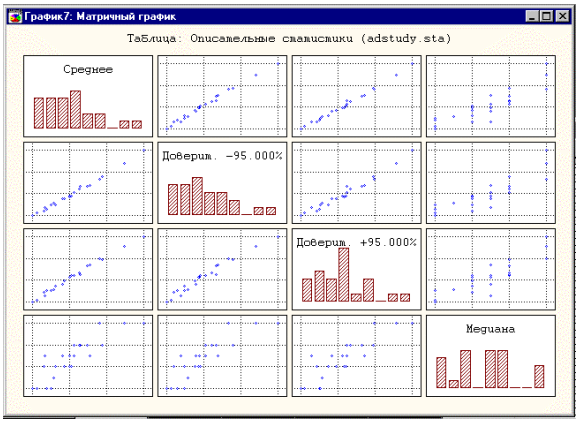
Значения параметров, задаваемые программой при открытии диалогового окна, определяются положением курсора или выделением блока данных в текущей таблице исходных данных. По умолчанию будет построен график для выделенного блока.

15. Пользовательские матричные графики (таблица результатов)

Эта кнопка вызывает диалоговое окно *Пользовательские матричные графики*. Здесь задаются параметры для построения различных матричных графиков по любой комбинации строк и/или столбцов (или их групп) текущей таблицы результатов.



На каждом из графиков матрицы выбранные данные (строки или столбцы таблицы результатов) рассматриваются как "переменные" и строится их зависимость друг от друга. Вид этой зависимости определяется выбранным типом графика.



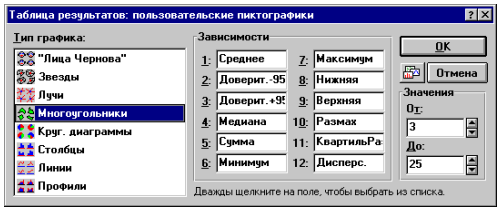
На таких графиках можно представить корреляционные связи не только между столбцами, но и между строками таблицы.

16. Пользовательские пиктографики (таблица результатов)

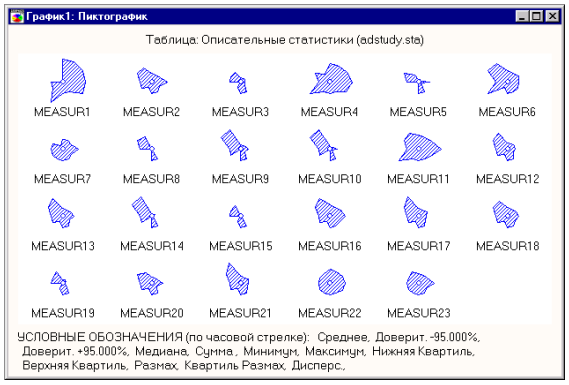
Эта кнопка вызывает диалоговое окно *Пользовательские пиктографики* Здесь задаются параметры для построения различных пиктографиков по любой комбинации строк и/или столбцов (или их групп) текущей таблицы результатов.

StatSoft®

Copyright © StatSoft, 1995



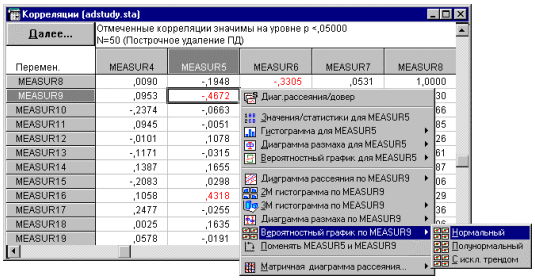
Последовательность пиктограмм на этих графиках представляет конфигурацию значений строк или столбцов, выбранных для построения графика. Пиктограммы обозначены названиями соответствующих строк или столбцов.



Значения параметров, задаваемые программой при открытии диалогового окна, определяются положением курсора или выделением блока данных в текущей таблице. По умолчанию будет построен график для выделенного блока.

17. Быстрые статистические графики (таблица результатов)

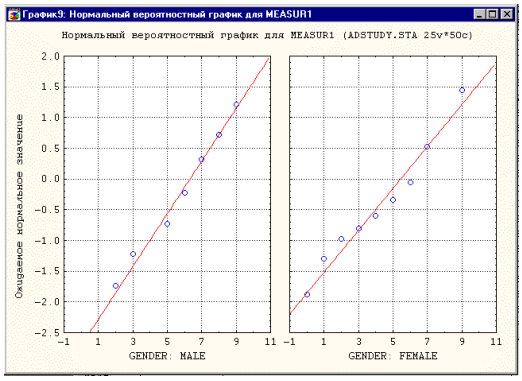
Эта кнопка вызывает список *Быстрые статистические графики* для текущей таблицы результатов.



В отличие от *пользовательских графиков* (см. пять предыдущих кнопок), которые дают возможность наглядно представлять численные результаты с помощью определенных пользователем типов графиков, *быстрые статистические графики* - это заранее заданные типы графического представления статистических данных. Они предлагают специальные типы графического представления результатов текущей таблицы результатов (например, пиктографики остатков) или стандартные методы графического представления исходных данных (например, графики рассеяния, гистограммы, графики центральных тенденций, категоризованные вероятностные зависимости, зависимости с исключенным трендом). Для некоторых таблиц результатов предлагаются нестандартные графики. В большинстве случаев переменная для отображения на графике определяется положением курсора в таблице результатов и типом этой таблицы.

Например, в таблицах результатов с описательными статистиками для графика будет выбрана переменная той строки, на которую указывает курсор. На графиках, где необходимо выбрать больше одной переменной (например, графики рассеяния или все категоризованные графики), этот выбор также определяется положением курсора или общей группирующей переменной для нескольких строк или столбцов таблицы результатов (в зависимости от данных этой таблицы).

Copyright © StatSoft, 1995
StatSoft



Если вторая переменная не может быть автоматически выбрана из таблицы по результатам анализа, то появится диалоговое окно выбора переменных, чтобы можно было задать недостающие параметры.

При построении этих графиков учитываются текущие условия выбора наблюдений и их веса (см. описание полей *Выбор* и *Вес* в строке состояния окна системы *STATISTICA*).

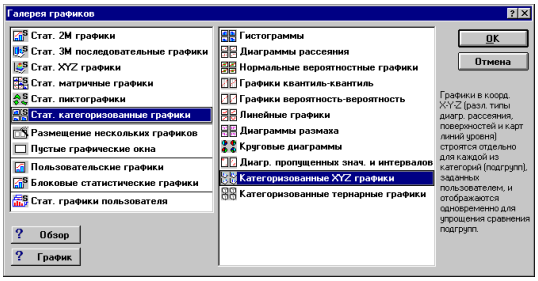
Сравнение *быстрых статистических* и *статистических графиков*. *Быстрые статистические графики* легко вызываются с панели инструментов кнопкой или из контекстных меню таблиц исходных данных и таблиц результатов. Они включают наиболее часто используемые типы *статистических графиков*. Сами *статистические графики* доступны только из *Галереи графиков* (см. описание следующей кнопки) и из выпадающего меню *Графика*. Меню *Быстрые статистические график* не предоставляет такой широкий спектр возможностей, как меню *Статистические графики*, но значительно облегчают и ускоряют процедуру построения графика. В отличие от *статистических графиков быстрые статистические графики*:

- вызываются из контекстных меню или с панели инструментов (обычно они не требуют обращения к выпадающим меню или диалоговым окнам),
- не требуют от пользователя выбора переменных (этот выбор определяется текущим положением курсора в таблице), и
- не требуют промежуточной настройки параметров (формат соответствующих графиков определяется по умолчанию).

Блоковые статистические графики. Эти графики можно вызвать только из контекстных меню или *Галереи графиков*. Они представляют собой статистическое обобщение данных по строкам или столбцам выделенного блока таблицы.

18. Галерея графиков системы *STATISTICA*

Открывает диалоговое окно *Галерея графиков STATISTICA*. Эта кнопка присутствует в диалоговом окне каждого типа графиков и позволяет быстро и легко вызывать любые *статистические и пользовательские графики*, *пустые графические окна* и *статистические графики пользователя*.

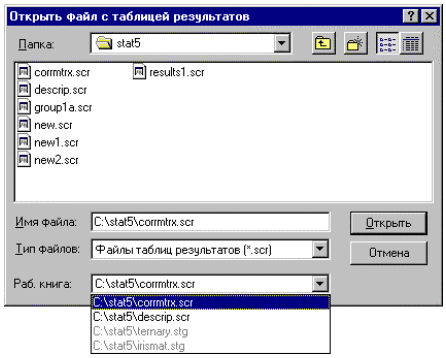


Для этого нужно выделить название нужного типа графика и дважды щелкнуть на нем (или нажать кнопку *OK*). Такой же выбор типов графиков представлен и в выпадающем меню *Графика*.

19. Открытие таблицы результатов

Открывает окно, в котором можно выбрать файл таблицы результатов (расширение **.scr*), и открыть его в новом окне таблицы результатов.

StatSoft®
Copyright © StatSoft, 1995



Обратите внимание, что каждое диалоговое окно *Открытие/Сохранение* системы *STATISTICA* предоставляет быстрый доступ к списку файлов, используемых с текущим набором данных (см. информацию по Рабочим книгам, стр. 1055). Поэтому все файлы (соответствующего типа) Рабочей книги можно легко найти и выбрать двойным щелчком. Файлы Рабочей книги, которые нельзя открыть в текущем диалоговом окне, будут затенены.

20. Сохранение таблицы результатов

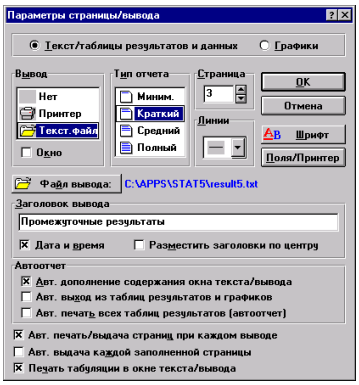
Сохраняет таблицу результатов под тем же именем (если оно есть) или под новым именем (по умолчанию файлы таблиц результатов имеют расширение **.scr*). Информацию о вызове списка файлов Рабочей книги из диалоговых окон *Открытие/Сохранение* см. в предыдущем разделе. При сохранении можно выбрать имя любого файла Рабочей книги и заменить его.

21. Печать таблицы результатов (или выделенного блока)

Печатает данные из текущей таблицы результатов. Это кратчайший способ отправить текущий файл таблицы результатов в канал вывода: принтер, файл или *Окно текста/вывода* (в зависимости от текущих параметров вывода, см. поле *Вывод* строки состояния в нижней части окна системы *STATISTICA*, стр. 1111); ниже описано диалоговое окно для задания этих параметров.

В отличие от команды *Печать* (или клавиш CTRL+P) выпадающего меню *Файл*, эта кнопка не открывает диалоговое окно *Печать таблицы результатов* (содержащее параметры печати таблиц). Вместо этого в канал вывода будет отправлена вся таблица результатов (или выделенный блок). Формат списка данных будет соответствовать текущим настройкам *Печати…* (доступным из выпадающего меню *Файл*).

Автоматическая печать. Обратите внимание, что если необходимо сохранять все таблицы результатов (и/или графики, см. стр. 1111) которые отображаются на экране, не нажимая при этом кнопку *Печать* на каждой таблице результатов, можно выбрать пункт *Авт. печать всех таблиц результатов* в диалоговом окне *Параметры страницы/вывода* (которое вызывается двойным щелчком на поле *Вывод* строки состояния в нижней части окна системы *STATISTICA*, см. стр. 1111).



В этом же диалоговом окне можно указать, куда будет производиться вывод: на принтер, в файл текста/вывода или прокручиваемое *Окно текста/вывода* (см. главу 5). Также можно указать, куда будет производиться графический вывод, который можно автоматически включать в отчет (подробности см. на стр.1111).

Copyright © StatSoft, 1995

StatSoft

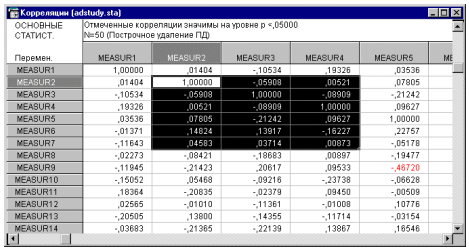
Информацию о стилях отчета, форматировании, печати линий сетки, настройке полей и других возможностях см. в разделе, посвященном описанию поля *Вывод* строки состояния (стр. 1111).

22. Вырезать (блок)

Вырезает (удаляет) содержание выделенных ячеек и копирует их в буфер обмена.

23. Копировать (блок)

Копирует выделенную часть таблицы результатов в буфер обмена (чтобы выделить всю таблицу результатов, нужно щелкнуть в ее левом верхнем углу). Эта функция не только копирует содержание выделенного блока ячеек, но и имена (или номера) наблюдений и переменных, связанных с выделенными ячейками. Чтобы скопировать только содержание выделенного блока ячеек без соответствующих имен наблюдений и переменных, нужно воспользоваться командой *Копировать содержание* выпадающего меню *Правка*).



Копировать:

	MEASUR2	MEASUR3	MEASUR4
MEASUR2	1.00000	-.05908	.00521
MEASUR3	-.05908	1.00000	-.08909
MEASUR4	.00521	-.08909	1.00000
MEASUR5	.07805	-.21242	.09627
MEASUR6	.14824	.13917	-.16227
MEASUR7	.04583	.03714	.00873

Копировать содержание:

1.00000	-.05908	.00521
-.05908	1.00000	-.08909
.00521	-.08909	1.00000
.07805	-.21242	.09627
.14824	.13917	-.16227
.04583	.03714	.00873

Обратите внимание, что блоки, скопированные из таблиц результатов системы *STATISTICA,* разделены символами табуляции. Поэтому при вставке в документ в текстовом редакторе они также будут разделены символами табуляции (совместимыми с таблицами текстовых редакторов Windows, например, MS Word).

Техническое замечание**.** Для повышения совместимости с другими приложениями система *STATISTICA* автоматически создает различные форматы буфера обмена. Например, создаются форматы электронных таблиц, которые используются при вставке содержания буфера обмена (скопированного из системы *STATISTICA*) в электронные таблицы MS Excel или Lotus для Windows.

24. Вставить (Блок)

Вставляет текущее содержание буфера обмена в таблицу результатов, начиная с текущего положения курсора.

25. Справка

Эта кнопка (последняя на всех панелях инструментов) вызывает *Электронное руководство* системы *STATISTICA*. Система предоставляет исчерпывающую электронную документацию на все функции программы и все контекстно-зависимые опции. Оно вызывается также с помощью клавиши F1 или кнопки ? заголовка любого диалогового окна.

StatSoft®

Copyright © StatSoft, 1995

ПАНЕЛЬ ИНСТРУМЕНТОВ ОКНА ТЕКСТА/ВЫВОДА

Окно текста/вывода работает так же, как и текстовый редактор; оно совместимо с форматом *RTF* (расширенный текстовый формат) и поддерживает форматированный текст и графики, а также позволяет управлять текстовыми/графическими отчетами практически неограниченного размера. Это окно можно использовать:

(1) для просмотра и изменения протокола вывода, который отображается в последующих таблицах результатов и/или графиках в процессе анализа (предусмотрена опция для автоматического перенаправления текстового и/или графического вывода в этот отчет);

(2) для открытия (и изменения/просмотра) файлов вывода или любых других текстовых или содержащих графику *RTF* –файлов; а также

(3) как локальный текстовый и графический блокнот.

Три перечисленные выше возможности кратко описаны в следующих абзацах.

(1) Протокол некоторых или всех таблиц результатов и графиков. Чтобы открыть для вывода *Окно текста/вывода*, нужно выбрать пункт *Окно текста/вывода* из выпадающего меню *Вид*. Как только это окно откроется, в нем будет отображаться вывод, который производится из таблиц исходных данных, графиков, таблиц результатов, или выделенного блока ячеек. Этот канал вывода также можно задать в диалоговом окне *Параметры страницы/вывода*, которое вызывается двойным щелчком на поле *Вывод* строки состояния в нижней части окна системы *STATISTICA* (см. стр. 1111). Если в этом диалоговом окне также выбран вывод на *принтер* или в *файл*, вывод будет одновременно направляться в *Окно текста/вывода* и в этот канал. Кроме того, можно выбрать один из режимов *Автоотчета* (см. стр. 1111). В этом случае при отображении на экране таблицы результатов и/или графика они будут одновременно направлены в заданные каналы вывода, а, следовательно, и в *Окно текста/вывода* (если оно открыто).

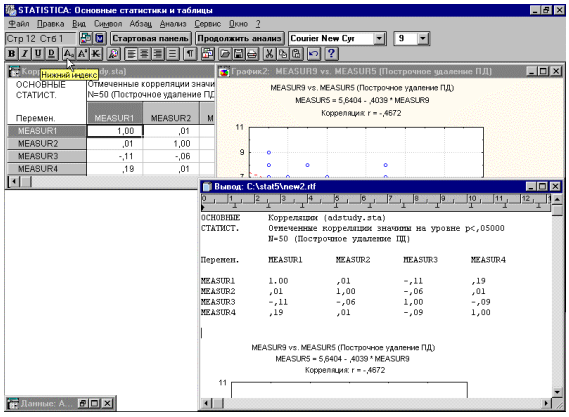
(2) Просмотр текстовых файлов. Независимо от текущего *Окна текста/вывода*, которое было открыто для вывода текущих результатов (см. предыдущий раздел), сохраненные ранее текстовые файлы можно открыть в новых текстовых окнах, выбрав пункт *Файл текста/вывода…* из списка команды *Открыть другие* (выпадающее меню *Файл*). Здесь также можно редактировать любые текстовые (ASCII) и форматированные (**.rtf*) файлы, содержащие графику.

(3) Блокноты. Независимо от того, как было открыто *Окно текста/вывода* (см. два предыдущих абзаца), его всегда можно использовать в качестве блокнота (в нем можно вводить, изменять или вставлять текст, а также добавлять графики или другие рисунки через буфер обмена).

Формат таблиц. Все элементы форматированных таблиц *Окна текста/вывода* разделяются символами табуляции (по умолчанию), и поэтому их можно копировать в электронные таблицы (как значения, но не как текст) и в текстовые редакторы (как исходные данные для формирования таблиц).

Панель инструментов. Панель инструментов *Окна текста/вывода* предоставляет быстрый доступ к наиболее часто используемым функциям управления выводом текста.

Copyright © StatSoft, 1995



Все функции, доступные через приведенные ниже кнопки панели инструментов, также могут быть вызваны из меню и с клавиатуры, что позволяет использовать их в клавиатурных макрокомандах.

1. Поле просмотра

В окошке *Поле просмотра Окна текста/вывода* отображается положение (строка и столбец) курсора внутри окна. Ширину окна *Поле просмотра* можно изменить, щелкая мышью на этом поле.

2. Переключатель модулей

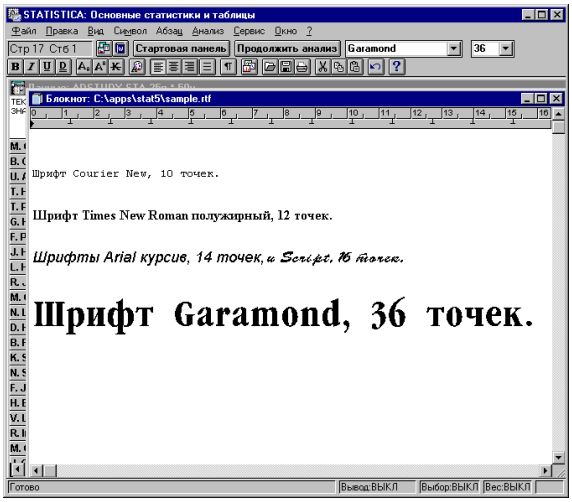
Вызывает окно *Переключатель модулей* системы *STATISTICA* (эта кнопка есть в каждой панели инструментов). *Переключатель* предоставляет возможность быстрого доступа ко всем “модулям” (т.е., группам аналитических процедур), доступным в используемой версии системы *STATISTICA*. В зависимости от текущей конфигурации он может открывать новые модули в этом же или в новых окнах.

3. Управление Рабочей книгой

Открывает диалоговое окно *Заголовок файла данных, примечания и Рабочая книга* (это вторая кнопка на любой панели инструментов). Это диалоговое окно позволяет ввести одну строку в поле *Заголовок файла*, которая отображается в верхней части таблицы исходных данных, подробную информацию о файле или замечания о содержании набора данных, их источнике и т.д., а также управлять всеми дополнительными файлами (например, графиками, отчетами, программами), которые используются с текущим набором данных. Подробную информацию об управлении файлами Рабочей книги см. на стр. 1055.

4. Имя и размер шрифта

Два поля *Выбор шрифта* позволяют задавать тип и размер шрифта для нового или выделенного текста.



Шрифты, размеры шрифтов и атрибуты символов в каждой строке текста можно комбинировать.

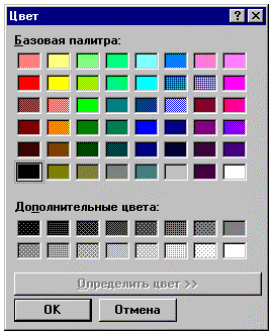
5. Форматирование символов

Эти кнопки позволяют установить форматирование символов для выделенного

StatSoft®
Copyright © StatSoft, 1995

текста (для вызова опций форматирования также существуют комбинации клавиш):

- **Полужирный** шрифт (комбинация клавиш: CTRL+B);
- *Курсив*(комбинация клавиш: CTRL+I);
- Подчеркивание (комбинация клавиш: CTRL+U);
- Двойное подчеркивание (комбинация клавиш: CTRL+D);
- *Нижний индекс* ($_{\text{нижний}}$ индекс);
- *Верхний индекс* (^{верхний} индекс);
- *Зачеркивание* (~~зачеркивание~~);
- Открывает диалоговое окно *Цвет*, в котором можно выбрать цвет для нового или выделенного текста.



Обратите внимание, что цвет и шрифт по умолчанию для текста в *Окне текста/вывода* - можно выбрать в диалоговом окне *Параметры по умолчанию: экран*, которое вызывается из выпадающего меню *Сервис*.

6. Форматирование абзацев

Эта группа кнопок позволяет управлять форматом как создаваемых, так и выделенных абзацев:

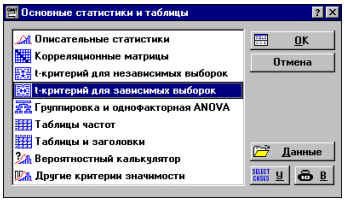
- *Выравнивание по левому полю* текста и графиков;
- *Выравнивание по центру* текста и графиков (между полями);
- *Выравнивание по правому полю* текста;
- Переключение между одиночным и двойным *межстрочным интервалом*;
- Включение и выключение отображения в окне символов абзаца (¶) и скрытого текста.

Стартовая панель

7. Стартовая панель

Вызывает окно *Стартовая панель*, которая является первым диалоговым окном выбора или задания анализа для текущего модуля (аналогично нажатию клавиш CTRL+T, или выбору пункта *Стартовая панель* выпадающего меню *Анализ*).

Ниже показана стартовая панель модуля *Основные статистики и таблицы*:



Если анализ уже выполняется и при его повторном запуске предыдущий выбор будет утерян, система *STATISTICA* выведет запрос на подтверждение запуска нового анализа. Кроме того, можно продолжить и текущий анализ (см. следующую кнопку).

Продолжить анализ

8. Продолжить анализ

Продолжает текущий анализ, открывая окно задания предыдущего анализа или выбора вывода [аналогично нажатию CTRL+*R*, выбору

Copyright © StatSoft, 1995

StatSoft

пункта *Продолжить анализ* выпадающего меню *Анализ* или нажатию кнопки *Далее*].

9. Печать блока

Печатает содержание выделенного блока текста (включая графику) в *Окне текста/вывода*.

Обратите внимание, что функции печати *Окна текста/вывода* имеют особый статус, отличный от таблиц исходных данных и таблиц результатов:

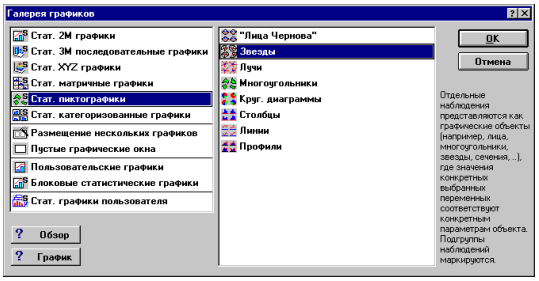
> Можно говорить, что *Окно текста/вывода* одновременно и "инструмент для просмотра" на экране (как таблицы результатов или окна графиков) и "канал вывода" (как принтер или файл вывода), в который можно направлять вывод анализа нажатием кнопки печати.

Печать из *Окна текста/вывода*. Из-за другого статуса этого окна печать из него производится особым образом, отличным от печати из таблиц исходных данных и таблиц результатов. При нажатии кнопки *Печать* все содержимое *Окна текста/вывода* (или выделенного блока) будет отправлено непосредственно на принтер, вне зависимости от текущих настроек поля *Вывод* строки состояния (располагается в нижней части окна системы *STATISTICA*). Например, оно будет отправлено на принтер или в файл, даже если вывод на *принтер* отключен (*Выкл.*). Если необходимо сохранить содержание *Окна текста/вывода* (включая графики, примечания, вставленный текст и т.д.), можно воспользоваться командой *Сохранить* (см. кнопку , ниже).

Самый быстрый и простой способ постоянно сохранять полный отчет обо всем выводе, отображаемом на экране, - выбрать некоторые режимы *Автоотчета* (см. поле *Вывод* строки состояния в нижней части окно системы *STATISTICA*, стр. 1111). При этом содержание всех отображаемых на экране таблиц результатов и/или окон с графиками автоматически перенаправляется в *Окно текста/вывода*. Информацию о форматировании, печати линий сетки и других параметрах вывода см. в разделе о строке состояния (стр. 1111).

10. Галерея графиков системы *STATISTICA*

Открывает диалоговое окно *Галерея графиков* системы *STATISTICA*.



Эта кнопка также присутствует в диалоговом окне каждого типа графиков и позволяет быстро и легко вызывать любые *статистические и пользовательские графики*, *пустые графические окна* и *статистические графики пользователя*.

11. Открытие файла отчета

Открывает в *Окне текста/вывода* выбранный файл отчета [по умолчанию с расширением **.rtf* (расширенный текстовый формат) или **.txt*].

Обратите внимание, что каждое диалоговое окно *Открытие/Сохранение* системы *STATISTICA* предоставляет быстрый доступ к списку всех файлов, используемых с текущим набором данных (см. информацию о Рабочих книгах, стр. 1055), позволяя легко найти и выбрать двойным щелчком любой файл (соответствующего типа) из списка файлов Рабочей книги.

StatSoft®
Copyright © StatSoft, 1995

12. Сохранение файла отчета

Сохраняет содержание текущего *Окна текста/вывода* (текст и графики) в файле отчета (по умолчанию с расширением **.rtf*). Подробную информацию о печати и сохранении файлов текста/вывода можно найти в описании кнопки *Печать* (, выше).

Информацию о вызове списка файлов Рабочей книги из диалоговых окон *Открытие/Сохранение* см. в предыдущем разделе. При сохранении можно выбрать имя существующего файла из списка файлов Рабочей книги, в этом случае он будет заменен.

13. Вырезать блок текста и/или график

Вырезает выбранный (выделенный) блок текста и/или график и помещает его в буфер обмена (информацию о форматах данных буфера обмена см. в следующем разделе).

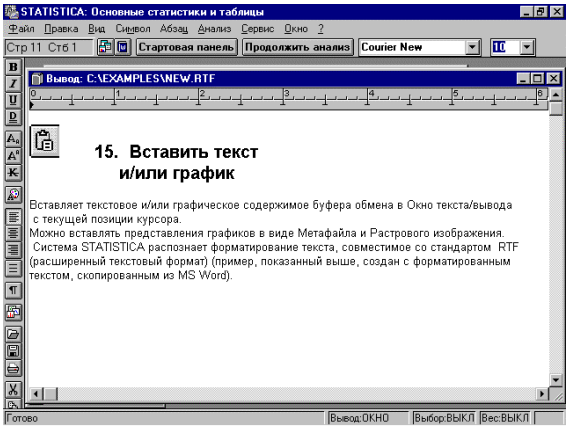
14. Копировать блок текста и/или график

Копирует выбранный (выделенный) блок текста и/или график в буфер обмена.

Техническое замечание: Одним из форматов копируемых объектов (текста и графиков), создаваемых системой *STATISTICA* в буфере обмена, является *RTF* (расширенный текстовый формат); благодаря этому вся информация о форматировании текста легко передается в текстовый редактор и другие приложения, поддерживающие формат *RTF*.

15. Вставить текст и/или график

Вставляет текстовое и/или графическое содержимое буфера обмена в *Окно текста/вывода* с текущей позиции курсора.



Можно вставлять представления графиков в виде *метафайла* и *растрового изображения*. Система *STATISTICA* распознает форматирование текста, совместимое со стандартом *RTF* (расширенный текстовый формат) (пример, показанный выше, создан с форматированным текстом, скопированным из MS Word).

16. Отменить

Отменяет последнее изменение, внесенное в отчет в *Окне текста/вывода*.

17. Справка

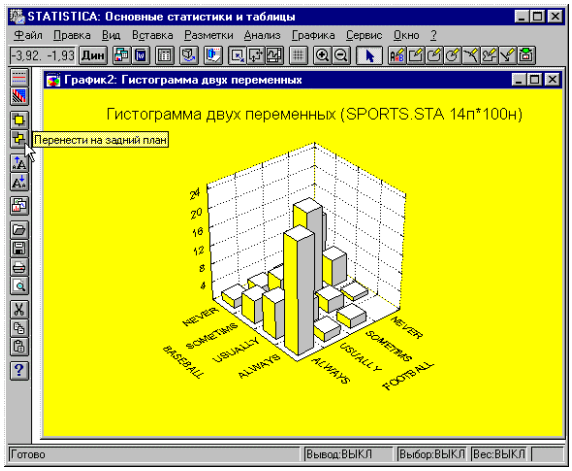
Эта кнопка (последняя на всех панелях инструментов) вызывает *Электронное руководство* системы *STATISTICA*. Система предоставляет исчерпывающую электронную документацию на все функции программы и все контекстно-зависимые опции. Его можно вызвать также с помощью клавиши F1 или кнопки в заголовке любого диалогового окна.

Copyright © StatSoft, 1995

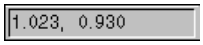
StatSoft

ГРАФИЧЕСКАЯ ПАНЕЛЬ ИНСТРУМЕНТОВ

Панель инструментов графического окна предоставляет возможности для рисования, настройки, а также управления одновременно несколькими графиками, связывания и внедрения графических объектов.



Большинство описанных ниже средств управления графикой с помощью кнопок панели инструментов доступны и из обычных меню (они могут быть записаны в виде клавиатурных макрокоманд). Число кнопок на любой панели инструментов и порядок их расположения определяется пользователем (см. стр. 1053).



0. Поле просмотра

В окне *Поле просмотра* на графической панели инструментов показаны текущие координаты курсора в графическом окне (для таблиц данных и результатов в нем показано число из выделенной ячейки, см. разделы о панелях инструментов таблиц). Это поле выполняет функцию "считывателя экрана", данные в нем автоматически обновляются при движении курсора (мыши) в текущем (активном) графическом окне и соответствуют значениям координат. По умолчанию (в *динамическом* режиме, см. ниже) эти показания соответствуют координатам осей *X* и *Y* текущего графика. Можно переключиться в *фиксированный* режим, нажав кнопку, следующую за окном *Поле просмотра* (см. ниже), при этом здесь будут показаны абсолютные координаты окна (в %), которые не зависят от шкалы графика:*[0, 0]%* соответствует левому нижнему углу окна, а *[100, 100]%* соответствует правому верхнему углу. При повороте текста на экране (если перемещать его, потянув за угол) в окне *Поле просмотра* указывается угол поворота.

Ширину этого поля можно изменить. Для этого надо щелкнуть на нем мышью или выбрать команду *Изменить поле просмотра* выпадающего меню *Вид*.

Дин / Фкс 1. Динамический /Фиксированный режимы (переключатель)

Эта кнопка обеспечивает переключение между *динамическим* и *фиксированным* режимами отображения координат для настройки графических объектов (см. предыдущий абзац).

***Динамический* режим.** В *динамическом* режиме, установленном по умолчанию все пользовательские объекты в любом месте графического окна (рисунки, стрелки, текст, внедренные и связанные графики и т.д.) привязаны к координатам текущих осей *X* и *Y*.

Таким образом, они динамически связаны с конкретным "логическим" местом графика (а не с абсолютным расположением внутри графического окна). При изменении масштаба или перемещении всего графика относительное расположение пользовательских (*динамических*) объектов и соответствующих элементов графика

StatSoft
Copyright © StatSoft, 1995

не изменяется (например, стрелка будет указывать на ту же точку линейного графика после изменения его масштаба, хотя сама эта точка окажется в другом месте окна).

***Фиксированный* режим.** В *фиксированном* режиме все пользовательские объекты сохраняют свое абсолютное положение, которое определяется только пропорциями графического окна.

Например, если внизу добавить сноску, то при любом изменении масштаба самого графика она будет оставаться на прежнем месте.

Индикатор статуса объекта. Эта кнопка используется в качестве индикатора состояния конкретного пользовательского графического объекта. Если выделить объект (например, щелкнуть на текстовой метке), то положение кнопки будет соответствовать статусу (состоянию) данного объекта (*динамическому* или *фиксированному*).

Изменение статуса объектов. Для изменения статуса конкретного объекта достаточно нажать эту кнопку, когда данный объект выделен.

2. Переключатель модулей

Эта кнопка (общая для всех панелей инструментов) вызывает окно *Переключатель модулей* системы *STATISTICA*. Он обеспечивает быстрый доступ ко всем модулям (т.е. пакетам программ), включенным в данную версию системы. В зависимости от текущей конфигурации *переключатель* запускает новый модуль либо в том же окне, либо открывает новое окно приложения.

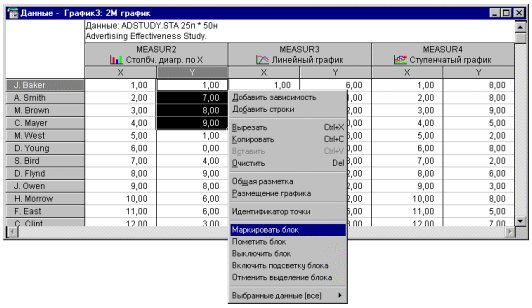
3. Управление Рабочей книгой

Эта кнопка всегда располагается второй на всех панелях инструментов и вызывает диалоговое окно *Заголовок файла данных, примечания и Рабочая книга*. В этом диалоговом окне можно ввести однострочный *заголовок файла* (который появляется в верхней части таблицы данных), подробную информацию или замечания о содержании данных, их источнике и т.д. Здесь можно контролировать все файлы, сопутствующие этому набору данных (например, графики, отчеты, программы). Управление Рабочей книгой файлов более подробно рассматривается на стр. 1055.

4. Редактор данных графика

Данная кнопка вызывает окно *Редактор данных графика* , в котором содержатся все данные, отображенные на текущем графике. В системе *STATISTICA* можно просматривать и редактировать любые значения, представленные на каждом из графиков, независимо от того, исходные ли это данные или часть таблицы результатов, или какие-либо рассчитанные величины (например, плотность вероятности). Все эти значения всегда доступны для редактирования во встроенном *Редакторе данных графика*.

Он организован в виде групп столбцов, представляющих отдельные зависимости (т.е. наборы данных) текущего графика.

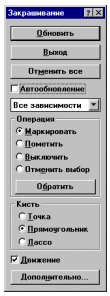


Copyright © StatSoft, 1995
StatSoft

Для смешанных графиков каждая группа столбцов может представлять зависимости различных типов (например, линейные графики и диаграммы рассеяния), при этом соответствующий каждому типу значок располагается около названия столбца. Группа может состоять из одного, двух, трех или четырех столбцов в зависимости от типа графика. Содержание редактора может быть дополнено, объединено с другими данными, записано в файл и т.п. (чтобы получить более подробную информацию, нажмите клавишу F1, находясь в *Редакторе*). При сохранении данных полное содержание *Редактора* записывается также и в графический файл, чтобы потом можно было продолжить интерактивный анализ данных (например, закрашивание, см. ниже). Кроме того, *Редактор* позволяет оперировать с данными, помеченными с помощью закрашивания (см. следующую кнопку), например, их можно выборочно копировать в буфер обмена, переместить на другой график и т. д.

5. Кисть

Эта кнопка включает режим *закрашивания* и вызывает диалоговое окно *Закрашивание* . При этом курсор принимает форму "перекрестья прицела" (можно также использовать кисть в форме прямоугольника или лассо, см. диалоговое окно *Закрашивание*). В этом режиме появляется возможность выделять точки на графике, щелкая на них курсором в виде перекрестья (или заключая их в прямоугольник или лассо).



Система *STATISTICA* предоставляет широкий выбор способов работы с кистью и с точками данных на экране, а также различные варианты действий с закрашенными (выделенными) точками.

Как работает инструмент *Кисть*: различие между выделением данных и обновлением. Выделение данных происходит перед совершением какой-либо *операции*, аналогично тому, как выделение текста предшествует какой-либо операции с ним (например, изменению шрифта). Подобным же образом, "указывая на" (выделяя) конкретные точки на графике, мы "сообщаем программе", с какими данными будем оперировать.

График не изменится, пока не дана команда *Обновить* (т.е. выполнить конкретную *операцию*). Когда нужные точки выделены, нажатие кнопки *Обновить* в диалоговом окне *Закрашивание* (см. ниже) приведет к выполнению действия, определенного в поле *Операция* этого диалогового окна. Существуют различные типы действий: точки могут быть *маркированы* (различными маркерами), *помечены* или *выключены* (т.е. временно удалены с графика, например, для исследования влияния выбросов на аппроксимацию функции).

Еще одна доступная операция *Отменить выбор* позволяет изменить все атрибуты точек данного графика (*маркированная*, *помеченная*, *выключенная* или *подсвеченная*), которые получились в результате предыдущих операций. Это может быть сделано одновременно для всех зависимостей, если в соответствующем поле списка диалогового окна *Закрашивание* выбран режим *Все зависимости*.

Отменить весь выбор. Нажатие этой кнопки приведет к удалению всех атрибутов (*маркированная*, *помеченная*, *выключенная* или *подсвеченная*) всех выбранных точек для всего графика (независимо от выбора текущей

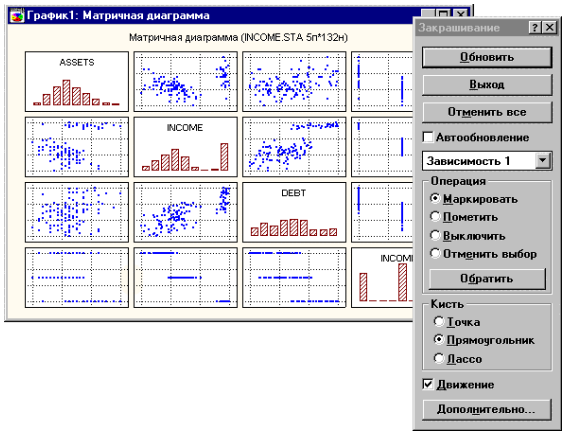
StatSoft®
Copyright © StatSoft, 1995

зависимости в списке). При этом для всех точек будет восстановлен статус по умолчанию.

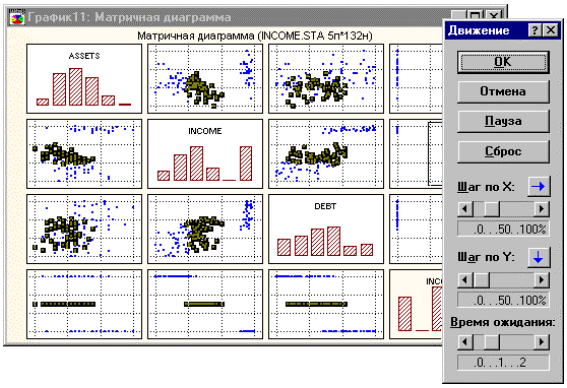
Автоматическое обновление. Для графиков, которые могут быть достаточно быстро перерисованы, полезно использовать режим *Автообновление*. В этом режиме после каждого применения кисти график автоматически перерисовывается (т.е. выполняется заданная *операция*). Поэтому при *автообновлении* точки никогда не выделяются, а сразу маркируются, отмечаются и т. д.

Обратное изменение. При нажатии на кнопку *Обратить* все *операции* меняются на противоположные. Теперь выделенные точки можно *не маркировать*, *не помечать* или *включить*.

Движение. Если выбрать тип кисти *Прямоугольник* или *Лассо*, то можно применить динамическое закрашивание (для этого надо выбрать режим *Движение*).

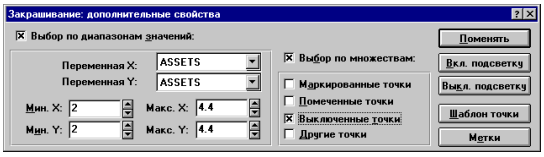


В этом режиме кисть автоматически перемещается по графику, а ее параметрами можно интерактивно управлять из диалогового окна *Движение*.



Такой метод очень удобен при разведочном анализе данных на матричных графиках.

Дополнительно... Кнопка *Дополнительно...* в исходном диалоговом окне *Закрашивание* вызывает диалоговое окно *Закрашивание: дополнительные свойства*. Здесь можно либо определить диапазон координат выбираемых точек, либо использовать их ранее заданные атрибуты. Для них можно *включить подсветку* (или *выключить подсветку* или *поменять* эти атрибуты). Все *дополнительные свойства* применяются одновременно, т.е. точки могут выбираться по диапазону значений и в то же время по атрибутам или комбинации атрибутов.



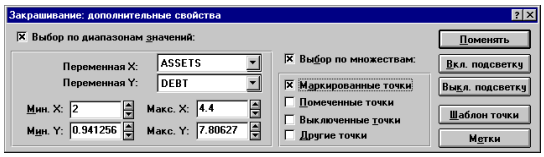
Над точками, выделенными в окне *Дополнительные свойства*, после нажатия кнопки *Обновить* в диалоговом окне *Закрашивание* совершается операция, определенная в этом диалоговом окне в поле *Операция*. Предположим, нужно *включить подсветку* точек, которые предварительно *маркированы* (т.е. их маркеры изменены). Для этого в диалоговом окне *Дополнительные свойства* отметьте поле *Маркированные точки* и нажмите кнопку *Вкл. подсветку*. Теперь можно

Copyright © StatSoft, 1995

StatSoft

временно удалить эти точки с графика (и из процедуры подгонки), если выбрать пункт *Выключить* в диалоговом окне *Закрашивание*.

Обратите внимание, что в режиме *Автообновление* (см. выше) кнопки *Вкл. подсветку* и *Выкл. подсветку* недоступны (они затенены). Это связано с тем, что в данном режиме не происходит выделения данных, а все действия над ними осуществляются сразу после нажатия кнопки *OK*.

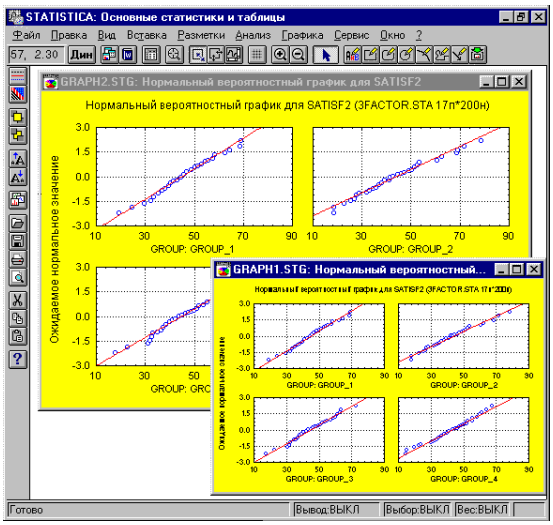


Выход из режима закрашивания. Выйти из режима *закрашивания* можно следующими способами: нажать кнопку *Выход*, или нажать кнопку *Режим выделения* (этот режим установлен по умолчанию, см. кнопку ниже), или нажать клавишу ESC. Заметим, что после выхода из режима *закрашивания подсветка* точек, над которыми не были совершены никакие операции (маркировка, пометка и т.п.), не будет показана на графике до тех пор, пока мы снова не вернемся в этот режим.

Операции над выбранными точками данных (применение). Как правило, закрашивание используется при разведочном анализе данных для интерактивного исследования вкладов отдельных точек в общую картину взаимосвязей переменных (например, на матрицах диаграмм рассеяния) или для идентификации выбросов. Этими выбранными точками часто необходимо оперировать (например, выборочно копировать их в буфер обмена или создать из них новый столбец данных). Такие операции осуществляются в *Редакторе данных графика*.

6. Фиксировать пропорции графика

Эта кнопка включает установленный по умолчанию режим *Фиксировать пропорции* при изменении размеров графического окна.



В этом режиме при изменении размеров не изменяются пропорции графика, т.е. не происходит его искажения.

Сохранение установленного коэффициента пропорциональности. По умолчанию эта кнопка нажата. Она выключается при переходе в другой режим (изменения пропорций, см. следующую кнопку). Нажатие кнопки *Фиксировать пропорции* после изменения пропорций графика не восстановит исходные соотношения сторон, но при этом будут зафиксированы текущие (новые) пропорции. Это позволит в дальнейшем сохранять их при изменении размеров графического окна.

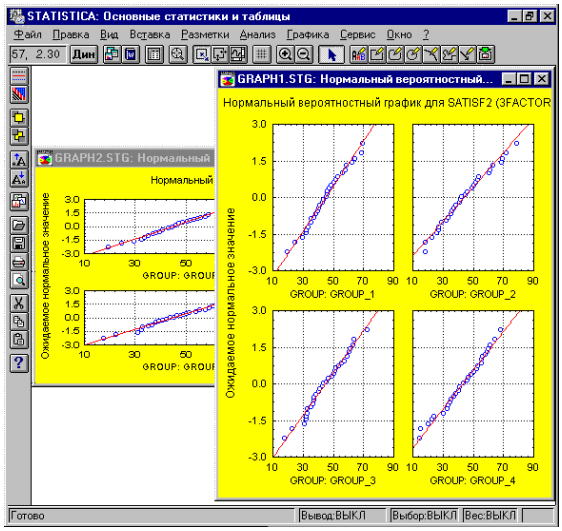
Пропорции, установленные по умолчанию. Устанавливаемые по умолчанию пропорции графического окна можно изменить в диалоговом окне *Отображение графика*, которое вызывается из выпадающего меню *Вид*;

StatSoft®
Copyright © StatSoft, 1995

например, вместо стандартных значений можно установить соотношение сторон графического окна *1:1*.

7. Изменить пропорции графика

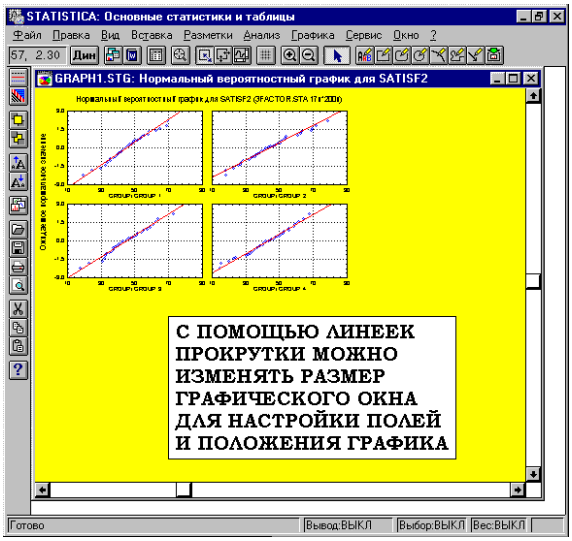
Данная кнопка включает режим свободного (непропорционального) изменения размеров графического окна. При этом можно " сжать " или "растянуть" график в любом направлении, изменить соотношения координат X и Y на двумерном изображении. (Заметим, что этот режим позволяет интерактивно подбирать пропорции, в отличие от настройки в диалоговом окне *Отображение графика*, см. меню *Вид*.)



При этом будут изменены и пропорции графика на печати (это можно увидеть в диалоговом окне *Предварительный просмотр,* нажав кнопку на панели инструментов или выбрав пункт *Предварительный просмотр* в меню *Файл*). После того как пропорции подобраны, их можно зафиксировать, снова включив кнопку *Фиксировать пропорции графика*, см. описание предыдущей кнопки). Вновь установленные пропорции графического окна будут сохранены при изменении его размеров. Процедура изменения установленных по умолчанию пропорций (например, *1:1*), описана в предыдущем параграфе.

8. Подобрать область графика и поля

Данная кнопка включает режим настройки положения графика и его полей . При этом появляются линейки прокрутки, позволяющие менять расположение графика в пределах графического окна.



В отличие от двух предыдущих режимов (см. описание двух предыдущих кнопок), в данном случае сам график не изменяется при изменении размеров графического окна (т.е. его размер и пропорции остаются постоянными), а изменяется (увеличивается или уменьшается) область вокруг графика. Таким способом можно, например, подобрать поля или создать дополнительное пространство с одной стороны для размещения текста, записи, внедренного объекта и т.п.

В этом режиме системой отсчета для всех операций настройки служит окно приложения системы *STATISTICA*. Поэтому относительный

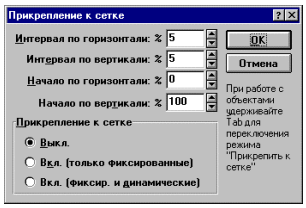
Copyright © StatSoft, 1995
StatSoft

размер графического окна и его расположение в окне приложения будут влиять на диапазон перемещений графика и величину дополнительного добавляемого пространства. Чтобы иметь возможность расширения во всех направлениях, прежде чем войти в этот режим, разместите графическое окно в центре окна приложения и убедитесь, что его размер составляет не более 25% размера этого окна (см. также пункт *Пустое графическое окно* в меню *Графика*).

Все устанавливаемые по умолчанию размеры полей (т.е. положение графика внутри графического окна), могут быть заданы в диалоговом окне *Отображение графика*, которое вызывается из меню *Вид*.

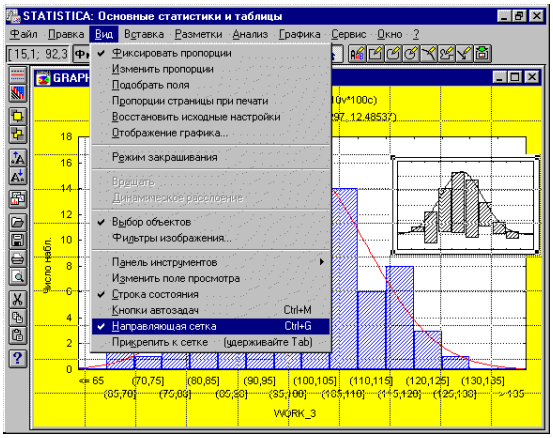
9. Прикрепить к сетке

При нажатии на эту кнопку появляется диалоговое окно для настройки и отображения невидимой сетки, позволяющей выравнивать рисованные объекты (линии, прямоугольники, вставки, текст и др.)



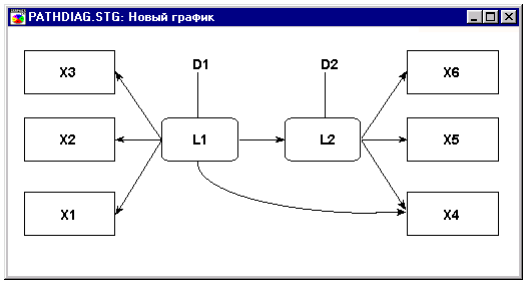
По умолчанию объект выравнивается по ближайшему пересечению линий сетки (как будто разрешение экрана "уменьшилось"). Это и называется прикреплением к сетке.

Изображение направляющей сетки на графике может быть включено и выключено с помощью комбинации клавиш CTRL+G или с помощью команды *Направляющая сетка* в меню *Вид*.



Если режим *Прикрепить с сетке* отключен, то сетка не влияет на рисованные объекты. Можно включить этот режим при размещении конкретного объекта, если нажать клавишу TAB (даже если кнопка *Прикрепить к сетке* выключена, нажатие клавиши TAB включает этот режим). Для отображения сетки выберите пункт *Направляющая сетка* в меню *Вид*. Режим *Прикрепить к сетке* можно разрешить только для объектов в *фиксированном* режиме или для всех объектов (подробнее о двух способах расположения объектов на графике см. страницу 1086).

Использование. Обычно режим *Прикрепить к сетке* используется для выравнивания объектов (например, ряда надписей, независимых элементов условных обозначений или внедренных графиков);

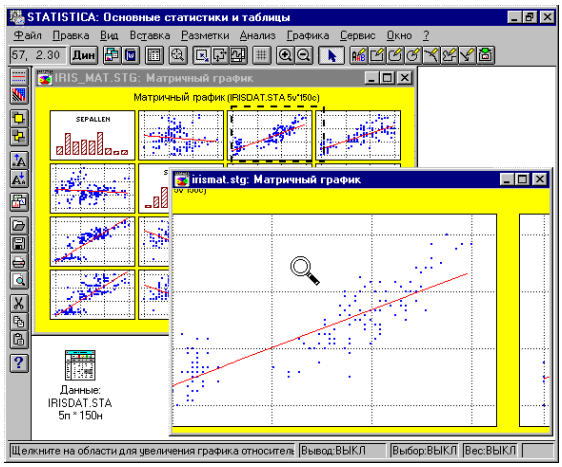


а также для составления диаграмм, списков и т.п.

StatSoft
Copyright © StatSoft, 1995

10. Увеличение

Эта кнопка дает возможность увеличивать выбранную область графического окна. При выборе этого режима (*Увеличение*) курсор принимает вид лупы (), и появляется возможность пропорционального увеличения текущего графика. Для этого надо поместить курсор (лупу) на ту область графика, которую хотелось бы подробно рассмотреть, и щелкнуть один раз левой клавишей мыши.



При этом данная область увеличится, а точка, в которую был помещен курсор, будет центром нового графического окна. Пока вы находитесь в этом режиме (имея курсор в виде лупы), каждый щелчок кнопкой мыши приводит к увеличению соответствующей области графического окна примерно в два раза. Можно сделать максимум 5 щелчков, и получить коэффициент увеличения 32:1.

Отличие логического увеличения от "механического расширения". Заметим, что режим увеличения предоставляет больше возможностей, нежели "механическое растяжение" картинки. При логическом увеличении выбранной области сохраняются размеры маркеров точек, шрифтов и ширина каждой линии. Таким образом, режим *Увеличение* эффективно повышает "функциональное разрешение" дисплея и позволяет исследовать области, которые плохо читаются при обычном разрешении (1:1) из-за наложения меток или маркеров точек. В противоположность этому, "механическое" растяжение приводит к пропорциональному увеличению всех элементов, что не позволяет разглядеть, например, перекрытия точек, которые показаны на приведенной выше картинке.

Пользовательское или установленное по умолчанию отображение графика. По умолчанию в режиме *Увеличение* сохраняются размеры шрифтов, маркеров и расстояния между компонентами графика (для выявления большего количества деталей в отличие от механического растяжения). Однако, в режиме *Уменьшение* для правильного уменьшения всей области графика шрифты, маркеры и расстояния пропорционально сокращаются (см. следующий пункт). В диалоговом окне *Отображение графика*, вызываемого из меню *Вид*, как для процедуры увеличения, так и для уменьшения можно установить один из двух режимов: либо изменения, либо сохранения размеров шрифтов, маркеров и расстояний.

Рисование в режиме увеличения. При логическом увеличении (в режиме *Увеличение*, см. предыдущий параграф) черные квадратики области редактирования и перемасштабирования сохраняют свои размеры независимо от коэффициента увеличения (для примера см. раздел *Пропорциональное и непропорциональное изменение размеров*, стр. 1098). Поэтому этот режим может быть использован для точного редактирования произвольных рисунков.

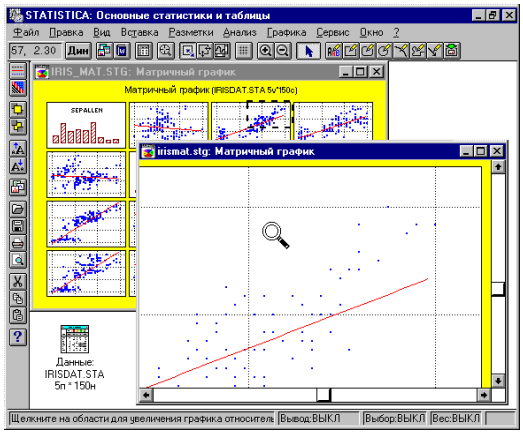
Просмотр графика в режиме увеличения (перемещение "лупы" по графику). В системе *STATISTICA* при увеличение применяется метод *виртуальной области графика* (а не способ "частичной перерисовки"),

Copyright © StatSoft, 1995

StatSoft

поэтому режим *Увеличение* может быть использован одновременно с режимом *настройки области графика и полей* (см. описание предыдущей кнопки).

При включении этого режима вокруг графика появляются линейки прокрутки (см. ниже), что позволяет передвигать увеличенную область в графическом окне. Это эквивалентно перемещению "лупы" вдоль графика.

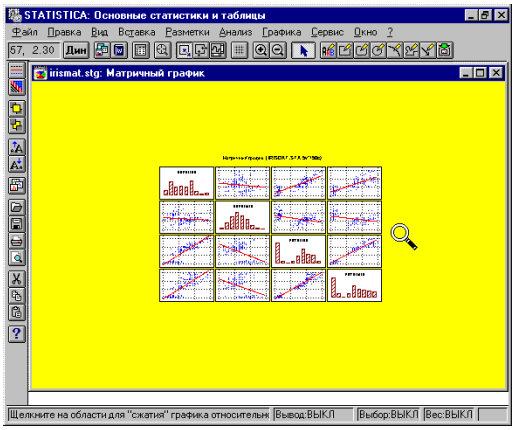


Обращение операции увеличения.
С помощью режима *Уменьшение* можно уменьшить "размер" графика (такими же шагами, какими происходило увеличение) (см. следующую кнопку). Обратите внимание, что если производить операцию уменьшения не в той же центральной точке, где график был перед этим увеличен, то положение графика в окне не будет восстановлено. Для того чтобы вернуться к первоначальному положению графика (а также к установленным по умолчанию полям и коэффициенту увеличения), необходимо использовать команду *Восстановить исходные настройки* из меню *Вид*.

11. Уменьшение

Эта кнопка позволяет включить режим уменьшения размеров выбранной области графика. Режим *Уменьшение* действует аналогично режиму *Увеличение* (см. описание предыдущей кнопки).

Этот режим удобен, когда нужны не детали, а свободное пространство вокруг графика для вставки текста или картинки или для подготовки составного документа. (См. пункт *Пустое графическое окно* в меню *Графика*).



Пользовательское или установленное по умолчанию отображение графика.
Обратите внимание, что в отличие от режима *Увеличение* (см. предыдущий пункт) в режиме *Уменьшение* все размеры шрифтов, маркеров и интервалы на графике уменьшаются пропорционально с целью правильного сокращения графика в целом. В диалоговом окне *Отображение графика*, вызываемого из меню *Вид*, можно задать либо перемасштабирование, либо сохранение размеров шрифтов, маркеров и интервалов как для режима *Увеличение*, так и для режима *Уменьшение*.

12. Режим выделения (указатель)

Данная кнопка включает режим указателя, который установлен по умолчанию для всех графических окон. В этом режиме с помощью левой кнопки мыши можно выделить для настройки, редактирования или интерактивного

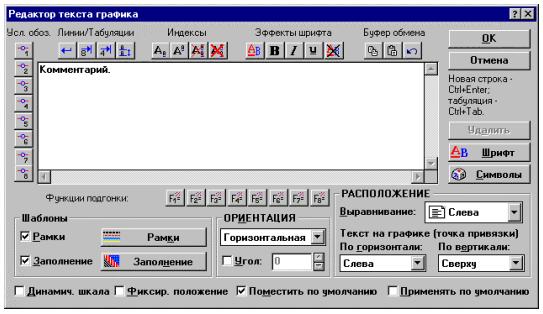
StatSoft
Copyright © StatSoft, 1995

анализа (например, изображения слоев с помощью "просвечивания" или последовательных трехмерных графиков) графические объекты (например, стрелки, внедренные или связанные объекты и т.п.) или различные части графика (например, заголовки, шкалы, линии сетки и т.п.). Режим выделения на время отключается при выборе любого из интерактивных режимов настройки графика (например, закрашивания, добавления текста, рисования стрелок, внедрения графических объектов). Нажатие на кнопку *Режим выделения (указатель)* возвращает систему *STATISTICA* к этому режиму, установленному по умолчанию.

Клавиша ESC. Вернуться в устанавливаемый по умолчанию режим из режима рисования, закрашивания или вставки можно также с помощью клавиши ESC. Обратите внимание, что краткая справка о текущем режиме (или операции с графическими объектами) всегда высвечивается в области сообщений строки состояния системы *STATISTICA*.

13. Редактор текста графика

Эта кнопка открывает *Редактор текста графика*. В этом редакторе можно вводить, изменять или форматировать пользовательский текст (например, заметки, комментарии или большие объемы текста) на графиках и рисунках системы *STATISTICA*.



Введенный в редактор текст может быть изменен или отформатирован, а затем помещен на график после нажатия кнопки *OK*. При этом окно редактора закроется, а курсор примет вид "креста со стрелками", переместите его в нужное место на графике и нажмите левую клавишу мыши для размещения текста. После этого текст, как и любой другой графический объект, может быть обычным образом выбран и перемещен (с помощью мыши). Двойной щелчок на уже существующем текстовом объекте вызовет *Редактор текста графика*, и все атрибуты этого текста можно будет отредактировать. В зависимости от положения переключателя *Динамический/Фиксированный* (см. стр. 1086), новый текст будет располагаться на графике динамически (т.е. его положение будет подстраиваться под все последующие изменения графика) или занимать фиксированное положение в графическом окне (например, всегда в нижнем левом углу на расстоянии от левой границы, равном 10% от ширины окна). Этот статус изменяется либо с помощью опции редактора *Динамич. шкала*, либо установкой по умолчанию в диалоговом окне *Стили текстовых объектов* (вызываемом из меню *Вставка*). Это можно сделать также, выделив текст и нажав кнопку Дин/Фкс панели инструментов (см. стр. 1086).

Форматирование, управляющие символы, встроенная панель инструментов. *Редактор текста графика* дает возможность различными способами форматировать, размещать и выравнивать текст, включать в него индексы и степени различных уровней, символы условных обозначений (соответствующие значкам для последовательных зависимостей на графике) и многое другое. Самый простой способ ввести какие-либо атрибуты (например, курсив или степени) - это выделить нужную часть текста и нажать соответствующую кнопку на встроенной

Copyright © StatSoft, 1995
StatSoft

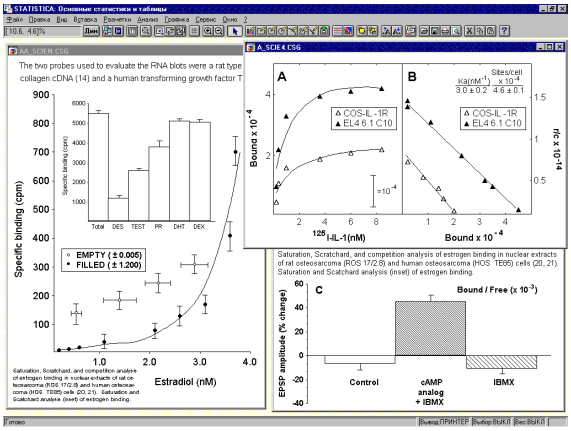
панели инструментов (ниже кратко описана каждая из ее кнопок).

При этом система *STATISTICA* вставит в текст нужную последовательность управляющих символов. Эту последовательность можно вставить в текст и непосредственно. Например, символ **@B** обозначает начало и конец жирного шрифта (поэтому **@B**TEXT**@B** будет напечатан как **TEXT**).

Управляющие символы. Во всех полях редактирования системы *STATISTICA* для помещения на график текста (заголовков, меток шкалы и данных, названий групп и значений, условных обозначений и т.п.) применяются одни и те же управляющие символы. Поэтому любой текст может быть соответствующим образом настроен. Кроме того, можно скопировать текст из *Редактора текста графика* и поместить его в любой из текстовых объектов. Более подробное описание этих возможностей и примеры можно найти в *Электронном руководстве*, которое вызывается кнопкой ? или клавишей F1 из окна *Редактор текста графика*.

Расширенные возможности форматирования текста, формулы. Кроме наиболее часто используемых средств форматирования, доступных со встроенной панели инструментов, в *Редакторе текста графика* можно дополнительно использовать управляющие символы (например, для ввода и редактирования формул). Можно, к примеру, управлять небольшими изменениями позиций табуляции и межстрочного интервала, а также вставлять в текст условные обозначения отдельных значений переменных на графиках поверхностей и линий уровня. Кроме того, можно автоматически ввести в условные обозначения значения, соответствующие конкретным скрытым областям на графике. Более подробное описание этих возможностей и примеры можно в *Электронном руководстве*, которое вызывается кнопкой ? или клавишей F1 в окне *Редактор текста графика*.

Комбинирование текста и графиков. На графиках системы *STATISTICA* может быть расположено практически неограниченное количество текста, при этом можно создавать документы, составленные из текста, картинок и формул.

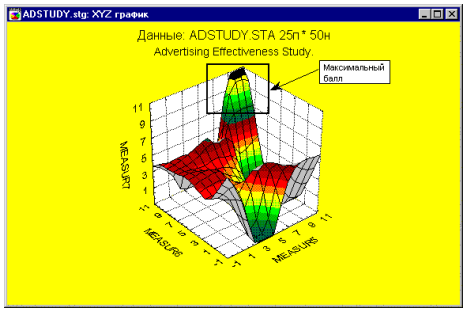


Как уже упоминалось выше, с помощью *Редактора текста графика* можно настроить любой пользовательский текст. Для этого надо либо дважды щелкнуть на нем левой клавишей мыши, либо щелкнуть правой клавишей, а затем выбрать соответствующую команду из появившегося контекстного меню. Это относится и к перемещаемым условным обозначениям, которые имеют статус пользовательского текста.

StatSoft®
Copyright © StatSoft, 1995

14. Вставка прямоугольника

С помощью этой кнопки включается *режим рисования прямоугольников* . При этом курсор принимает на графике вид креста. Перемещая его при нажатой клавише мыши, можно рисовать прямоугольники.



Рисование квадратов. Изобразить квадрат можно, если в процессе рисования держать нажатой клавишу SHIFT (при этом ширина и длина объекта будут одинаковы).

Выход. Для выхода из режима рисования нажмите кнопку *Режим выделения* (см. описание кнопки на стр. 1094) или клавишу ESC.

Шаблоны. Шаблоны линий и заполнения зависят от текущих установок по умолчанию, сделанных в диалоговом окне *Стиль фигур*, которое вызывается из меню *Вставка*). Для их изменения нужно щелкнуть на соответствующем прямоугольнике правой кнопкой мыши и выбрать нужную опцию настройки из появившегося контекстного меню. Можно сделать это и другим способом: выбрать прямоугольник, щелкнув на нем (вокруг него появятся масштабирующие квадратики), а затем подобрать его вид с помощью кнопок настройки шаблонов линий и заполнения (см. описание этих кнопок на стр. 1102 и 1102 соответственно).

Динамический или фиксированный режим. В зависимости от того, установлен ли режим *Динамическая шкала* в диалоговом окне *Стили фигур*, каждый новый прямоугольник будет располагаться на графике либо динамически (т.е. его положение будет зависеть от изменения графика), либо в фиксированном месте (например, в нижнем левом углу на расстоянии от левой границы, равном 10% от ширины окна).

15. Вставка скругленного прямоугольника

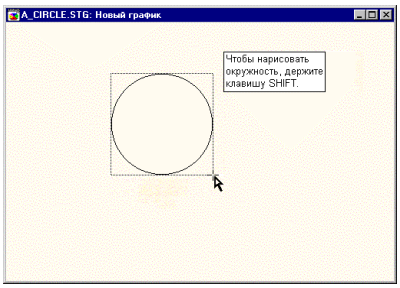
Эта кнопка включает *режим рисования скругленных прямоугольников* (см. описанные выше способы рисования и настройки прямоугольников).

Рисование скругленных квадратов. Если во время рисования в этом режиме держать нажатой клавишу SHIFT, то будет нарисован квадрат со скругленными углами (будет сохраняться одинаковая ширина и длина сторон).

16. Вставка эллипса

Данная кнопка включает *режим рисования эллипсов* (см. описанные выше способы рисования и настройки прямоугольников).

Рисование окружностей. Если во время рисования держать нажатой клавишу SHIFT, то будет нарисована окружность.

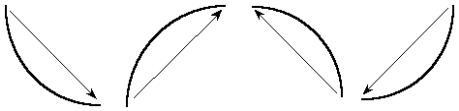


17. Вставка дуги

Эта кнопка включает *режим рисования дуг* (см. описанные выше способы рисования и настройки прямоугольников). При этом курсор принимает

Copyright © StatSoft, 1995
StatSoft

вид креста, и можно нарисовать дугу, перемещая его по графику при нажатой клавише мыши. Нарисованная дуга представляет собой четверть эллипса, ею можно управлять, как показано ниже.

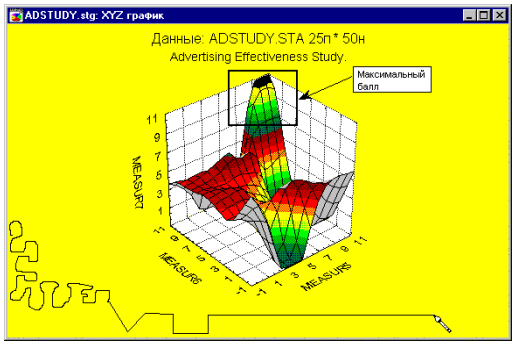


Пропорции эллипса зависят от угла, под которым перемещается курсор мыши (при угле 45° будет построена четверть окружности; другим способом этого можно добиться, если во время рисования держать нажатой клавишу SHIFT).

Ориентация дуги (вогнутость или выпуклость) зависит от направления движения курсора (вверх или вниз соответственно, как показано выше на рисунке).

18. Вставка ломаной линии, произвольное рисование

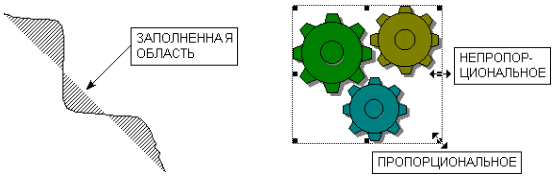
При нажатии этой кнопки включается *режим произвольного рисования, рисования ломаных линий или многоугольников* (см. описанные выше способы рисования и настройки прямоугольников). При нажатии на эту кнопку курсор примет вид тонкой "кисти" (см. ниже).



Произвольные линии. Для рисования произвольной линии нужно двигать кисть при нажатой левой клавише мыши.

Прямые линии. Чтобы нарисовать отрезок прямой, поместите кончик кисти в начальную точку отрезка и щелкните кнопкой мыши, затем, отпустив кнопку, переместите кисть в конечную точку отрезка и щелкните еще раз. Чтобы закончить рисование, сделайте двойной щелчок кнопкой мыши.

Закрашенные фигуры. Рисовать закрашенные фигуры можно тогда, когда заданы шаблоны заполнения (см. описание настройки шаблонов заполнения и кнопки на стр. 1102). Обратите внимание, что для заполнения ни кривая, ни ломаная линия не должны быть замкнутыми, достаточно, чтобы они имели хотя бы одну выпуклость, как показано ниже (т.е. не были прямыми).

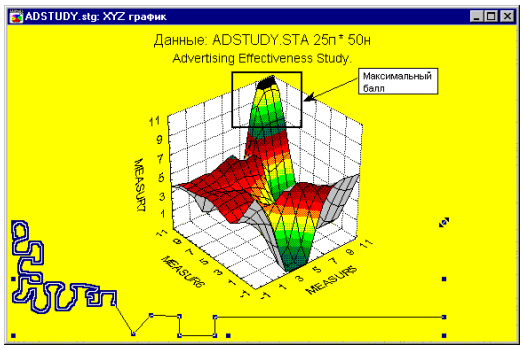


Пропорциональное и непропорциональное изменение размеров. Помимо обычных способов настройки графических объектов, существуют дополнительные методы рисования линий (см. приведенное выше описание рисования и настройки прямоугольников и кнопки). При выделении объекта он окружается восемью масштабирующими квадратиками вместо четырех. Если ухватить и потянуть за угловой квадратик, то рисунок изменится пропорционально, а если за один из средних квадратиков - то изменение произойдет в одном из направлений, горизонтальном или вертикальном (т.е. расширится или сожмется).

Вид курсора меняется при перемещении по квадратикам, показывая, какому режиму изменения размеров они соответствуют. К примеру, на графике, приведенном на

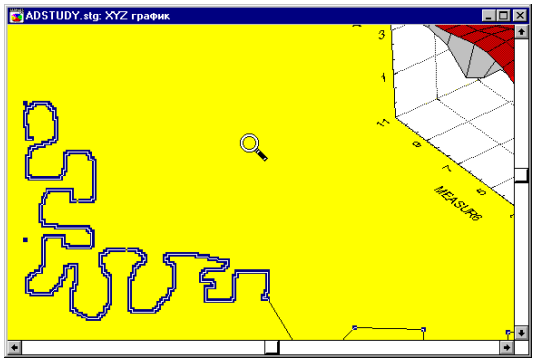
StatSoft®
Copyright © StatSoft, 1995

следующей странице, курсор расположен в правом верхнем квадратике. Это диагональный квадратик и форма курсора в виде двух стрелочек говорят о том, что, потянув за него, мы изменим объект пропорционально.



Редактирование формы линии. В режиме рисования линий возможно редактирование элементов *формы* линии. В каждой точке, где линия меняет свое направление, появляется масштабирующий квадратик (см. иллюстрацию выше). Если потянуть за него курсором мыши, то можно изменить форму линии в данном конкретном месте.

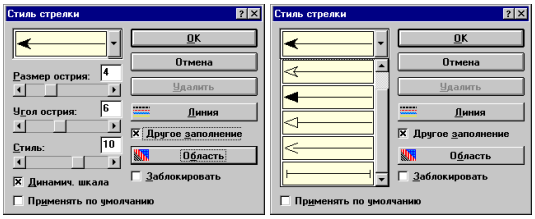
Редактирование в режиме увеличения. Как правило, произвольные рисунки плотно покрыты квадратиками, поскольку состоят из достаточно коротких фрагментов. Для редактирования таких кривых нужно увеличить рисунок (см. кнопку *Увеличение* на стр. 1093).



Рисунки могут состоять из большого числа независимых или перекрывающихся элементов, при этом последовательность, в которой они перерисовываются (и, следовательно, перерисовывают друг друга) определяется с помощью функции *Управление порядком перерисовывания* (см. соответствующие кнопки *Вынести на передний план* и *Поместить на задний план* на стр. 1102). Рисование, как и все другие способы настройки графика (например, внедрение объектов), можно осуществлять в любом режиме изменения размеров, включая режим увеличения (при котором можно добиться большей точности рисования, как показано выше на рисунке).

19. Вставка стрелок или отрезков

Эта кнопка включает *режим рисования стрелок и обозначений ошибок* . При ее нажатии появляется диалоговое окно для настройки стрелок. Здесь из множества заранее определенных стилей можно выбрать один нужный. Кроме того, этот метод настройки позволяет создавать неограниченное количество пользовательских типов стрелок и обозначений ошибок.



После выбора нужного стиля нажмите *OK*, поместите курсор в начальную точку (острие) стрелки, нажмите кнопку мыши и перетащите курсор в то место, где будет расположен конец (хвост) стрелки.

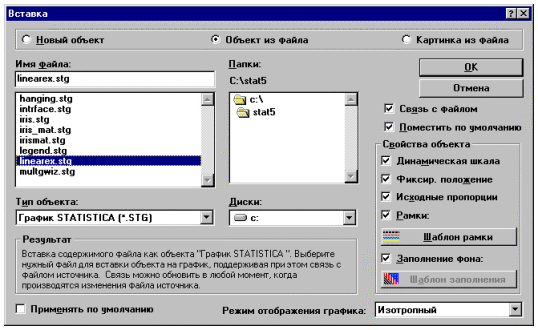
Настройка положения стрелки. Стиль, точный размер и положение стрелки (или обозначения ошибки) может быть подобран

Copyright © StatSoft, 1995
StatSoft

позднее. Щелкните на ней правой кнопкой мыши и выберите способ настройки из появившегося меню. Чтобы изменить положение стрелки, выберите ее с помощью двойного щелчка мышью, при этом на концах ее появятся масштабирующие квадратики. Если потянуть за любой из них, то можно изменить длину и/или угол поворота стрелки. Для перемещения всей стрелки расположите курсор мыши непосредственно на ней (курсор примет вид четырех стрелочек) и перетаскивайте объект.

20. Вставка объектов (средства *OLE*)

Данная кнопка включает *режим связывания или внедрения объекта* и вызывает диалоговое окно *Вставка*, которое можно вызвать и из выпадающего меню *Вставка*. Этот режим дает возможность объединять на одном изображении различные объекты (текст, таблицы, метафайлы, растровые изображения, графики системы *STATISTICA* и т.п.).



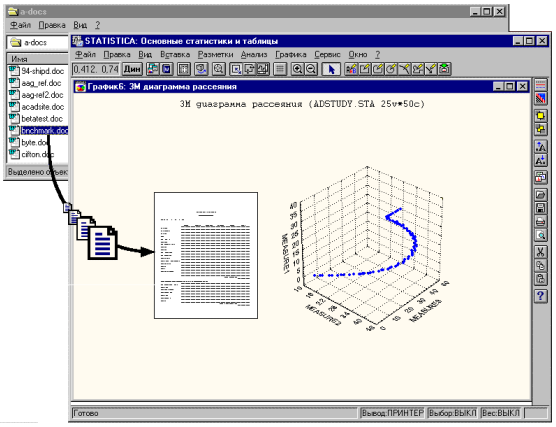
Типы объектов. Если задан режим *Новый объект*, то тип создаваемого объекта может быть выбран из списка приложений Windows, которые поддерживают средства *OLE*. После выбора типа и нажатия кнопки *OK* будет открыто окно соответствующего приложения для создания нового объекта.

Если задан режим *Объект из файла*, то тип объекта для вставки также выбирается из списка приложений Windows, поддерживающих средства *OLE*; после выбора типа будут показаны все предварительно сохраненные файлы этого приложения. В режиме *Картинка из файла* можно вставить объект, несовместимый с методом *OLE*, но записанный в одном из графических форматов Windows: в формате метафайла (файл с расширением **.wmf*) или растрового изображения (файл с расширением **.bmp*). Для получения более подробной информации нажмите кнопку ? или клавишу F1, которые вызывают *Электронное руководство*.

Связывание и внедрение (обзор). Система *STATISTICA* поддерживает средства *OLE* (Связывания и внедрения объектов) как в режиме клиента, так и в режиме сервера. Это означает, что возможна не только динамическая настройка графиков системы *STATISTICA* в других приложениях (режим сервера), но также внедрение и последующее преобразование *OLE*-совместимых объектов других приложений (например, графиков или таблиц) или собственных объектов в графики системы *STATISTICA*. Другими словами, помимо присоединения внешних элементов к графикам системы *STATISTICA* с помощью вставки (см. описание кнопки на стр. 1106), можно обращаться непосредственно к объектам, содержащимся в файле на диске (например, перетащить их непосредственно из окна *Диспетчер файлов* или *Проводник (Windows*

StatSoft®
Copyright © StatSoft, 1995

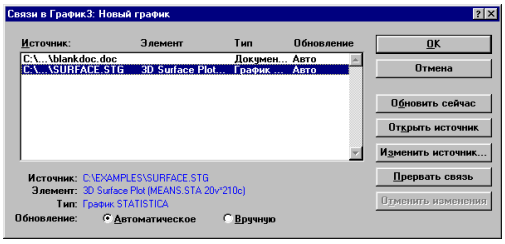
Explorer) и поместить на график системы *STATISTICA*).



Система *STATISTICA* поддерживает как связанные (т.е. динамически присоединенные), так и внедренные (т.е. статически "встроенные") объекты. При этом они могут быть расположены в любом файле, созданном приложениями Windows, включая файлы в собственном графическом формате системы *STATISTICA* (с расширением **.stg*). Более того, система *STATISTICA* одновременно может являться как клиентом, так и сервером в методе *OLE*, поддерживая при этом уникальную возможность создания *вложенных составных документов* (до четвертого порядка включительно), т.е. документ системы *STATISTICA* с внедренным документом может быть в свою очередь внедрен в другой документ этой системы.

Заметим, что каждый и этих двух способов присоединения (*связывание* и *внедрение*) имеет свои преимущества и недостатки.

Связанные объекты. Графики со связанными объектами медленнее перерисовываются, поскольку при этом могут быть задействованы связи с внешними файлами. В то же время эти графики обновляются автоматически (статус связей может быть установлен в диалоговом окне *Связи...*, которое вызывается из графического меню *Правка*).



Это позволяет легко создавать составные документы, которые включают именно "текущее" содержимое других файлов.

Внедренные объекты. Графики с внедренными объектами перерисовываются быстрее, чем со связанными объектами, поскольку здесь отсутствуют связи с обновляемыми внешними файлами. Если дважды щелкнуть на внедренном объекте, то будет вызвано приложение-сервер (т.е. источник), в котором можно изменить данный объект. При этом обновить внедренный объект можно двумя способами: отредактировать его или заменить вручную.

В меню *Правка* можно настроить все параметры внешних объектов (*связанных* или *внедренных*), а также их связи с другими компонентами графика. Кроме того, можно щелкнуть на объекте правой кнопкой мыши и выбрать нужные опции настройки из контекстного меню. Единственным исключением является способ присоединения объекта (*связывание* или *внедрение*), который определяется в момент подключения файла (после этого только связанный объект можно преобразовать во внедренный, но не наоборот, см. команду *Преобразовать во внедренный* из меню *Правка*).

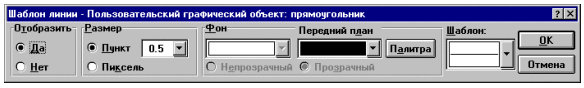
Автоматическое изменение масштаба и настройка встроенных графиков системы *STATISTICA*. При вставке графика системы *STATISTICA* его масштаб автоматически изменяется таким образом, чтобы все размеры,

Copyright © StatSoft, 1995
StatSoft

шрифты, метки и интервалы наилучшим образом соответствовали размеру объекта по отношению к размеру окна, в которое он помещается. Этот параметр может быть задан в поле *Режим отображения графика* в диалоговом окне *Вставка* (подробнее см. *Электронное руководство*).

21. Шаблон линии

Эта кнопка позволяет выбрать шаблон линии (стиль, цвет, размер, единицы измерения и режим) для конкретного графического объекта. Нажатие на нее приводит к результату только в том случае, когда выбран какой-либо объект. [Для подбора шаблонов новой последовательности объектов используйте диалоговые окна *Стили объектов*, которые вызываются из меню *Вставка*].

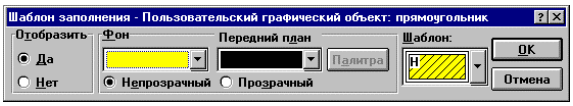


Показанное выше диалоговое окно для выбора стиля линии используется не только для пользовательских графических объектов, но и для настройки многих других линейных элементов на графиках (например, оно появляется, если дважды щелкнуть на любой линии или кривой какого-либо графика).

Специальные средства настройки линий. В этом диалоговом окне помимо обычных параметров настройки предлагаются некоторые уникальные средства, например, двухцветный шаблон линии (выберите любой шаблон разрывной линии, установите режим "непрозрачный" и выберите нужный "цвет фона", который заполнит пробелы в шаблоне линии). Толщину линии можно задавать в различных единицах: в пикселях (при этом задается число точек, размер которых определяется конкретным устройством - экраном, принтером или записывающим устройством) или в пунктах (1 пункт = $^1/_{72}$ дюйма) с шагом в четверть пункта (т.е. $^1/_{288}$ дюйма). Чтобы добиться желаемого результата при печати и избежать появления "слишком тонких" линий, шириной в один пиксель, которые по умолчанию рисует лазерный принтер, необходимо в диалоговом окне *Параметры печати* ввести общую минимальную толщину линии (см. ниже описание кнопки).

22. Шаблон заполнения

Данная кнопка дает возможность выбрать тип шаблона заполнения (его структуру, цвет и режим) для конкретного графического объекта. Нажатие на нее не дает результата, если не выбран какой-либо объект. [Для выбора шаблонов новой последовательности объектов используйте диалоговые окна *Стили объектов*, которые вызываются из меню *Вставка*].



Показанное выше диалоговое окно настройки шаблона заполнения используется не только для пользовательских, но и для многих других графических объектов системы *STATISTICA*, в которых содержатся закрашиваемые области.

Специальные средства заполнения. Кроме обычных способов настройки, здесь имеются и уникальные средства, такие как "прозрачная штриховка" или двухцветный шаблон заполнения (например, установите не сплошной и не прозрачный шаблон заполнения и выберите нужный "цвет фона", который заполнит промежутки).

23. Вынести на передний план

С помощью данной кнопки выбранный графический объект можно поместить на передний план. Операция *Вынести на передний план* изменяет порядок перерисовывания и позволяет расположить нужный графический

StatSoft®
Copyright © StatSoft, 1995

объект "поверх" всех остальных. Объект, находящийся на переднем плане, не загораживается другими объектами, поскольку перерисовывается в последнюю очередь (см. описание следующей кнопки *Поместить на задний план*).

Например, если необходимо поместить текст поверх нарисованного позже (или внедренного) чертежа, то нужно выделить этот текстовый объект и нажать кнопку *Вынести на передний план* (или можно выделить фоновый объект и нажать кнопку *Поместить на задний план*, см. ниже).

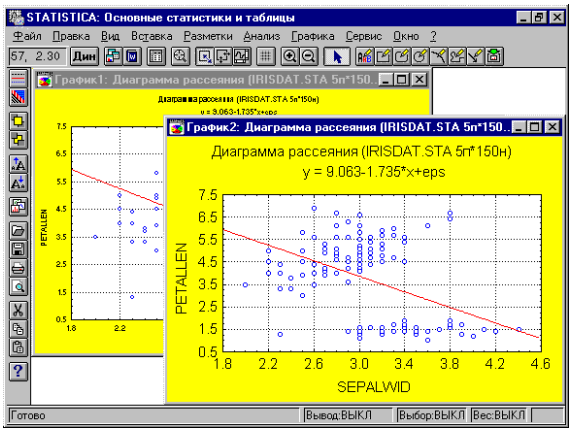
24. Поместить на задний план

С помощью этой кнопки можно поместить выбранный пользовательский графический объект на задний план. Это означает, что данный объект будет теперь перерисовываться в первую очередь и, следовательно, будет закрыт другими пользовательскими объектами.

Операция *Поместить на задний план* противоположна по действию операции *Вынести на передний план* (см. описание предыдущей кнопки).

25. Увеличить все шрифты на 16% (при нажатой клавише CTRL - на 4%)

Эта кнопка позволяет пропорционально увеличить размеры всех шрифтов на графике (а также размеры маркеров, все расстояния и интервалы, заданные в пунктах, или связанные с размером шрифтов или маркеров). Однократное нажатие кнопки увеличивает все эти размеры на 16%. Если одновременно нажать клавишу CTRL , то увеличение будет меньше и составит 4%.



Copyright © StatSoft, 1995
StatSoft

Метод *Глобальное увеличение шрифтов* используется при создании маленьких графиков (например, для публикаций), когда требуется вводить нестандартные размеры шрифтов по отношению к размерам самого графика.

На приведенном выше рисунке для графика на переднем плане использовано *глобальное увеличение шрифтов* (на графике, расположенном сзади, использован шрифт по умолчанию).

Техническое замечание: логические и физические размеры шрифтов (отображение размеров шрифта в различных по размерам областях графика). Данный параметр определяет величину *базы отображения* (его можно также задать в диалоговом окне *Отображение графика*, вызываемом из меню *Вид*, где можно изменять установки по умолчанию). Система *STATISTICA* поддерживает возможность непрерывного изменения масштаба графика вместе со всеми шрифтами при выводе на экран или на печать. В этом случае физический размер шрифтов (в пунктах, 1 пункт = $^1/_{72}$ дюйма) сохраняется при выполнении следующих условий: не изменена *база отображения* (с помощью описываемой в этом параграфе кнопки), использована бумага обычного размера (A4) в книжной ориентации с установленными по умолчанию полями в 1 дюйм и не менялись пропорции графика.

Быстрый способ изменения размера шрифта с помощью кнопки позволяет увеличивать или уменьшать все шрифты (как на экране, так и на принтере), не меняя их настройки (т.е. не меняя установленный размер в 12 пунктов на размер в 14 пунктов). Это достигается в данном случае другим способом - с помощью глобального изменения свойств отображения логического размера шрифтов в область графика на экране или на печати (т.е. изменения *базы отображения*). При нажатии на эту кнопку размер текстового объекта в пунктах останется неизменным (например, 10), но теперь этой величине (10 пунктов) на экране или на печати будут соответствовать буквы примерно на 16% большие, чем прежде, что определяется новым значением *базы отображения*.

Внедренные или связанные объекты. Если при вставке (внедрении или связывании) графика был использован режим *Фиксир. размер шрифта*, это означает, что была осуществлена его привязка к величине *базы отображения* основного документа (см. диалоговое окно *Свойства объекта* для дочернего графика). Шрифты такого встроенного графика тоже будут увеличены при нажатии на кнопку *Увеличить все шрифты на 16% (*CTRL – *на 4%)*

Установки по умолчанию. Установленное по умолчанию значение *базы отображения* может быть изменено в диалоговом окне *Отображение графика*, вызываемом из меню *Вид*.

26. Уменьшить все шрифты на 16% (при нажатой клавише CTRL – на 4%)

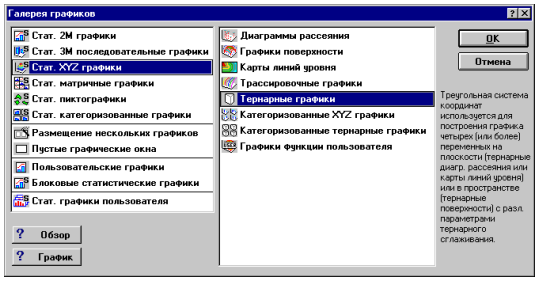
При помощи этой кнопки можно пропорционально уменьшить на графике размеры всех шрифтов (метод *глобального уменьшения шрифтов*). Этот метод является обратным по отношению к методу *глобального увеличения шрифтов* (см. описание предыдущей кнопки).

27. Галерея графиков системы *STATISTICA*

С помощью данной кнопки открывается диалоговое окно *Галерея графиков* системы *STATISTICA*. Она также присутствует в каждом диалоговом окне задания параметров графика.

StatSoft®

Copyright © StatSoft, 1995



Из этого списка можно быстро и легко вызвать любые *статистические и пользовательские графики*, *пустое графическое окно*, а также *статистические графики пользователя*.

28. Открытие графического файла

Эта кнопка позволяет открыть графический файл системы *STATISTICA* (по умолчанию имеющий расширение **.stg*). Новый график открывается в отдельном окне. Отметим, что графические файлы системы *STATISTICA* могут быть открыты непосредственно из окна *Диспетчер файлов*, если дважды щелкнуть на соответствующем имени.

Для присоединения графика или чертежа в другом формате используйте меню *Вставка*. Это средство обеспечивает доступ к графическим метафайлам Windows (с расширением **.wmf*), к растровым графическим файлам (с расширением **.bmp*), а также к графикам, записанным в различных других форматах (см. выше описание кнопки *Вставка объектов*).

Обратите внимание, что в любом диалоговом окне *Открытие/Сохранение* системы *STATISTICA* легко вызвать список файлов, которые используются текущим набором данных (см. информацию о Рабочей книге в томе I). При этом любой файл нужного типа может быть легко выбран из этого списка, если дважды щелкнуть на его имени.

29. Сохранение графического файла

При нажатии на эту кнопку в графическом файле системы *STATISTICA* (имеющем по умолчанию расширение **.stg*) сохраняется текущее графическое окно. В этом файле будет записан график со всеми его настройками и параметрами, включая текущий, отображенный на нем набор данных (см. выше описание кнопки *Редактор данных графика*). При перезаписи графика программа предоставит вам возможность сохранить его в файле с другим расширением (**.bak* вместо **.stg*).

Если вы хотите перезаписать файл, то его имя можно выбрать из списка Рабочей книги. В предыдущем пункте описано, как получить этот список в диалоговом окне *Открытие/Сохранение*.

30. Печать графика

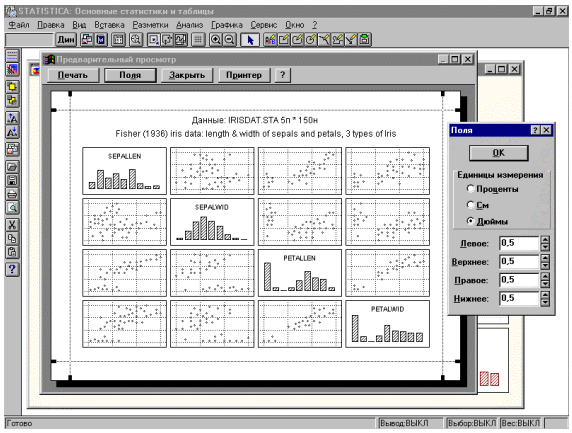
Данная кнопка обеспечивает вывод графика на печать. Это быстрый способ направить график в заранее выбранный канал вывода (*Принтер* и/или *Окно текста/вывода*) с параметрами, установленными по умолчанию или заданными в диалоговом окне *Параметры страницы/вывода*. Изменить их можно в этом же диалоговом окне или с помощью команды *Печать* из меню *Файл*. При нажатии кнопки *Печать* вывод графика начинается немедленно. В то время как при обращении к соответствующему пункту меню сначала вызывается диалоговое окно *Параметры печати*, в котором можно сделать различные настройки. Чтобы подобрать размер полей и увидеть, как график будет выглядеть на печати, используйте режим *Предварительный просмотр* (нажмите кнопку на панели инструментов или используйте команды меню *Файл*). С помощью команды *Печать файлов* из меню *Файл* можно задать печать в командном режиме. Кроме того, можно потребовать автоматическую печать всех графиков и/или

Copyright © StatSoft, 1995
StatSoft

помещение их в файл отчета в *Окне текста/вывода* (см. стр. 1111).

31. Предварительный просмотр

Эта кнопка включает режим предварительного просмотра печати. Здесь можно увидеть, как будет выглядеть график на печатной странице, а также подобрать его расположение и поля.



Можно изменять размеры окна *Предварительный просмотр*, чтобы рассмотреть мелкие детали данного графика. Для настройки полей нажмите кнопку *Поля*. Теперь можно перетащить линии полей на нужное место, либо задать их размер (в дюймах, сантиметрах или процентах).

32. Вырезать графический объект

При нажатии на эту кнопку вырезается (т.е. удаляется) и копируется в буфер обмена выбранный (выделенный в данный момент) графический объект (текст, метка, рисунок, внедренный/связанный график и т.п.). То же самое происходит и при нажатии клавиш CTRL+X. Для совместимости с другими приложениями Windows в буфер помимо объекта в собственном формате системы *STATISTICA* копируются также его растровое, текстовое представления и метафайл.

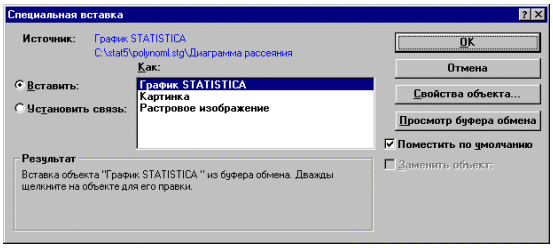
33. Копировать график, графический объект или текст

С помощью данной кнопки можно скопировать выбранный объект или весь график системы *STATISTICA* в буфер обмена (то же самое происходит при нажатии CTRL+C). Если выбран (выделен) какой-либо графический объект (текст, метка, рисунок, внедренный график или чертеж), то только он будет скопирован в буфер. При этом поддерживаются те же форматы записи, что и в операции *Вырезать* (см. описание предыдущей кнопки).

34. Вставить график, графический объект или текст

С помощью данной кнопки в верхний левый угол графика помещается содержимое буфера обмена в формате по умолчанию (см. ниже *Техническое замечание*) (то же самое происходит при нажатии CTRL+V). Потом положение и размер объекта можно нужным образом изменить.

Техническое замечание (Специальная вставка): методы помещения объектов на графики системы *STATISTICA*. Режим *Специальная вставка* из меню *Правка* позволяет выбрать формат, обрамление и положение для объекта, помещаемого на график системы *STATISTICA* из буфера обмена, а также дает возможность установить его связь с исходным файлом.



StatSoft®
Copyright © StatSoft, 1995

Режим *Установить связь* действует только в том случае, когда запущено исходное приложение и сохранен файл, из которого копируется объект.

В этом диалоговом окне можно просмотреть или изменить следующие параметры:

Источник: Показывает источник объекта из буфера обмена.

Вставить: Выбор режима *Вставить* с форматом по умолчанию эквивалентен команде *Вставить* из меню *Правка* (или нажатию кнопки или комбинации клавиш CTRL+V). Если формат вставляемого объекта отличается от заданного по умолчанию, то его надо определить (см. ниже параметр *Как:*).

Установить связь: Обеспечивает связь объекта, вставляемого из буфера, с исходным документом (аналогично команде *Установить связь* из меню *Правка*).

Как: Позволяет выбрать формат объекта, вставляемого из буфера обмена.

Заменить объект: Эта операция доступна только в том случае, когда на графике системы *STATISTICA* выбран какой-либо объект. При этом вставка помещается вместо выбранного объекта (замещает его).

Поместить по умолчанию: При выборе этого режима объект помещается в левый верхний угол графического окна (это положение задано по умолчанию и может быть изменено, как и размеры объекта). Если отменить этот режим, то положение вставки на графике будет определяться положением курсора (т.е. объект будет помещен на график только после того, как будет "нарисован" прямоугольник и отпущена кнопка мыши).

Свойства объекта: Эта кнопка вызывает диалоговое окно *Стиль картинок и сложных объектов*, в котором можно задать различные параметры вставляемого элемента.

Просмотр буфера обмена: Открывает окно *Просмотр буфера обмена*, в котором можно просмотреть содержимое буфера.

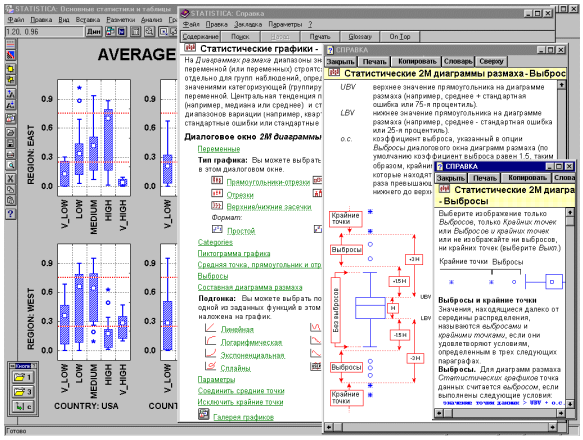
Техническое замечание: отображение вставляемых графиков системы *STATISTICA*. Для графика в формате системы *STATISTICA* доступны некоторые дополнительные опции вставки, параметры которых определяются в диалоговом окне *Стиль картинок и сложных объектов*. Это окно вызывается перед вставкой объекта из диалогового окна *Специальная вставка*. Его можно также открыть и после помещения объекта на график с помощью контекстного меню, вызываемого правой клавишей мыши, или с помощью комбинации клавиш ALT+ENTER (если объект выделен). Например, при выборе режима *Фиксир. размер шрифта* (в поле *Режим отображения*) шрифт вставляемого графического объекта будет привязан к значению *базы отображения* основного окна, и, следовательно, он будет автоматически изменяться при нажатии кнопок *Увеличить (уменьшить) все шрифты* (см. стр. 1103).

Текст. Заметим, что *текст*, помещенный из буфера обмена на график системы *STATISTICA,* может быть в дальнейшем отредактирован и настроен с помощью *Редактора текста графика* (см. выше описание кнопки).

35. Справка

Эта кнопка располагается последней на любой панели инструментов и вызывает интерактивное *Электронное руководство*. При нажатии справочной кнопки или клавиши F1 система *STATISTICA* предоставляет исчерпывающее описание всех программных процедур и функций в контекстно-зависимом режиме.

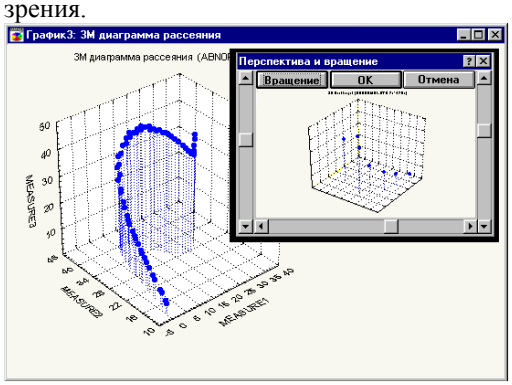
Copyright © StatSoft, 1995



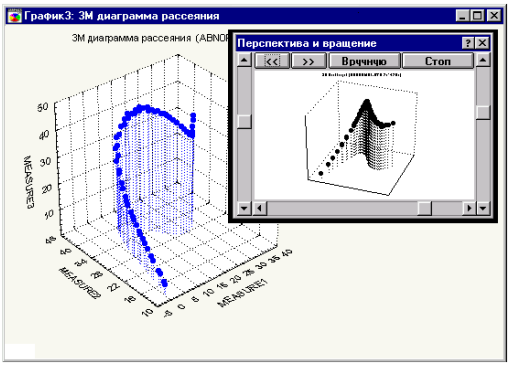
Другие кнопки графических панелей инструментов. Описанные ниже кнопки присутствуют только на панелях инструментов трехмерных графиков.

36. Вращение графика

Эта кнопка вызывает диалоговое окно *Перспектива и вращение* (см. ниже), в котором можно поворачивать трехмерное изображение и интерактивно подбирать угол зрения.



Если нажать кнопку *Вращение*, то изображение начнет вращаться по или против часовой стрелки.

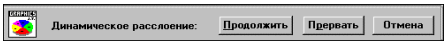


Первоначально в этом окне будет перерисовываться упрощенное отображение графика (для экономии времени). Если использовать функцию *Вращение*, то на картинке будет представлен полный набор данных.

Заметим, что конкретные настройки угла зрения и перспективы (например, для точного воспроизведения изображения) можно сделать и с помощью задания числовых значений, нажав кнопку *Дополнительно…* в диалоговом окне *Общая разметка* (для этого используйте меню *Разметки* или дважды щелкните на графике за пределами осей).

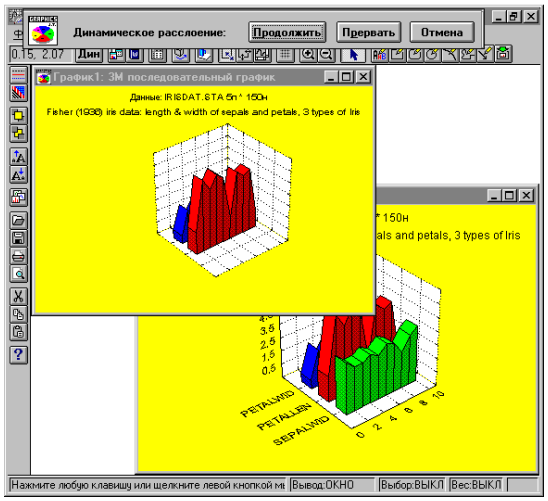
37. Динамическое расслоение

Эта кнопка задает способ перерисовывания графика в режиме *Динамическое расслоение* . Он доступен для всех трехмерных последовательных графиков. В этом режиме график перерисовывается послойно (последовательно "слой за слоем"), а для его подробного просмотра и анализа делается пауза после изображения каждого "сечения" (например, слоя поверхности, полосы, ряда блоков или столбцов диаграммы). Этим процессом можно управлять со специальной панели, которая возникает в верхней части экрана (см. ниже).



StatSoft
Copyright © StatSoft, 1995

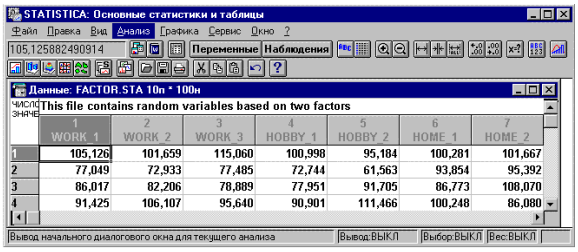
При каждом нажатии на кнопку *Продолжить* рисуется следующий "слой". Если нажать кнопку *Прервать*, то будут дорисованы все оси и метки, но остальные слои изображены не будут (это позволит вывести на печать текущее сечение). Кнопка *Отмена* означает выход из режима динамического расслоения, после чего график полностью перерисовывается.



Copyright © StatSoft, 1995
StatSoft®

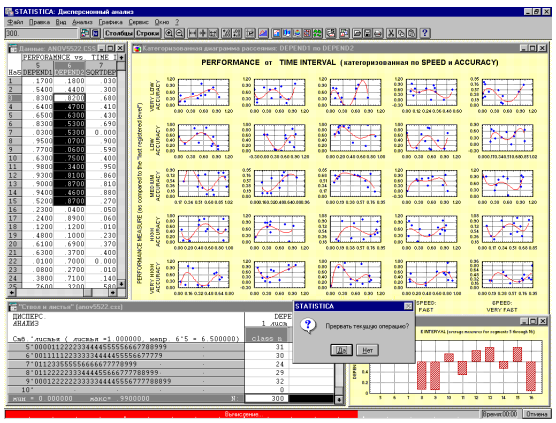
СТРОКА СОСТОЯНИЯ (В НИЖНЕЙ ЧАСТИ ОКНА СИСТЕМЫ *STATISTICA*)

Строка состояния, расположенная в нижней части окна системы *STATISTICA*, используется для вывода краткой справки или описания, а также предоставляет быстрый доступ к некоторым наиболее часто используемым функциям системы.



0. Индикатор состояния с таймером

Когда начинается обработка данных, строка состояния преобразуется в индикатор состояния с таймером. Кнопка *Отмена* в конце строки позволяет остановить текущую обработку данных.



Остановка анализа

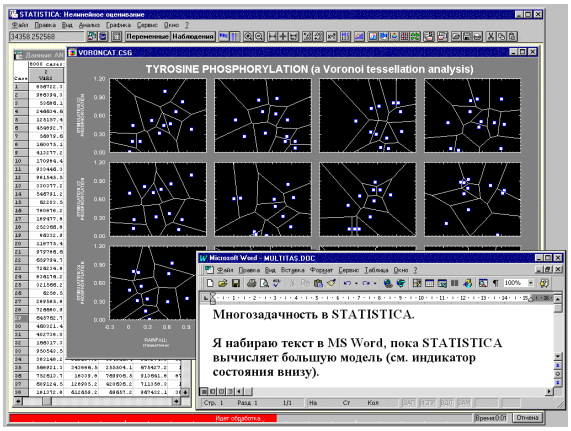
Таймер показывает время, которое прошло с начала обработки данных. Если временно прервать процесс обработки, таймер по прежнему будет работать, а в случае продолжения обработки автоматически вычтет из своих показаний время перерыва (для этого нужно ответить *Нет* на приведенный выше запрос).

Остановка перерисовывания графиков. Обратите внимание, что перерисовывание графиков, не связанных с отображением индикатора состояния, можно остановить, нажав любую клавишу или кнопку мыши (в любом месте экрана).

Остановка выполнения программы на языке *SCL*. Выполнение программы на языке *SCL* (*Командный язык системы STATISTICA*) можно остановить, нажав кнопку мыши или клавишу ESC или CTRL+BREAK.

Многозадачность

Система *STATISTICA* поддерживает режим многозадачности (между своими модулями или другими приложениями).



При обработке очень больших объемов данных или выполнении сложных анализов можно

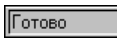
StatSoft®
Copyright © StatSoft, 1995

переключиться в другой модуль системы *STATISTICA* (или другое приложение Windows), при этом обработка будет производиться в фоновом режиме.

Поля контроля строки состояния

Как уже упоминалось ранее, если данные не обрабатываются, то в строке состояния отображается краткая справка или описание. Кроме того, она предоставляет быстрый доступ к некоторым часто используемым функциям, позволяющим контролировать вывод (например, принтер, файл вывода, *Окно текста/вывода*, стиль отчета) и обработку данных (условия выбора и веса наблюдений). В ней также отображается состояние записи или выполнения макрокоманды системы *STATISTICA*.

Все контролирующие функции, вызываемые через строку состояния, также доступны через меню и комбинации клавиш (поэтому их можно включать в клавиатурные макрокоманды).



1. Поле сообщений

Поле сообщений строки состояния может служить "кнопкой справки" - двойной щелчок на ней выводит на экран окно с предметным указателем справки. В этой области также отображаются:

- краткое пояснение к выделенному элементу меню или нажатой кнопке панели инструментов, как показано выше (чтобы получить краткое описание кнопки панели инструментов, не запуская соответствующую операцию, нужно удерживать кнопку нажатой во время чтения сообщения, затем, не отпуская кнопку мыши, перевести курсор за пределы кнопки и только тогда отпустить кнопку мыши);
- информация о выполняемой операции (например, *Выделение факторов…*, *Идет обработка…* или *Вычисление остатков…*);
- краткие инструкции, относящиеся к текущей стадии анализа или выполняемой операции (например, когда перерисовывается или обновляется большой график, сообщение будет таким: *Нажмите любую клавишу или щелкните левой кнопкой мыши для прерывания перерисовывания*; если выбран режим *Рисование прямоугольников*, сообщение будет таким: *Перемещайте курсор для создания прямоугольника (для отмены нажмите Esc или щелкните мышью на кнопке "Режим выделения")*).

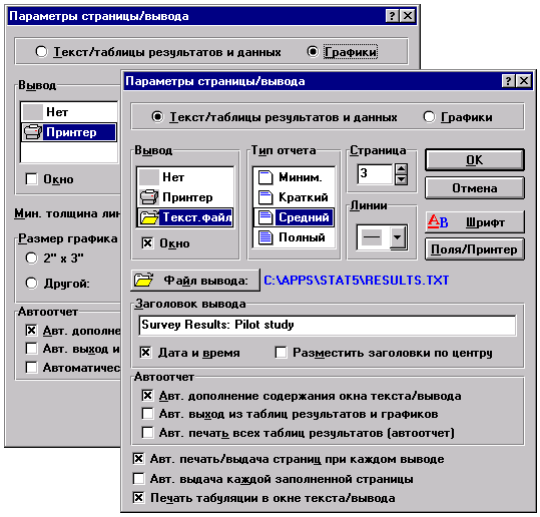
2. Контроль вывода

В поле *Вывод: ВЫКЛ* отображается текущее состояние канала вывода (например, *ПРИНТЕР* - вместо *ВЫКЛ*). Это поле также действует как кнопка: дважды щелкнув на ней, можно вызвать диалоговое окно *Параметры страницы/вывода,* чтобы распечатать таблицы результатов, таблицы исходных данных или графики. Это окно также отображается при запросе вывода (например, при нажатии кнопки *Принтер* панели инструментов), если канал вывода не указан (принтер, файл или *Окно текста/вывода*). Кроме того, оно вызывается одноименной командой выпадающего меню *Файл*. Диалоговое окно имеет две закладки *Текст/Таблицы результатов и данных* и *Графики*, которые открываются при выборе соответствующего пункта.

Печать текста/таблиц результатов/таблиц исходных данных. Значения параметров на этой закладке диалогового окна *Параметры страницы/вывода* влияют на печать всего текстового (не графического) вывода, отображаемого в таблицах результатов, таблицах

Copyright © StatSoft, 1995

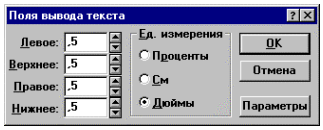
исходных данных, полученного на языке *STATISTICA BASIC* и т.п.



Канал вывода. В этом поле можно задать вывод на *Принтер*, в *Файл* и/или в *Окно* (экранный отчет, содержащий текст и графики вывода; информацию об *Окне текста/вывода*, см. в разделе *Панель инструментов Окна текста/вывода*, стр. 1081).

Типы отчета. Можно выбрать один из четырех *типов отчета*, которые определяют количество дополнительной информации о текущем файле и выбранных переменных, автоматически включаемой в таблицу результатов при выводе (*Миним.*, *Краткий*, *Средний* и *Полный*; описание типов можно вызвать, нажав кнопку ? или клавишу F1).

Поля/Принтер. Кнопка *Поля/Принтер* открывает диалоговое окно, в котором можно изменять поля вывода текста.



Чтобы выбрать ориентацию бумаги (книжную и альбомную) для вывода текста, можно нажать кнопку *Параметры* (это аналогично выбору опции *Принтер* выпадающего меню *Файл*).

Обратите внимание, что поля, ориентация бумаги и другие параметры печати графиков контролируются отдельно (см. ниже).

Другие параметры. Также можно изменять другие параметры распечатки, например, номер начальной страницы и печать линий сетки в таблицах (таблицах исходных данных и таблицах результатов), и изменение стиля линий сетки (см. ниже), используемый шрифт, а также текст заголовка, который будет печататься на каждой странице, дату, время и центрирование заголовков. Наконец, это диалоговое окно позволяет установить режим выдачи страниц и режим автоотчета (см. ниже).

Линии сетки в выводе и скорость печати. Обратите внимание, что линии сетки печатаются как графика высокого разрешения; поэтому печать вывода без линий сетки производится быстрее (если поле списка *Линии* оставлено пустым).

Режимы автоотчета. Следующие режимы *автоотчета* позволяют автоматически создавать отчет (например, отчет, содержащий текст и графики) обо всех проводимых процедурах анализа (автоотчет для графиков см. на стр. 1115).

Авт. выход из таблиц результатов и графиков. Когда таблицы результатов (и/или графики, см. ниже) создаются совместно как результат анализа (например, таблицы частот для рядов переменных), можно выбрать эту опцию, чтобы не отображать в таблицах результатов кнопку *Далее,* если очередь переполнена. Теперь если выбрать режим *Авт. печать всех таблиц результатов*, таблицы результатов будут напечатаны в пакетном режиме, без перерыва.

StatSoft®
Copyright © StatSoft, 1995

Авт. печать всех таблиц результатов. Этот переключатель позволяет автоматически создавать текстовый отчет из всех таблиц результатов, отображаемых на экране, освобождая от необходимости нажимать кнопку *Принтер* для печати отдельных таблиц. Содержание каждой таблицы результатов, отображаемой на экране, автоматически отправляется в текущий канал вывода (принтер, файл и/или *Окно текста/вывода*) с соблюдением стиля отчета и соглашений, указанных в этом диалоговом окне и диалоговом окне *Печать таблицы результатов* (см. главу 4). При печати в *Окно текста/вывода* совместно с использованием режима *Автоматически печатать все графики* (см. печать графиков, стр. 1115) можно создать комбинированный отчет, содержащий текстовые и графические результаты, полученные в ходе анализа. Обратите внимание, что также существует функция печати в пакетном режиме, которую можно использовать для совместной печати сохраненных ранее файлов таблиц результатов (расширение **.scr*), для этого необходимо выбрать пункт *Печать файлов...* выпадающего меню *Файл*, см. главу 3.

Авт. дополнение содержания Окна текста/вывода. Можно поставить галочку в этом поле, чтобы автоматически открывать сохраненный последним файл отчета в новом *Окне текста/вывода* (например, чтобы добавить текущий вывод к результатам предыдущего анализа). Если галочка в этом поле не установлена, будет открыто пустое *Окно текста/вывода* из диалогового окна *Параметры страницы/вывода*.

Авт. печать/выдача страниц при каждом выводе. Если включен этот режим, то после каждого текстового вывода (например, с помощью команд *Печать данных* или *Печать таблицы результатов* или кнопки *Печать* панели инструментов), страницы будут "выданы" на принтер (или в файл вывода). Поэтому каждая распечатка (таблица) будет начинаться с новой страницы. Если вывод достаточно короткий (например, маленькая таблица результатов), но часто повторяется, то выбор этого режима может замедлить работу, поскольку процесс печати будет запускаться очень часто (для каждой маленькой части вывода). В то же время, преимущество этого метода состоит в том, что выдача на печать происходит непосредственно после запроса.

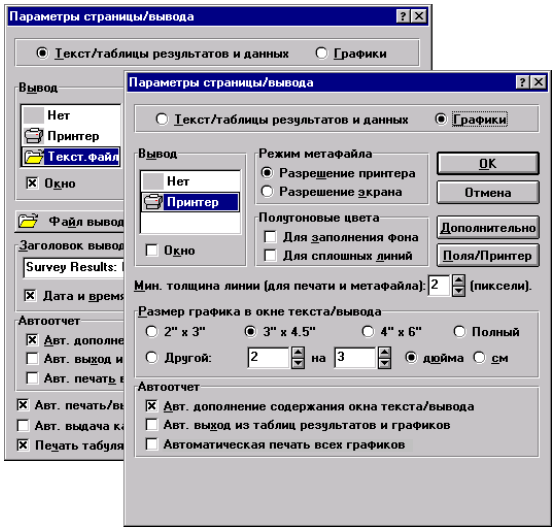
Авт. выдача каждой заполненной страницы. При включении этого режима каждая страница распечатки будет закончена (выдана) тогда, когда будет заполнена страница. Этот режим несколько медленнее, поскольку процесс печати запускается отдельно для каждой заполненной страницы. Однако, преимущество такой настройки в том, что распечатка вывода создается автоматически в процессе анализа. Кроме того, если выбран этот режим (совместно с режимом *Авт. печать/выдача*, см. предыдущий абзац), то весь вывод будет накапливаться во внутреннем буфере и отправляться на принтер один раз в конце сеанса (или когда будет выбрана опция *Печать/выдача текущих страниц* выпадающего меню *Файл*, которая позволяет "очистить" буфер вывода в любое время).

Печать табуляции в окне текста/вывода. Если выбран этот режим, столбцы таблицы результатов при печати в *Окно текста/вывода* будут разделены символами табуляции (это параметр по умолчанию). Форматированные таким образом таблицы можно легко преобразовать в таблицы текстовых редакторов (например, MS Word).

Печать графиков. Возможности настройки, предлагаемые в этой части диалогового окна

Copyright © StatSoft, 1995

Параметры страницы/вывода определяют режим печати графиков.



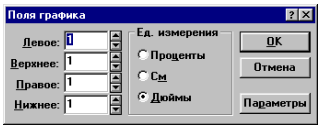
Канал вывода. Выбрав в этом поле соответствующие значения, можно направить вывод на *Принтер* и/или в *Окно* - (экранный отчет, содержащий текст и графики вывода; информацию об *Окне текста/вывода*, см. в разделе *Панель инструментов Окна текста/вывода* на стр. 1081).

Режим метафайла. Все графические метафайлы, получаемые в системе *STATISTICA* (представления буфера обмена, сохраненные метафайлы и графики, выводимые в *Окно текста/вывода*), могут быть представлены либо в разрешении принтера (которое предоставляет наилучшее качество, но требует больше памяти или дискового пространства), либо в разрешении экрана (которое предоставляет лучшее отображение на экране).

Полутоновые цвета. Эти параметры позволяют контролировать использование полутоновых цветов при заполнении фона графика, пользовательского текста, условных обозначений, внедренных объектов и т.п. (по умолчанию этот режим не выбран, чтобы вид графика при печати был оптимальным), а также при печати сплошных линий (выбор этого режима имеет смысл, если линии достаточно толстые; в противном случае они могут оказаться “разорванными”).

***Дополнительные* возможности.** Кнопка *Дополнительно* открывает диалоговое окно с дополнительными, редко используемыми опциями, которые позволяют контролировать создание графического вывода, включающего графические объекты *OLE*.

***Поля/Принтер*.** Кнопка *Поля/Принтер* открывает диалоговое окно, в котором можно изменять поля вывода графиков.



Чтобы выбрать ориентацию бумаги (книжную и альбомную) для вывода графика, можно нажать кнопку *Параметры* (это аналогично выбору команды *Принтер* выпадающего меню *Файл*).

Обратите внимание, что поля, ориентация бумаги и другие параметры печати текста контролируются отдельно (см. выше).

Мин. толщина линии. Здесь указывается минимальная толщина линии для распечаток и метафайлов (в пикселях принтера или устройства). Эта опция особенно полезна при использовании устройства вывода высокого разрешения (например, 1200 DPI, где линии толщиной в один пиксель очень тонкие), а также когда по умолчанию размер графика в *Окне текста/вывода* установлен очень маленьким (например, 2x3, см. ниже описание поля *Размер графика в Окне текста/вывода*).

Если не увеличить значение этого параметра, установленное по умолчанию (2 пикселя принтера), то ширина всех линий, печатаемых системой *STATISTICA*, будет определяться либо

StatSoft®
Copyright © StatSoft, 1995

значениями ширины линии для определенных компонент графика (по умолчанию 0,5 пункта), либо будет равна 2 пикселя (устройства), в зависимости от того, какая из этих величин больше. Эта опция позволяет определять минимальную толщину печатаемой линии. Например, если установить значение 3, то все печатаемые линии будут иметь ширину по крайней мере 3 пикселя принтера, даже если в некоторых диалоговых окнах *Точки/Линии/Шаблоны областей* для них установлены меньшие значения.

Если *Размер графика в Окне текста/вывода* имеет значение меньше, чем *полный*, то толщина линий будет пропорционально уменьшена. Для компенсации этого, возможно, понадобится увеличить минимальную толщину линии.

Размер графика. В этом поле можно выбрать размер по умолчанию, который используется для отображения графиков при печати в *Окно текста/вывода* (после печати размер графиков можно изменить).

Режимы автоотчета. Следующие режимы *автоотчета* позволяют автоматически создавать отчет (содержащий, например, текст и графики) обо всех производимых анализах (автоотчет для таблиц результатов и таблиц исходных данных см. выше).

Авт. дополнение содержания окна текста/вывода. Можно поставить галочку в этом поле, чтобы в новом *Окне текста/вывода* автоматически открывать сохраненный последним файл отчета (например, для добавления текущего вывода к результатам предыдущего анализа). Если галочка в этом поле не установлена, будет открыто пустое *Окно текста/вывода* из диалогового окна *Параметры страницы/вывода*.

Авт. выход из таблиц результатов и графиков. Когда графики (и/или таблицы результатов, см. ниже) создаются в пакетном режиме как результат анализа (например, кнопка *Категориз. диаграммы рассеяния* в диалоговом окне *Описательные статистики*), можно выбрать этот режим, чтобы при переполнении очереди не отображать в таблицах результатов кнопку *Далее*. Теперь если выбрать режим *Автоматическая печать всех графиков*, графики будут напечатаны в пакетном режиме, без перерыва.

Автоматическая печать всех графиков. Этот переключатель позволяет автоматически создавать распечатки всех графиков, отображаемых на экране. Вывод будет направляться в текущий канал вывода (принтер и/или *Окно текста/вывода*). При этом будут использованы текущие настройки (например, тип принтера, поля, ориентация бумаги и т.д.), выбранные в диалоговом окне *Параметры печати* для графиков. Обратите внимание, что также существует функция печати в пакетном режиме, которую можно использовать для печати в пакетном режиме сохраненных ранее файлов таблиц результатов (расширение **.scr*), для этого необходимо выбрать команду *Печать файлов...* выпадающего меню *Файл*, см. главу 3.

При печати в *Окно текста/вывода* совместно с использованием режима *Авт. печать всех таблиц результатов* можно создать комбинированный отчет, содержащий текстовые и графические результаты, полученные в ходе анализа. Информацию о форматах отчета *Окна текста/вывода* см. на стр. 1081.

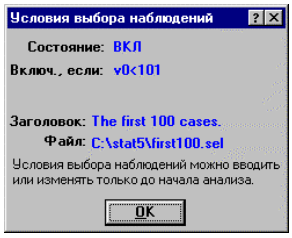
3. Условия выбора наблюдений

В поле *Выбор* отображается текущее состояние *условий выбора наблюдений*, то есть, заданных

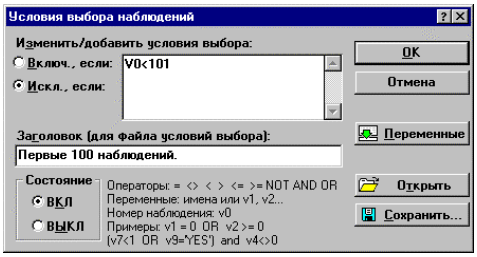
Copyright © StatSoft, 1995

пользователем условий (или “фильтров”), которые можно использовать для выбора конкретного подмножества наблюдений для анализа.

Обычно, при отсутствии пропущенных данных обрабатываются все наблюдения файла данных. Однако можно временно ограничить анализ только выбранным подмножеством из набора данных (например, выделить только женщин до 60 лет, у которых либо высокий уровень холестерина, либо высокое кровяное давление). Поле *Выбор* также действует как кнопка: дважды щелкнув на нем, можно открыть окно состояния *Условия выбора наблюдений*, в котором отображаются заданные условия (если они есть).



Если условия были взяты из файла, то отображается сообщение о нем (например, комментарий). В этом окне содержится только информация об *условиях выбора наблюдений* и не предусмотрена возможность изменения заданных условий в процессе анализа, поскольку в противном случае на различных стадиях вычислений использовались бы различные наборы данных. Однако если дважды щелкнуть на этом поле строки состояния, например, до начала анализа (когда условия выбора можно изменить), то откроется другое окно, в котором можно ввести или отредактировать текст условий.



Синтаксис этих условий очень простой (обратите внимание на примеры в поле справки диалогового окна).

Вызывать переменные можно по их номерам (например, *v1*, *v2*, *v3*, ...; обратите внимание, что *v0* это имя наблюдения) или именам (например, *Income*, *Profit*). Поэтому, например, условие:

Включ., если: `v0<101 and v1=1`

включает в анализ только первые 100 наблюдений (т.е., номер наблюдения должен быть меньше, чем *101*), в которых значение переменной 1 (т.е., *v1*) равно *1*. Обратите внимание, что если имя переменной 1 было *Gender* и для этой переменной *1* было эквивалентно текстовому значению *MALE* (см. раздел о текстовой/числовой “двойной записи” значений в системе *STATISTICA*, глава 7), то аналогичное условие выбора наблюдений можно было бы ввести так:

Включ., если: `v0<101 and GENDER = 'MALE'`

Синтаксис условий поддерживает использование скобок и различных операторов и позволяет задать условия выбора наблюдений практически неограниченной сложности.

Заданные условия можно сохранить в файле на диске для последующего использования (при выходе из системы *STATISTICA* будет выведен запрос о сохранении текущих условий).

Другие применения *условий выбора наблюдений*. Условия также можно сохранить и использовать в системе *STATISTICA* с другими наборами данных или для других целей.

StatSoft®

Copyright © StatSoft, 1995

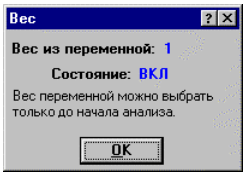
Последнее возможно благодаря одинаковому синтаксису и соглашениям, применяемым во всех функциях системы *STATISTICA*, которые используют или позволяют задавать пользовательские наборы наблюдений. Например, одинаковые соглашения применяются в функции *перекодировки* данных (вызываемой нажатием кнопки **Переменные** таблицы исходных данных, см. стр. 1056), в функции проверки данных (доступной в модуле *Управление данными*, см. главу 7), а также во всех процедурах, позволяющих задавать пользовательские наборы данных (такие как таблицы частот или графики, отображающие несколько подгрупп).

Условия выбора наблюдений можно также задать, нажав кнопку *Выбор наблюдений* . Эта кнопка включается во все стартовые панели и все диалоговые окна задания параметров анализа, которые открываются до начала обработки данных.

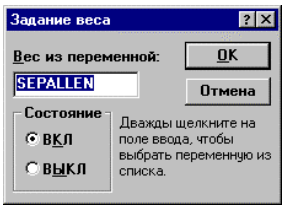
4. Вес наблюдений

Поле *Вес:* показывает текущее состояние функции *Вес наблюдения*, которая позволяет рассматривать значения выбранной переменной при обработке данных как (целый) множитель наблюдения. Вес можно использовать для аналитических процедур (например, некоторые наблюдения могут оказаться значительно более “важными” и эту важность можно представить некоторым значением веса), а также для экономии хранения данных (например, в некоторых больших наборах данных, таких как данные переписи или исследования, многие наблюдения могут оказаться идентичными и поэтому могут быть представлены одним наблюдением с соответствующим весом).

Поле *Вес:* также действует как кнопка: дважды щелкнув на поле *Вес*, можно открыть окно состояния *Вес* , в котором отображается переменная взвешивания (если она была задана).



Если с начала анализа уже прошло некоторое время, отображается только информация о весах наблюдений без возможности их изменения, поскольку в противном случае различные стадии вычислений были бы основаны на различной конфигурации данных. Однако если дважды щелкнуть на этом поле строки состояния, например, до начала анализа (когда вес можно изменить), то появится другое окно, в котором можно указать переменную для взвешивания.



Вес наблюдений можно также указать, нажав кнопку *Вес*: . Эта кнопка включена во все стартовые панели и диалоговые окна задания параметров анализа, которые отображаются до начала обработки данных.

5. Макрокоманды (текущее состояние)

Поле состояния *макрокоманды* (последнее поле строки состояния) обычно остается пустым, если выполняемая операция не включает записанную макрокоманду. При записи макрокоманды в поле состояния *макрокоманды* мигает надпись *ЗАП.*

Copyright © StatSoft, 1995

Чтобы остановить запись макрокоманды, можно нажать клавиши CTRL+F3.

В системе *STATISTICA* поддерживается два типа макрокоманд: изменяемые последовательности нажатий клавиш и записанные действия (последовательности движений мыши и/или нажатий клавиш). Записанные макрокоманды можно открыть и изменить в редакторе нажатий клавиш. Подробную информацию об этих типах макрокоманд можно найти в главе 6.

StatSoft®

Copyright © StatSoft, 1995

ПРЕДМЕТНЫЙ УКАЗАТЕЛЬ

А

Автоматическая выдача каждой заполненной страницы, 1113
Автоматическая выдача каждой страницы, 1113
Автоматическая печать всех графиков, 1115
Автоматическая печать всех таблиц результатов, 1113
Автоматический выход из таблиц результатов и графиков, 1115
Автоотчет (автоматическая печать всех графиков), 1115
Автоотчет (автоматическая печать всех таблиц результатов), 1113

Б

Быстрые основные статистики, 1060
  выбор переменных, 1060
  калькулятор вероятностных распределений, 1061
Быстрые статистические графики
  в таблицах исходных данных, 1065
  в таблицах результатов, 1077

В

Вероятностный калькулятор, 1061
Вложенные составные документы, 1101
Внедрение и связывание (на графиках), 1100
Вращение трехмерных графиков, 1108
Вставить
  в графическом окне, 1106
  в Окне текста/вывода, 1085
  в таблицах исходных данных, 1069
  в таблицах результатов, 1080
Выделенные значения в окне *Редактор данных графика*, 1073
Вынести на передний план (на графиках), 1102
Вырезать
  в графическом окне, 1106
  в Окне текста/вывода, 1085
  в таблицах исходных данных, 1067
  в таблицах результатов, 1080

Г

Галерея графиков системы *STATISTICA*, 1066, 1078, 1084, 1104
Глобальное изменение/реструктуризация наблюдений, 1057
Глобальное изменение/реструктуризация переменных, 1056
Глобальное увеличение шрифтов (на графиках), 1104
Глобальное уменьшение шрифтов на графиках, 1104
Графики
  связывание и внедрение, 1100
  текст, 1095
Графики XYZ, 1064, 1075
Графики на плоскости XY, 1062
Графическая панель инструментов, 1086

Д

Деловые графики, 1062
Диалоговое окно *Параметры страницы/вывода*, 1111
Динамическая связь с данными, 1062
Динамический режим (отображение координат) на графиках, 1086
Динамическое расслоение (на графиках), 1108
Добавить десятичный разряд, 1059

И

Индикатор состояния с таймером, 1110

К

Канал вывода, 1112, 1114
Кисть, 1088
Кнопка *Вес* (наблюдений), 1117
Кнопка *Диспетчер текстовых значений* (в таблицах исходных данных), 1060
Кнопки автозадач, 1053
Кнопки вызова графиков таблицы результатов, 1074
Копировать
  в Окне текста/вывода, 1085
  в таблицах исходных данных, 1068
  в таблицах результатов, 1080
  графики, 1106
Копировать и копировать содержание, 1068, 1080

Л

Линии сетки при выводе и скорость печати, 1112
Логические и физические размеры шрифта (на графиках), 1104
Логическое увеличение на графиках, 1093
Лупа (на графиках), 1093

М

Маркировать (снять маркировку) значения (в таблицах результатов), 1073
Метод DDE (Динамический обмен данными), 1062
Метод *Перетащить и отпустить* операции с переменными, 1057
Методы OLE, 1100
Многозадачность, 1110

Н

Наблюдения, условия выбора, 1115
Настройка положения графика и полей, 1091

О

Обращение операции увеличения (на графиках), 1094
Объединенное окно спецификаций переменных, 1056
Окно блокнота, 1081
Окно текста/вывода, 1081
Операции с буфером обмена в таблице исходных данных, 1069

Copyright © StatSoft, 1995

Операции с буфером обмена и глобальные операции, 1067
с наблюдениями, 1057
с переменными, 1056
Остановка анализа, 1110
Отображение размеров шрифта в различных по размерам областях графика, 1104

П

Панели инструментов
Окно текста/вывода, 1081
графическое окно, 1086
таблица исходных данных, 1054
таблица результатов, 1071
Панель состояния *Вес*, 1117
Переключение отображения имен наблюдений/номеров, 1058
Переключение отображения числовых/текстовых значений (в таблицах исходных данных), 1058
Пересчитать переменные, 1059
Подсказки, 1053
Поле *Вес* строки состояния, 1117
Поле *Выбор* строки состояния, 1115
Поле *Вывод* (в строке состояния), 1111
Поле просмотра (в таблицах исходных данных), 1054
Поле просмотра (в таблицах результатов), 1071
Поле просмотра на графической панели, 1086
Поле сообщений строки состояния, 1111
Поле состояния макрокоманды (строка состояния), 1117
Полутоновые цвета, 1114
Пользовательские 2M графики
в таблицах исходных данных, 1062
в таблицах результатов, 1074
Пользовательские 3M последовательные графики
в таблицах исходных данных, 1063
в таблицах результатов, 1075
Пользовательские XYZ графики
в таблицах исходных данных, 1063
в таблицах результатов, 1075
Пользовательские матричные графики
в таблицах исходных данных, 1064
в таблицах результатов, 1076
Пользовательские пиктографики
в таблицах исходных данных, 1065
в таблицах результатов, 1076
Поля, 1112, 1114
Поля текста/вывода, 1112, 1114
Поместить на задний план (на графиках), 1103
Прикрепление к сетке, 1092
Произвольное рисование, 1098
Пропорции графического окна по умолчанию, 1090
Просмотр графика в режиме увеличения, 1093
Протокол таблиц результатов и графиков, 1081

Р

Размер графика, уменьшение, 1094
Размер полей по умолчанию, 1092
Расширенный текстовый формат (RTF), 1081
Редактирование формы линии на графиках, 1099
Редактор данных графика (кнопка), 1087
Редактор текста графика, 1095
Режим выделения (указатель) на графиках, 1094
Результаты, отображение графики, 1086
окно текста/вывода, 1081
таблицы результатов, 1071
Рисование в режиме увеличения, 1093
Рисование дуг, 1097
Рисование ломаных линии/многоугольников, 1098
Рисование обозначений ошибок, 1099
Рисование прямоугольников, 1097
Рисование скругленных прямоугольников, 1097
Рисование стрелок и обозначений ошибок, 1099
Рисование эллипсов, 1097

С

Связывание (на графиках), 1100
Связывание и внедрение, обзор, 1100
Специальные средства заполнения, 1102
Специальные средства настройки линий, 1102
Статистический советник, 1070
Стиль линии на графиках, 1102
Столбец, ширина (в таблицах исходных данных), 1059
Строка состояния, 1110

Т

Таблица исходных данных панель инструментов, 1054
Таблица результатов панель инструментов, 1071
Таймер, 1110
Текстовые файлы, 1081
Тип шаблона на графиках, 1102
Типы отчета, 1112

У

Увеличение (в таблицах исходных данных), 1058
Увеличение (в таблицах результатов), 1072
Увеличение (на графиках), 1093
Удалить десятичный разряд, 1059
Уменьшение (в таблицах исходных данных), 1059
Уменьшение (в таблицах результатов), 1072
Уменьшение (на графиках), 1094
Управляющие символы, 1096
Условия выбора, 1115

Ф

Файлы Рабочей книги, 1055
автоматически открывать, 1055
заблокировать, 1055
Фиксированный режим (отображение координат) на графиках, 1087
Фиксировать пропорции графика, 1091
Фиксировать пропорции на графиках, 1090

StatSoft®
Copyright © StatSoft, 1995

Ш

Ширина имени строки (в таблицах результатов), 1072
Ширина полей (в таблицах результатов), 1072
Ширина столбцов (в таблицах результатов), 1072
Ширина текущего столбца (в таблицах исходных данных), 1059

Я

Ячейки, маркированные программой, 1073

Copyright © StatSoft, 1995

StatSoft®

Copyright © StatSoft, 1995

Глава 3:

ОКНО ТАБЛИЦЫ ИСХОДНЫХ ДАННЫХ: МЕНЮ, ДИАЛОГОВЫЕ ОКНА, СВОЙСТВА

Оглавление

Операции с блоками и *автозаполнение* в таблицах исходных данных ... 1132
Разделение прокрутки в таблицах исходных данных и таблицах результатов ... 1136
Динамический обмен данными (*DDE*) ... 1138
Выпадающее меню *Файл* таблицы исходных данных ... 1132
Выпадающее меню *Правка* таблицы исходных данных ... 1167
Выпадающее меню *Вид* таблицы исходных данных ... 1208
Выпадающее меню *Анализ* таблицы исходных данных ... 1215
Выпадающее меню *Графика* таблицы исходных данных ... 1219
Выпадающее меню *Сервис* таблицы исходных данных ... 1223
Выпадающее меню *Окно* таблицы исходных данных ... 1235
Выпадающее меню *?(Справка)* таблицы исходных данных ... 1238
Предметный указатель ... 1241

Подробное оглавление

ОПЕРАЦИИ С БЛОКАМИ И *АВТОЗАПОЛНЕНИЕ* В ТАБЛИЦАХ ИСХОДНЫХ ДАННЫХ ... 1132
Возможности метода *Перетащить и отпустить* ... 1132
Выделение блока ... 1132
Перемещение блока ... 1133
Копирование блока ... 1133
Вставка блока ... 1133
Очистка блока ... 1134
Расширение файла данных методом *Перетащить и отпустить* ... 1134
Экстраполяция блока (*автозаполнение*) ... 1135
Изменение скорости выделения блока ... 1136

РАЗДЕЛЕНИЕ ПРОКРУТКИ В ТАБЛИЦАХ ИСХОДНЫХ ДАННЫХ И ТАБЛИЦАХ РЕЗУЛЬТАТОВ 1136

ДИНАМИЧЕСКИЙ ОБМЕН ДАННЫМИ (*DDE*) ... 1138
Введение ... 1138
Создание связей ... 1138
Установить связь ... 1138
Длинное имя переменой ... 1139
Диспетчер связей ... 1140
Связи *DDE*, *Командный язык* системы *STATISTICA* и *Модуль запуска* ... 1140

Copyright © StatSoft, 1995

StatSoft®

ВЫПАДАЮЩЕЕ МЕНЮ *ФАЙЛ* ТАБЛИЦЫ ИСХОДНЫХ ДАННЫХ.....1142
Создать данные 1142
Открыть другие... 1142
Открыть файл данных 1143
Открыть графический файл 1143
Открыть файл таблицы результатов..... 1143
Открыть файл текста/вывода 1143
Открыть программу на языке *STATISTICA BASIC*..... 1144
Открыть файл командного языка (*SCL*)..... 1144
Диалоговое окно *Открытие/Сохранение файла* 1144
Имя файла..... 1145
Список файлов по типу 1145
Папки 1145
Диски 1145
Рабочая книга..... 1145
Как текст..... 1145
Импорт данных – Быстрый импорт..... 1146
Быстрый импорт - Таблицы..... 1147
Быстрый импорт – Файлы баз данных..... 1148
Быстрый импорт – Текстовые файлы 1148
Импорт - *ODBC*..... 1149
Выбор полей и условий для выбора строк 1149
Сохранить таблицу исходных данных 1150
Сохранение таблицы исходных данных под другим именем..... 1150
Быстрый экспорт..... 1151
Быстрый экспорт в форматы Excel/Lotus/Quattro 1151
Быстрый экспорт текстовых (ASCII) файлов..... 1152
Преобразование в таблицу результатов 1153
Первое наблюдение 1153
Последнее наблюдение 1153
Первая переменная 1153
Последняя переменная 1153
Печать/выдача текущих страниц 1153
Печать данных..... 1154
Печатать..... 1154
Переменные..... 1154
Все данные 1155
Печать значений данных..... 1155
Печать/выдача страниц при выводе 1155
Параметры..... 1155
Параметры вывода текста на принтер..... 1155
Печать графиков и таблиц результатов..... 1156
Список файлов для вывода на печать 1156
Добавить файлы..... 1156

StatSoft®

Copyright © StatSoft, 1995

Удалить ... 1157
Отмена ... 1157
Параметры страницы/вывода (текст/таблицы результатов и данных) ... 1157
Вывод ... 1157
Тип отчета ... 1158
Файл вывода ... 1158
Страница ... 1159
Линии ... 1159
Заголовок вывода ... 1159
Дата и время ... 1159
Разместить заголовки по центру ... 1159
Шрифт ... 1159
Поля/Принтер ... 1159
Автоматическая печать/выдача страниц при каждом выводе ... 1159
Автоматическая выдача каждой заполненной страницы ... 1160
Режимы автоотчета ... 1160
Автоматическая печать/выдача страниц при каждом выводе ... 1160
Автоматическая выдача каждой заполненной страницы ... 1161
Печать табуляции в окне текста/вывода ... 1161
Параметры страницы/вывода (графики) ... 1161
Вывод ... 1161
Режим метафайла ... 1161
Полутоновые цвета ... 1162
Поля/Принтер ... 1162
Минимальная толщина линии ... 1162
Размер графика в окне текста/вывода ... 1162
Автоотчет ... 1162
Дополнительные параметры ... 1163
Метафайлы в Окне текста/вывода ... 1163
Растровый рисунок (буфер обмена, сохранение) ... 1164
Режим сервера *OLE* ... 1164
Проверять ошибки в метафайлах ... 1164
Поля вывода текста ... 1164
Поля графика ... 1164
Принтер… ... 1165
Принтер ... 1165
Ориентация ... 1165
Бумага ... 1165
Параметры ... 1165
Выход ... 1166

ВЫПАДАЮЩЕЕ МЕНЮ *ПРАВКА* ТАБЛИЦЫ ИСХОДНЫХ ДАННЫХ ... 1167
Изменение ячейки ... 1167
Добавить переменные ... 1167

Copyright © StatSoft, 1995

Число переменных ... 1167
Вставить после ... 1167
Переместить переменные ... 1168
С переменной ... 1168
По переменную ... 1168
Вставить после ... 1168
Копировать переменные ... 1169
С переменной ... 1169
По переменную ... 1169
Вставить после ... 1169
Удалить переменные ... 1170
С переменной ... 1170
По переменную ... 1170
Изменение (текущих) спецификаций ... 1170
Имя ... 1171
Код ПД ... 1171
Формат отображения ... 1171
Длинное имя - Метка ... 1173
Длинное имя - Формула ... 1173
Длинное имя - Связь ... 1174
Все параметры ... 1174
Текстовые значения ... 1174
Значения/статистики ... 1174
Графики ... 1175
Кнопки << , >> ... 1175
Формулы в таблице исходных данных: Мастер функций ... 1175
Обзор синтаксиса формул системы *STATISTICA* ... 1176
Все спецификации ... 1178
Изменение текстовых значений ... 1178
Новое “текст-числовое” присвоение ... 1179
Изменение “текст-числового” присвоения ... 1179
Текст ... 1179
Число ... 1179
Метка значения ... 1179
Удалить строку ... 1180
Очистить все ... 1180
Сортировать: Текст - возрастание ... 1180
Сортировать: Текст - убывание ... 1180
Сортировать: Числа - возрастание ... 1180
Сортировать: Числа - убывание ... 1180
Копировать из ... 1180
Операции над датами ... 1180
Создать из 2 или 3 переменных ... 1181
Разбить на 2 или 3 переменные ... 1181

StatSoft®

Copyright © StatSoft, 1995

Текстовые значения/Даты ... 1181
Исходные переменные ... 1181
Результаты ... 1181
Пример 1 (создание даты) ... 1182
Пример 2 (разделение переменной, содержащей дату) ... 1182
Текстовые значения- /Даты ... 1183
Преобразование переменной ... 1183
Исходная переменная ... 1183
Результат ... 1183
Формат ... 1183
Пересчитать ... 1184
Все переменные ... 1184
Переменная ... 1184
Подмножество ... 1184
Автоматический пересчет при изменении данных ... 1184
Сдвинуть (лаг) ... 1184
Переменная ... 1185
Лаг ... 1185
Направление ... 1185
Ранжировать ... 1185
Переменные ... 1185
Наблюдения ... 1185
Задать вес ... 1185
Присвоить ранг 1 ... 1186
Ранги для совпадающих значений ... 1186
Тип рангов ... 1186
Перекодировать ... 1186
Категория ... 1187
Переменные ... 1188
Открыть все ... 1188
Сохранить ... 1188
Печать ... 1188
Другие ... 1188
Добавить наблюдения ... 1188
Число наблюдений ... 1189
Вставить после ... 1189
Переместить наблюдения ... 1189
С наблюдения ... 1189
По наблюдение ... 1189
Вставить после ... 1189
Копировать наблюдения ... 1190
С наблюдения ... 1190
По наблюдение ... 1190
Вставить после ... 1190

Copyright © StatSoft, 1995

Удалить наблюдения ... 1191
С наблюдения ... 1191
По наблюдение ... 1191
Изменение имен наблюдений ... 1191
Имена наблюдений ... 1191
Параметры ... 1192
Параметры имен наблюдений ... 1192
Ширина ... 1192
Взять из ... 1192
Буфер обмена ... 1192
Заголовок файла данных, примечания и Рабочая книга ... 1192
Строка заголовка файла данных ... 1193
Информация о файле и примечания ... 1193
Рабочая книга ... 1193
Вырезать ... 1195
Копировать ... 1195
Копировать содержание ... 1196
Вставить ... 1196
Вставить в графики ... 1196
Очистить ... 1197
Установить связь ... 1197
Ошибки *DDE* ... 1198
Автоматически ... 1198
Вручную ... 1198
Отмена ... 1198
Найти ... 1198
Найти ... 1198
Направление ... 1199
Искать ... 1199
Найти далее ... 1199
Заменить ... 1199
Найти ... 1199
Заменить на ... 1199
Направление ... 1199
Поиск ... 1199
Найти далее ... 1199
Заменить все ... 1200
Транспонировать блок ... 1200
Транспонирование файла данных ... 1200
Заполнить/стандартизировать блок - Заполнить случайными значениями ... 1201
Заполнить/стандартизировать блок - Заполнить/копировать вниз ... 1201
Заполнить/стандартизировать блок - Заполнить/копировать вправо ... 1201
Заполнить/стандартизировать блок – Стандартизировать столбцы ... 1202
Заполнить/стандартизировать блок – Стандартизировать строки ... 1202

StatSoft®
Copyright © StatSoft, 1995

Статистики блока по столбцам ... 1202
Статистики блока по строкам ... 1203
Блоковые статистические графики ... 1204
Гистограммы ... 1204
Захват экрана ... 1205
Связи ... 1206
Связи ... 1206
Обновление ... 1206
Обновить сейчас ... 1207
Отменить связь ... 1207
Изменить связь ... 1207

ВЫПАДАЮЩЕЕ МЕНЮ *ВИД* ТАБЛИЦЫ ИСХОДНЫХ ДАННЫХ ... 1208

Текстовые значения ... 1208
Имена наблюдений ... 1208
Полные имена переменных ... 1208
Диалоговое окно *Шрифт* ... 1208
Шрифт ... 1209
Начертание ... 1209
Размер ... 1209
По умолчанию ... 1209
Применить ... 1209
Пользовательские цвета для таблицы результатов и таблицы исходных данных ... 1209
Элемент экрана ... 1210
Линии сетки ... 1210
Применять по умолчанию ... 1210
Диалоговое окно *Цвет* ... 1210
Окно текста/вывода ... 1211
Заголовки ... 1212
Панель инструментов ... 1212
Изменить поле просмотра ... 1213
Строка состояния ... 1213
Комментарий ... 1213
Вывод ... 1213
Выбор ... 1214
Вес ... 1214
Запись ... 1214
Индикатор состояния ... 1214
Состояние ... 1214
Затраченное время ... 1214
Отмена ... 1214

ВЫПАДАЮЩЕЕ МЕНЮ *АНАЛИЗ* ТАБЛИЦЫ ИСХОДНЫХ ДАННЫХ ... 1215

Стартовая панель ... 1215
Данные ... 1215

Copyright © StatSoft, 1995

Выбор наблюдений (У) ... 1215
Задать вес (В) ... 1215
Продолжить анализ ... 1215
Язык *STATISTICA BASIC* ... 1216
Командный язык (SCL) ... 1216
Быстрые основные статистики ... 1216
Другие виды анализа (Переключатель модулей системы *STATISTICA*) ... 1217
Переключиться в ... 1217
Закрыть и переключиться в ... 1218
Изменить список… ... 1218
Настроить список модулей ... 1218
Добавить ... 1218
Заменить ... 1218
Добавить/Удалить ... 1218

ВЫПАДАЮЩЕЕ МЕНЮ *ГРАФИКА* ТАБЛИЦЫ ИСХОДНЫХ ДАННЫХ ... 1219

ВЫПАДАЮЩЕЕ МЕНЮ *СЕРВИС* ТАБЛИЦЫ ИСХОДНЫХ ДАННЫХ ... 1223

Выбор наблюдений ... 1223
Изменить/добавить условия выбора ... 1223
Примеры ... 1223
Описание синтаксиса ... 1224
Заголовок ... 1224
Состояние ... 1224
Переменные ... 1224
Открыть ... 1224
Сохранить… ... 1224
Выбор переменной для задания веса ... 1225
Вес из переменной ... 1225
Состояние ... 1225
Параметры по умолчанию: общие настройки ... 1225
Стартовые параметры системы *STATISTICA* ... 1225
Очереди окон ... 1226
Связи данных графика (автообновление) ... 1226
Рабочая книга ... 1227
Помнить список ... 1227
Изменение текстовых значений ... 1227
Автоматическое перемещение курсора после нажатия ENTER ... 1228
Метод ”*Перетащить и отпустить*” ... 1228
Многозадачность (при анализе данных) ... 1229
Запрос перед закрытием таблицы результатов ... 1229
Переключатель модулей: одно приложение ... 1229
Процентили ... 1230
Параметры по умолчанию: экран ... 1232
Отображать строку состояния ... 1232

StatSoft®

Copyright © StatSoft, 1995

Полные меню ... 1232
Значки в меню ... 1232
Значки в диалоговых окнах ... 1232
Подсказки для кнопок ... 1232
Кнопки автозадач ... 1232
Объемные диалоговые окна ... 1233
Шрифты ... 1233
Цвета ... 1233
Панели инструментов ... 1233
Текст результатов - Цвета ... 1233
Параметры по умолчанию: принтер ... 1234
Сервис - Макрокоманды ... 1234
Командный язык (*SCL*) ... 1234
Язык *STATISTICA BASIC* ... 1234
Быстрый вызов файла ... 1234
Клавиатурная макрокоманда ... 1234
Макрокоманда записи действий ... 1234
Настройка кнопок автозадач ... 1234

ВЫПАДАЮЩЕЕ МЕНЮ *ОКНО* ТАБЛИЦЫ ИСХОДНЫХ ДАННЫХ ... 1235

Диалоговые окна *Диспетчер таблиц результатов* и *Диспетчер графиков* ... 1235
Очередь ... 1235
Незаблокированные окна ... 1235
Заблокированные окна ... 1235
Блокировать ... 1236
Упорядочить каскадом ... 1236
Упорядочить сверху вниз ... 1236
Упорядочить слева направо ... 1236
Упорядочить значки ... 1237
Закрыть все (незаблокированные) окна, кроме данных ... 1237

ВЫПАДАЮЩЕЕ МЕНЮ *?(СПРАВКА)* ТАБЛИЦЫ ИСХОДНЫХ ДАННЫХ ... 1238

Меню *?(Справка)* – Предметный указатель ... 1238
Меню *?(Справка)* - Таблица с данными ... 1238
Меню *?(Справка)* - Таблица результатов ... 1238
Меню *?(Справка)* - График ... 1238
Меню *?(Справка)* - Текст/вывод ... 1238
Меню *?(Справка)* - Вызов справки… ... 1239
Меню *?(Справка)* - Словарь терминов ... 1239
Меню *?(Справка)* - Клавиатура ... 1239
Меню *?(Справка)* - Советник ... 1239
Меню *?(Справка)* - О программе ... 1240
Ссылка на систему *STATISTICA* ... 1240

ПРЕДМЕТНЫЙ УКАЗАТЕЛЬ ... 1241

Copyright © StatSoft, 1995

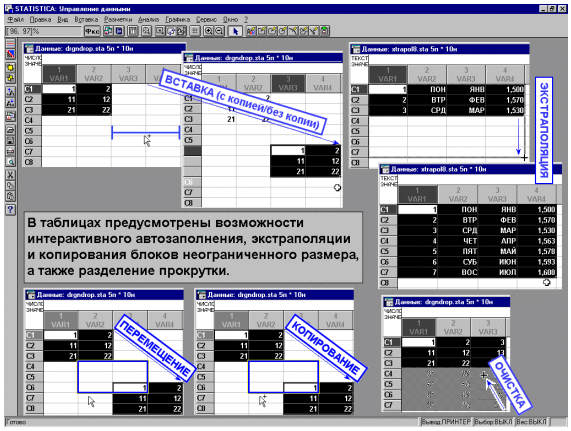
Глава 3:

ОКНО ТАБЛИЦЫ ИСХОДНЫХ ДАННЫХ: МЕНЮ, ДИАЛОГОВЫЕ ОКНА, СВОЙСТВА

ОПЕРАЦИИ С БЛОКАМИ И *АВТОЗАПОЛНЕНИЕ* В ТАБЛИЦАХ ИСХОДНЫХ ДАННЫХ

Возможности метода *Перетащить и отпустить*

Для таблицы исходных данных система *STATISTICA* предоставляет полный набор возможностей метода *Перетащить и отпустить*, которые доступны при работе со стандартными электронными таблицами (например, Excel).



Таблицы результатов поддерживают те же возможности метода *Перетащить и отпустить* за исключением вставки.

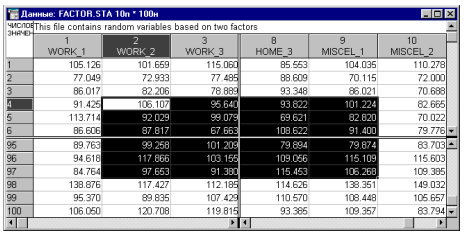
Обратите внимание, что операции, использующие метод *Перетащить и отпустить,* могут быть отключены выбором режима *Отключить все функции метода Перетащить и отпустить* в диалоговом окне *Параметры по умолчанию: общие настройки* системы *STATISTICA* (см. стр. 1225).

Выделение блока

Блоки можно выделять следующими методами: (1) провести курсором мыши, удерживая при этом нажатой левую кнопку, (2) щелкнуть на одном из углов выделяемого блока, переместить курсор на противоположный угол (первая ячейка при этом останется выделенной) и щелкнуть на последней ячейке, нажав клавишу SHIFT (щелкнуть с клавишей SHIFT), или (3) использовать клавиши управления курсором, удерживая при этом нажатой клавишу SHIFT.

Для расширения выделенного ранее блока можно воспользоваться клавишами управления курсором при нажатой клавише SHIFT или прокрутить таблицу курсором мыши и щелкнуть на нужном углу блока, нажав клавишу SHIFT.

Чтобы выделить блок большого размера в режиме разделения окон (как показано ниже), щелкните на ячейке в одном окне, затем прокрутите таблицу до противоположного угла в другом окне и щелкните на нем с клавишей SHIFT.



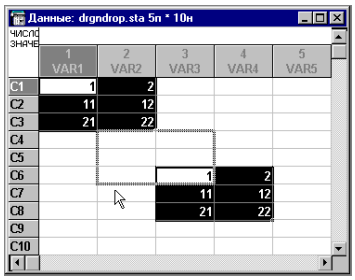
Как только в таблице исходных данных или таблице результатов будет выделен блок, можно будет произвести одну из описанных ниже операций метода *Перетащить и отпустить*.

StatSoft

Copyright © StatSoft, 1995

Перемещение блока

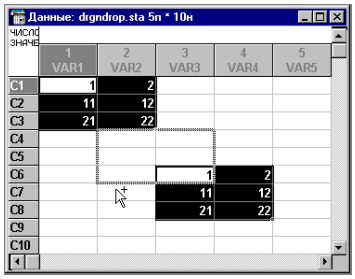
Для перемещения выделенного блока необходимо подвести курсор мыши к границе выделения, и когда он превратится из креста в стрелку, перетащить блок на новое место.



При перемещении блока данных, оставшиеся ячейки заполняются пропущенными значениями. См. также параграфы *Перемещение и копирование значений в таблице исходных данных*.

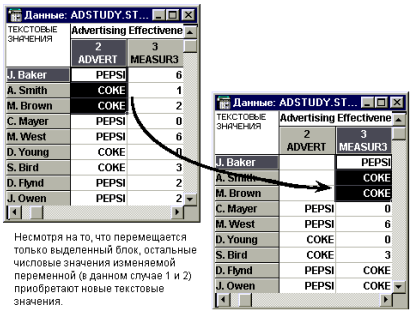
Копирование блока

Для копирования нужно подвести курсор мыши к границе выделения, чтобы он превратился из креста в стрелку, и при перетаскивании копии блока на новое место держать нажатой клавишу CTRL (рядом с курсором мыши появится знак “плюс”).



Действие, по умолчанию происходящее при перетаскивании блока, если не нажата клавиша CTRL, можно задать (изменить) в диалоговом окне *Параметры по умолчанию: общие настройки* системы *STATISTICA* (см. стр. 1225).

Перемещение и копирование значений в таблице исходных данных. При копировании или перемещении блока в таблице исходных данных методом *Перетащить и отпустить*, значения, которые копируются или перемещаются, будут зависеть от режима отображения таблицы исходных данных. Если во время копирования или перемещения блока в таблице представлены числовые значения (или метки значений), то будут скопированы или перемещены только эти числовые значения (или метки значений). Если во время копирования или перемещения блока в таблице представлены текстовые значения, то будут скопированы или перемещены не только текстовые значения, но и соответствующие числовые значения и метки значений (если они есть).



Это может привести к присвоению текстовых значений тем числовым элементам, у которых раньше не было текстовых эквивалентов (как показано на иллюстрации выше).

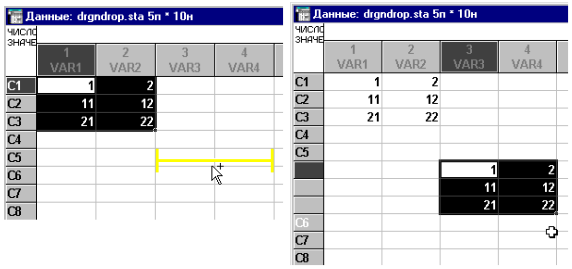
Вставка блока

Можно вставить блок между строками или столбцами, нажав клавишу SHIFT во время перетаскивания блока и поместив курсор мыши между строками или столбцами. Если при этом еще нажать клавишу CTRL, то вместо *перемещения и вставки* будет произведено *копирование и вставка*. Как и при копировании методом *Перетащить и отпустить* рядом со

Copyright © StatSoft, 1995

StatSoft®

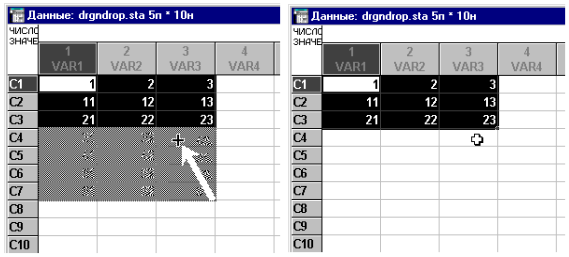
стрелкой курсора мыши появится плюс (см. рисунок ниже).



Обратите внимание, что все операции, производимые методом *Перетащить и отпустить со вставкой* (при нажатой клавише SHIFT) , изменяют размер файла данных (операция *Вставка* невозможна в таблицах результатов); это верно и для простого перетаскивания блоков за границы текущего файла данных. Однако, перетаскивание (без *вставки*) в пределах границ текущего файла данных приводит к результатам, идентичным использованию буфера обмена (только буфер обмена при этом не используется).

Очистка блока

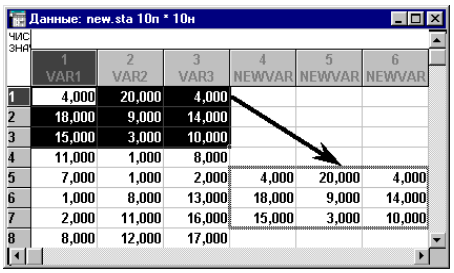
Часть блока (или весь блок) может быть очищена перетаскиванием в пределах выделенного блока.



Выделив блок, переместите курсор в правый нижний угол блока, к инструменту *Ручка заполнения*. Когда он изменится на знак “плюс”, нажмите левую кнопку мыши и потяните внутрь выделенного блока. Во время этой процедуры значения внутри блока станут затененными. Как только вы отпустите левую кнопку мыши, затененные значения будут удалены.

Расширение файла данных методом *Перетащить и отпустить*

При перетаскивании блока за текущие границы таблицы, файл данных расширится таким образом, чтобы появилось место для новых данных.



Таким же образом можно расширять таблицу с помощью операции *Вставка* (см. стр. 1133). Эта операция вставляет блок между наблюдениями или переменными, увеличивая тем самым размер таблицы.

Обратите внимание, что каждый раз, когда вы перемещаете, копируете или вставляете новые наблюдения или переменные, и при этом часть наблюдений или переменных перемещается, копируется или вставляется за текущие границы таблицы, система *STATISTICA* заполняет оставшиеся значения переменных и наблюдений пропущенными значениями (см. выше).

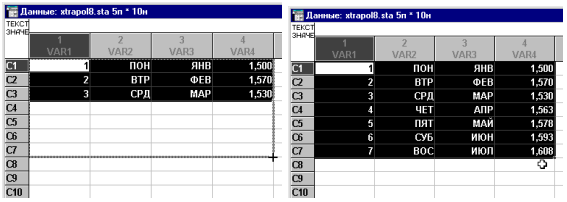
Также обратите внимание, что при работе с большими файлами данных (например, более 10 мегабайт), может понадобиться включения режима *Подтверждать вставку строк или столбцов* в диалоговом окне *Параметры по умолчанию: общие настройки* системы *STATISTICA* (см. стр. 1225).

StatSoft®
Copyright © StatSoft, 1995

Экстраполяция блока (*автозаполнение*)

Горизонтальные или вертикальные ряды значений блока могут быть экстраполированы с помощью перемещения инструмента блока *Ручка заполнения* (маленького квадрата в правом нижнем углу блока). Обратите внимание, что если вы полностью выделяете строку или столбец, то инструмент *Ручка заполнения* будет отображаться и вверху, и внизу столбца или в начале и в конце строки.

Система *STATISTICA* может создавать серии значений, таких как последовательные числа, линейные экстраполяции и даты (например, можно дополнить серию *1*, *2*, *3* так, чтобы она содержала *4*, *5*, *6*, ...).



Ниже перечислены основные правила, используемые при экстраполяции (*автозаполнении*) блока.

- Если исходное выделение содержит повторяющиеся значения, то эти значения будут продублированы в расширенном блоке.
- Неповторяющиеся значения расширяются с помощью линейной регрессии для вычисления экстраполированных значений серий.
- Если блок содержит текстовые значения, то в расширенный блок будет скопировано текстовое значение последней ячейки.
- Если переменная в блоке содержит даты (т.е. в диалоговом окне *текущих спецификаций* для переменной установлен тип формата *Даты*, см. стр. 1172), то система *STATISTICA* экстраполирует даты соответствующим образом, основываясь на шаблоне дат.
- Если переменная в блоке содержит не даты, а текстовые названия месяцев (например, *ЯНВ*, *ФЕВ*, *МАР*, ...), дней (например, *ПОН*, *ВТО*, *СРЕ*, ...), или кварталов (например, *К1*, *К2*, ...), то система *STATISTICA* экстраполирует (в заданном направлении) остальные названия в серии (например, *АПР*, *МАЙ*, ..., *ДЕК*; *ЧЕТ*, *ПЯТ*, *СУБ*, *ВОС*; *К3*, *К4* соответственно). Ниже приведен синтаксис написания этих названий. Обратите внимание, что хотя программа и понимает написание как заглавными, так и прописными буквами, но при записи каждой последовательной серии нужно обязательно использовать один и тот же формат (например, используйте *Янв*, *Фев*, *Мар*, ..., вместо *Янв*, *фев*, *Мар*, ...):

Месяцы. Названия месяцев могут быть записаны следующим образом (обратите внимание, что длина названия ограничена 8 символами:

- *Январь*, *Февраль*, *Март*, *Апрель*, *Май*, *Июнь*, *Июль*, *Август*, *Сентябрь*, *Октябрь*, *Ноябрь*, *Декабрь*
- *Янв*, *Фев*, *Мар*, *Апр*, *Май*, *Июн*, *Июл*, *Авг*, *Сен*, *Окт*, *Ноя*, *Дек*

Дни. Названия дней недели могут быть записаны следующим образом (обратите внимание, что длина названия ограничена 8 символами; поэтому слово “понедельник” будет сокращено до “понедел.”):

- *Понедел.*, *Вторник*, *Среда*, *Четверг*, *Пятница*, *Суббота*, *Воскр.*
- *Пон*, *Вто*, *Сре*, *Чет*, *Пят*, *Суб*, *Вос*
- *Пн*, *Вт*, *Ср*, *Ч*, *Пт*, *Сб*, *Вс*

Copyright © StatSoft, 1995

StatSoft®

Кварталы. Названия кварталов года в переменных могут быть заданы следующим образом:

- *K1*, *K2*, *K3*, *K4*

Экстраполяция (*автозаполнение*) блока влево или вверх. Экстраполяцию влево или вверх можно произвести тем же способом, что и экстраполяцию вправо или вниз, потянув блок за элемент *Ручка заполнения* вверх или влево от начала блока. Обратите внимание, что если вы потянете инструмент *Ручка заполнения* вверх или влево и остановитесь в пределах исходного выделения, не переходя верхнюю или левую границу, то данные внутри выделения будут уничтожены (уничтожаемые данные показываются серым цветом , см. стр. 1134).

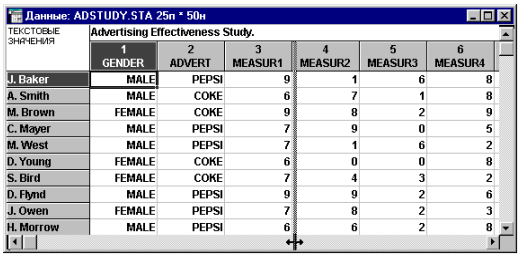
Кроме того, можно воспользоваться командами *Заполнить блок* из контекстного меню или выпадающего меню *Правка*, чтобы скопировать первую строку или первый столбец ячеек блока в смежные ячейки внутри текущего выбранного блока.

Изменение скорости выделения блока

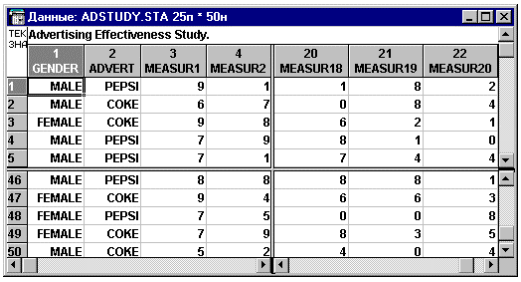
Можно задавать скорость прокрутки таблицы в случае расширения блока за пределы текущего окна. Перемещая курсор на небольшое расстояние от таблицы исходных данных (или таблицы результатов), можно прокручивать таблицу за один раз на одну линию, если блок выделен; переместив курсор на большее расстояние от таблицы исходных данных (или таблицы результатов), можно прокручивать таблицу постранично.

РАЗДЕЛЕНИЕ ПРОКРУТКИ В ТАБЛИЦАХ ИСХОДНЫХ ДАННЫХ И ТАБЛИЦАХ РЕЗУЛЬТАТОВ

Таблицы исходных данных и таблицы результатов могут быть разделены на две или четыре части (окна) путем перетаскивания разделителя (черного прямоугольника в верхней части вертикальной полосы прокрутки или в левой части горизонтальной). Это удобно при наличии большого объема информации для просмотра результатов или данных из различных частей таблицы.



Чтобы разделить таблицу исходных данных или таблицу результатов по вертикали или по горизонтали, поместите курсор мыши на вертикальный или горизонтальный разделитель (черный прямоугольник в верхней части вертикальной полосы прокрутки или в левой части горизонтальной) и потяните курсор-разделитель (↔ или ↕) к строке или столбцу, где вы хотите, чтобы окно разделилось. Можно разделить таблицу исходных данных или таблицу результатов на два (см. выше) или четыре (см. ниже) “окна”.



StatSoft

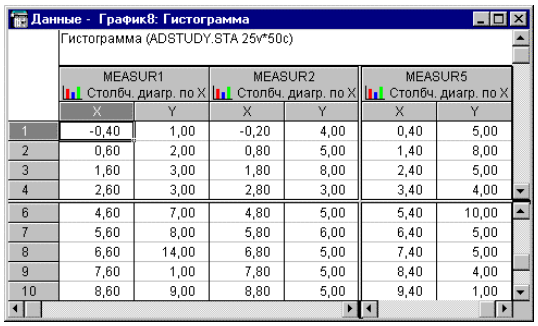
Copyright © StatSoft, 1995

Обратите внимание, что разделенные по вертикали окна прокручиваются синхронно, когда вы прокручиваете таблицу вертикально, а окна, разделенные по горизонтали, прокручиваются синхронно при горизонтальной прокрутке.

Чтобы убрать разделение окон из таблицы исходных данных или таблицы результатов, перетащите разделитель к кнопкам прокрутки на любом конце полосы прокрутки.

Для выделения большого блока в режиме разделения окон, щелкните на ячейке в одном окне, прокрутите таблицу до противоположного угла и щелкните с клавишей SHIFT (см. ниже параграф *Выделение блока*). Далее для выбранных блоков можно использовать операции с блоками системы *STATISTICA* (т.е. метод *Перетащить и отпустить*, см. ниже, и команды *Блоковые статистики/Столбцы-/Строки/Графики*, см. стр. 1202).

Разделение окон также возможно в окне *Редактор данных графика*,



в таблицах *Менеджера мегафайлов* и других таблицах результатов.

Copyright © StatSoft, 1995

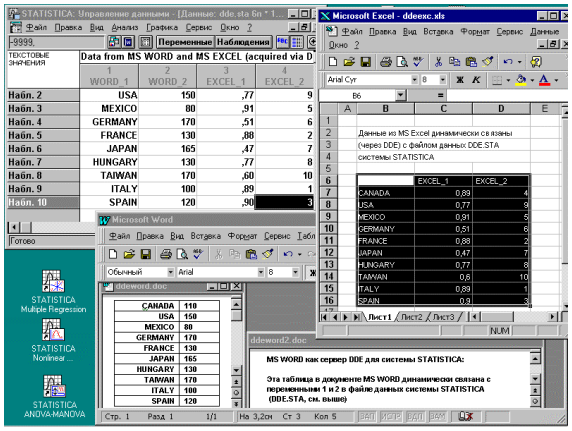
ДИНАМИЧЕСКИЙ ОБМЕН ДАННЫМИ (*DDE*)

Введение

Система *STATISTICA* поддерживает соглашения *динамического обмена данными* (*DDE*). Это означает, что данные из таблиц можно динамически присоединять к данным из других приложений (Windows).

Создание связей

Можно установить связь *DDE* (динамического обмена данными) между файлом "источника" (файлом-сервером) (например, таблицей Excel) и файлом данных системы *STATISTICA* (файлом-клиентом), таким образом, что при изменении данных в источнике автоматически производятся изменения данных в таблице системы *STATISTICA* (клиенте).



В качестве простейшего использования двух динамически связанных файлов может быть рассмотрен пример с производственными установками, когда файл данных системы *STATISTICA* будет динамически связан с измерительным устройством (например, для ежечасного автоматического обновления определенных измерений).

Связи *DDE* могут быть установлены с помощью быстрой команды *Установить связь* выпадающего меню *Правка* (см. ниже), или с помощью описания связи в поле *Длинное имя (метка, связь или формула)* в диалоговом окне *текущих спецификаций* или в окне *Диспетчер связей* (см. стр. 1206). Также связи *DDE* можно использовать вместе с *Командным языком* системы *STATISTICA* (через *Модуль запуска*, см. ниже.) для работы с результатами системы *STATISTICA* в других приложениях Windows (например, Excel). Можно динамически связывать файл-источник с файлом данных системы *STATISTICA* тремя различными способами, описанными в следующих разделах.

Установить связь

Эта команда выпадающего меню *Правка* (см. *Установить связь*, стр. 1197) является наиболее быстрым и простым способом установить связь *DDE* .

Шаг 1. Откройте источник и скопируйте блок данных (или текст), который вы хотите связать с файлом-клиентом (из системы *STATISTICA*), в буфер обмена. Далее, не закрывая файл-сервер, переключитесь в нужный модуль системы *STATISTICA*.

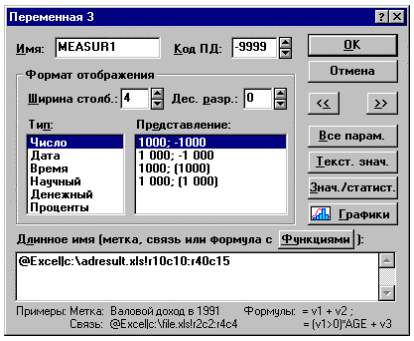
Шаг 2. В таблице системы *STATISTICA* отметьте (выделите) ячейку, к которой будут "прикреплены" данные, вставляемые в таблицу (данные будут вставлены вниз и вправо от этой ячейки).

Шаг 3. Выберите пункт *Установить связь* для установления связи. Система *STATISTICA* автоматически отметит путь *DDE* (см. ниже) и сохранит его в *длинном имени переменной* (см. ниже) и в окне *Диспетчер связей* (см. стр. 1206).

StatSoft®
Copyright © StatSoft, 1995

Длинное имя переменой

Связи *DDE* сохраняются и могут быть установлены (и изменены) в поле *Длинное имя* в диалоговом окне *текущих спецификаций*.



Чтобы установить здесь связь *DDE*, впишите оператор связи (путь *DDE*). Части оператора связи (например, *@Excel|c:\Excel\File.xls!r2c2:r4c4;c1;r1*) описаны ниже.

(1) @Название сервера| . Сначала надо задать приложение-источник (сервер) (в приведенном примере это Excel). *Название сервера* должно отделяться от остальных частей оператора связи символом "|". Также надо определить, является ли связь *DDE* "горячей" (поставьте символ "@" перед *названием сервера* для указания того, что файл-клиент будет обновляться каждый раз, когда в файл-сервер будут вноситься изменения) или "холодной" (поставьте символ "^" перед *названием сервера* для указания того, что файл-клиент будет обновляться только по запросу пользователя или при изменении связи на "горячую").

(2) Путь! . *Путь* обозначает имя (и путь) файла-источника и должен отделяться от остальных частей оператора символом "!" (убедитесь, что задан полный путь к файлу). В приведенном примере *путь* (*c:\Excel\File.xls!*) обозначает не только название файла Excel (*File.xls*), но и устройство и каталог, в котором находится этот файл (например, *c:\Excel*). Определение файла-источника зависит от приложения, которое вы используете (обратитесь к документации к этому приложению, чтобы узнать, как задать *путь*. Если используется команда *Установить связь* (см. выше), то он будет задан автоматически).

(3) Координаты ячеек; . Здесь задается ячейка, блок или имя, заданное для приложения-источника, затем ставится символ ";" (чтобы отделить их от остальных частей пути *DDE*). То, что задается здесь, зависит от приложения-источника (обратитесь к документации к этому приложению, чтобы узнать, как задать координаты ячеек).

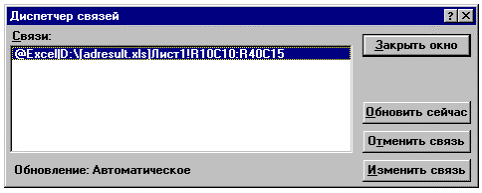
При использовании команды *Установить связь* (см. выше) они будут заданы автоматически. В приведенном примере задан блок таблицы Excel (например, *r2c2:r4c4;* задает блок "от строки 2, столбца 2 до строки 4, столбца 4").

(4) Столбец;строка . Эта часть пути *DDE* не является обязательной; по умолчанию (если не задано начального столбца или строки) система *STATISTICA* помещает первое значение в блоке привязанных данных в первую ячейку (строку) выделенной переменной в таблице и помещает остальные значения снизу и справа от этой "закрепляющей" ячейки. Если вы хотите "прикрепить" привязанные данные к другой ячейке (области) таблицы, можно задать номер столбца и номер строки (разделив их при помощи символа ";") для определения позиции этой ячейки. Кроме того, можно задать только номер столбца (переменную) и система *STATISTICA* использует первую ячейку (строку) в этом столбце как "закрепляющую". Поскольку эта часть пути *DDE* не является обязательной, она не будет автоматически определена при использовании команды *Установить связь* (см. выше).

Copyright © StatSoft, 1995

Диспетчер связей

Когда связь установлена, ею можно управлять с помощью диалогового окна *Диспетчер связей*, доступного из пункта *Связи...* выпадающего меню *Правка* для таблицы исходных данных.



Связи. Это поле отображает все связи *DDE* системы *STATISTICA*. Можно изменить связь, выделив ее в этом окне и нажав кнопку *Изменить связь* (см. ниже).

Обновление. Отображает состояние связи *DDE*, выделенной в поле *Связи*. Связи *DDE* могут быть "горячими" (начинаются с символа "@", что означает *автоматическое* обновление) или "холодными" (начинаются с символа"^", что означает *ручное* обновление). В *автоматическом* режиме присоединенные данные в таблице системы *STATISTICA* (файле-клиенте) будут обновляться автоматически при внесении изменений в данные приложения-сервера. В *ручном* режиме связь *DDE* временно отключена, и файл-клиент нужно обновлять вручную с помощью кнопки *Обновить сейчас* или изменить состояние на "горячее" (с помощью кнопки *Изменить связь*).

Обновить сейчас. Нажмите эту кнопку, чтобы вручную обновить связь *DDE*, которая выделена в поле *Связи* в файле-клиенте (таблице системы *STATISTICA*).

Отменить связь. Нажмите эту кнопку, чтобы временно отключить выделенную связь *DDE*. ПРИМЕЧАНИЕ: Если вы собираетесь произвести анализ привязанного файла данных, убедитесь, что связь временно отключена, чтобы данные не могли измениться в ходе анализа. Можно восстановить связь, нажав кнопку *Обновить сейчас*.

Изменить связь. Нажмите эту кнопку, чтобы открыть диалоговое окно *текущих спецификаций,* в котором можно изменить связь *DDE* с помощью параметра *Длинное имя-связь* (см. раздел *Создание связей (динамический обмен данными)*) для получения подробной информации о пути *DDE*).

В системе *STATISTICA* существуют также и другие приложения для использования возможностей динамического обмена данными (см. *Электронное руководство* по *DDE* для получения подробной информации).

Связи *DDE*, *Командный язык* системы *STATISTICA* и *Модуль запуска*

Можно вставлять команды *Командного языка* системы *STATISTICA* (*SCL*) или целые программы на этом языке в создаваемые макрокоманды, а затем запускать их из других приложений Windows (например, Excel, MS Word, Ami Pro, Quattro Pro). См. ниже пример макрокоманды Excel, которая использует команды *SCL* и *Модуль запуска* системы *STATISTICA* для расчета некоторых описательных статистик.

```
Excel macro DDE_EX.XLM:
channel=INITIATE("STA_RUN","SYSTEM")
=EXECUTE(channel, "DESC VARS=ALL")
=TERMINATE(channel)
=RETURN()
```

Строка 1. Открывает канал *DDE* для системы *STATISTICA* (конкретно *Модуль запуска*). Обратите внимание, что для инициализации канала *DDE* система *STATISTICA* должна находиться в пути поиска, или должен быть запущен файл *Sta_run.exe.*

Строка 2. Заставляет систему *STATISTICA* выполнить команду *SCL* "*DESC VARS=ALL*" (т.е.

StatSoft®

Copyright © StatSoft, 1995

рассчитать описательные статистики всех переменных).

Строка 3. Заставляет Excel прекратить обмен данными *DDE*.

Строка 4. Возврат в приложение-источник.

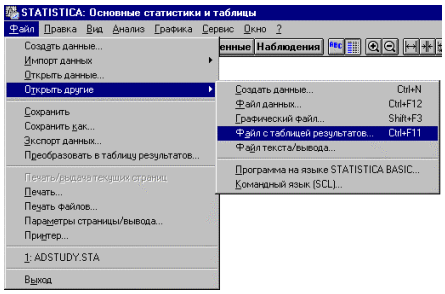
Этот простой пример показывает, как можно открыть модуль системы *STATISTICA*, запустить программу и вернуться в приложение Windows. Кроме того, можно вставить команды для копирования результатов из системы *STATISTICA* в виде текста и/или графических изображений в буфер обмена и использовать их в документе приложения Windows (например, как часть доклада, сделанного в MS Word).

Примечание. С помощью окна *Диспетчер связей* (см. выше) связи могут быть временно отключены, чтобы провести анализ привязанного файла данных. Если вы не отключите эту связь, а данные будут обновлены системой *STATISTICA* во время анализа, то, возможно, результат не будет иметь смысла.

Copyright © StatSoft, 1995

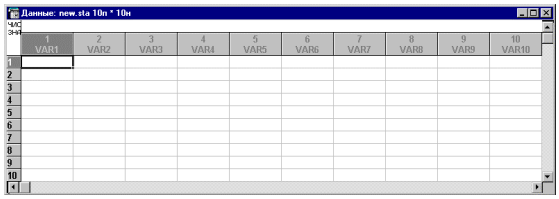
ВЫПАДАЮЩЕЕ МЕНЮ *ФАЙЛ* ТАБЛИЦЫ ИСХОДНЫХ ДАННЫХ

Из выпадающего меню *Файл* таблицы исходных данных доступны описанные ниже диалоговые окна и команды.



Создать данные

Этот пункт выпадающего меню *Файл* позволяет открыть новый файл для ввода данных. При выборе этого пункта откроется диалоговое окно *Создание данных*: *имя файла* (см. ниже), позволяя вам выбрать имя, под которым будет сохранена новая таблица. После выбора имени файла будет создана таблица размером 10 переменных на 10 наблюдений.

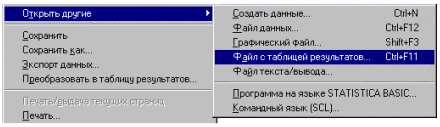


В эту таблицу можно вводить данные, добавлять и удалять наблюдения и переменные, изменять имена и спецификации переменных (см. раздел *Изменение спецификаций (текущей) таблицы*, стр. 1170) или имена наблюдений (см. раздел *Изменение имен наблюдений таблицы*, стр. 1191), а также производить другие операции (см. раздел *Меню Правка таблицы исходных данных*, стр. 1167).

Кроме того, в модуле *Управление данными* можно создать новый файл данных в соответствии с вашими подробными спецификациями [заданным числом переменных и наблюдений, автоматическим созданием определенного числа последовательных переменных, с параметрами по умолчанию (например, форматом) или с полным набором спецификаций, полученным из другого файла данных (включая присвоенные числовые и текстовые значения и метки значений)] (см. главу 7).

Открыть другие...

Это подменю является общим для всех меню *Файл* (для таблиц исходных данных, таблиц результатов, графиков, *Окна текста/вывода*) и предоставляет доступ ко все типам файлов системы *STATISTICA*.



Эти команды позволяют *создать данные* (стр. 1142), открыть *файл данных* (стр. 1143), *графический файл* (стр. 1143), *файл с таблицей результатов* (стр. 1143), *файл текста/вывода* (стр. 1144), *программу на языке STATISTICA BASIC* или на *Командном языке (SCL)* (см. *Электронное руководство* для получения более подробной информации об двух последних пунктах). Если вы выберете пункт *Создать данные*, откроется диалоговое окно *Создание данных: имя файла* (стр. 1144), в котором можно указать имя файла для сохранения новой таблицы. После выбора имени файла будет создана таблица размером 10 переменных на 10 наблюдений.

StatSoft®
Copyright © StatSoft, 1995

Выбор любого другого пункта из перечисленных выше приведет к открытию соответствующего диалогового окна *Открытие файла* (стр.1144), в котором можно выбрать файл, который нужно открыть.

Открыть файл данных

Выбор этого пункта из выпадающего меню *Файл* приведет к открытию диалогового окна *Открытие файла данных* (см. выше), в котором можно выбрать файл данных системы *STATISTICA* (с расширением **.sta*). Файлы данных системы *STATISTICA* содержат значения, метки и символы форматирования (см. раздел *Изменение (текущих) спецификаций таблицы исходных данных*, стр.1170), заголовки и примечания (см. раздел *Заголовок файла*, стр.1192), а также другую информацию. Чтобы открыть файл данных, сохраненный в формате, отличном от формата системы *STATISTICA* (например, Excel, Lotus, Quattro, dBASE, Paradox, ASCII, и других), воспользуйтесь пунктом *Импорт данных* из того же меню (см. стр. 1146).

Как открыть несколько файлов данных. В модуле может быть открыт только один файл данных за один раз. В то же время можно открыть несколько файлов (например, чтобы копировать данные между ними) в различных модулях или с помощью приложения *Сервер файлов* системы *STATISTICA*. Можно также одновременно запустить несколько модулей *Сервер файлов* и в каждом из них открыть свой файл данных.

Кнопка на панели инструментов. См. также кнопку панели инструментов таблицы исходных данных.

Комбинация клавиш. CTRL+O

Открыть графический файл

Выбор этого пункта из выпадающего меню *Файл* приведет к открытию диалогового окна *Открытие графического файла* (см. ниже), в котором можно выбрать графический файл системы *STATISTICA* (с расширением **.stg*). Эти файлы в формате *График системы STATISTICA* содержат не только графики, но и пользовательские объекты (например, рисунки, внедренные объекты и т.д.), а также значения, представленные на графике. После того как будет открыт графический файл, данные могут быть исправлены или использованы для других целей.

Графические файлы системы *STATISTICA* являются объектами *OLE*, поэтому вы не можете одновременно открыть один и тот же объект несколько раз (даже если использовать для этого разные модули). Поэтому при попытке получить доступ к графическому файлу, который уже открыт в каком-либо модуле, будет открыта его копия, и это будет записано в его заголовке (например, "*Копия Hist.stg: ...*").

Комбинация клавиш. SHIFT+F3

Открыть файл таблицы результатов

Выбор этого пункта из выпадающего меню *Файл* приведет к открытию диалогового окна *Открытие файла с таблицей результатов* (см. ниже), в котором можно выбрать файл, содержащий таблицу результатов (с расширением **.scr*). Эти файлы содержат результаты вычислений, информацию о форматировании и пользовательскую информацию (например, параметры выделения, цветов и т.д.).

Комбинация клавиш. CTRL+F11

Открыть файл текста/вывода

Выбор этого пункта из выпадающего меню *Файл* приведет к открытию диалогового окна *Открытие текстового файла* (см. ниже), в котором можно выбрать текстовый файл (с расширением **.txt*) или файл в формате *RTF* (с

Copyright © StatSoft, 1995

расширением **.rtf*), который вы хотите открыть. Выбранный здесь текстовый файл будет открыт в *Окне текста/вывода* (см. стр. 1211).

Открыть программу на языке *STATISTICA BASIC*

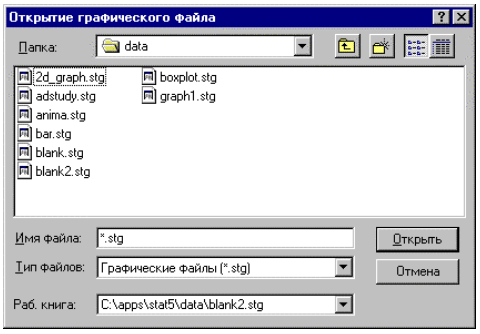
Выбор этого пункта из выпадающего меню *Файл* приведет к открытию диалогового окна *Открытие файла STATISTICA BASIC* (см. ниже), в котором можно выбрать файл программы на языке *STATISTICA BASIC* (с расширением **.stb*), который вы хотите открыть. Выбранный здесь файл будет открыт в *Окне STATISTICA BASIC* (подробную информацию см. в *Электронном руководстве*).

Открыть файл командного языка (*SCL*)

Выбор этого пункта из выпадающего меню *Файл* приведет к открытию диалогового окна *Открытие файла SCL* (см. ниже), в котором можно выбрать файл командного языка (*SCL*) (с расширением **.scl*). Выбранный здесь файл будет открыт в окне *Командный язык SCL* (подробную информацию см. в *Электронном руководстве*).

Диалоговое окно *Открытие/Сохранение файла*

С помощью этого диалогового окна можно выбрать и открыть (команды *Создать данные*, *Открыть данные*, *Открыть другие*) или сохранить (команды *Сохранить*, *Сохранить как*) новую таблицу исходных данных, таблицу результатов, график или файл текста/вывода.



Точно такое же диалоговое окно открывается, когда необходимо выбрать имя файла, который нужно открыть или сохранить. Каждый из перечисленных выше пунктов выпадающего меню открывает соответствующее диалоговое окно *Открытие/Сохранение файла* со следующими параметрами.

Открыть файлы. Файлы системы *STATISTICA* можно открыть одним из перечисленных ниже способов.

- Открыть диалоговое окно *Открытие/Сохранение файла* (описанное ниже), выбрав соответствующий пункт *Открыть* из выпадающего меню *Файл* или нажав кнопку панели инструментов.
- Дважды щелкнуть на имени файла в окне *Диспетчер файлов*, чтобы открыть файл в модуле *Сервер файлов* системы *STATISTICA*.
- Перетащить файл из окна *Диспетчер файлов (File Manager)* Windows и отпустить его на значке модуля системы *STATISTICA,* чтобы открыть этот модуль с нужным файлом.
- Установить для нужных файлов режим *Автооткрытие* в диалоговых окнах *Заголовок файла данных, примечания и Рабочая книга*. Когда будет открыт файл данных, эти файлы откроются автоматически.

StatSoft

Copyright © StatSoft, 1995

Имя файла

Здесь можно вписать имя файла или выбрать его из списка. Основные типы файлов, которые используются системой *STATISTICA,* имеют следующие расширения (см. *Расширения имен файлов* системы *STATISTICA* в *Электронном руководстве* для получения информации об обновленном списке расширений имен файлов):

- Файлы данных таблиц исходных данных системы *STATISTICA*, которые включают данные, имена переменных, длинные метки, формулы и связи, имена наблюдений, текстовые значения, спецификации переменных, примечания и другую информацию); расширение имени файла: ***.sta*.
- Окна вывода таблиц результатов системы *STATISTICA*, которые включают форматированную таблицу вывода; расширение имени файла: ***.scr*.
- Графические файлы, содержащие графики, параметры пользователя, внедренные графические файлы, связи и графические данные; расширение имени файла: ***.stg*.
- Текст (текстовые файлы, созданные или используемые в *Окне текста/вывода*, а также данные *Редактора текста графика* в текстовом виде); расширение имени файла: **.txt*.
- Файлы в расширенном текстовом формате (*RTF),* например, созданные или используемые в *Окне текста/вывода*; расширение имени файла: **.rtf.*
- файлы языка *STATISTICA BASIC;* расширение имени файла: **.stb*.
- файлы *Менеджера мегафайлов*; расширение имени файла: **.mfm*.

Список файлов по типу

В поле *Имя файла* выберите вывод всех файлов определенного типа (т.е. таблицы исходных данных/таблицы результатов/графические файлы, резервные файлы, файлы *Менеджера мегафайлов* или все файлы). Чтобы открыть файл данных, сохраненный в формате, отличном от формата системы *STATISTICA* (например, Excel, Lotus, Quattro, dBASE, Paradox, ASCII и другие), воспользуйтесь пунктом *Импорт данных* того же меню (см. ниже).

Папки

Укажите имя *папки*, в которой находится нужный файл.

Диски

Укажите имя *диска*, на котором находится нужный файл.

Рабочая книга

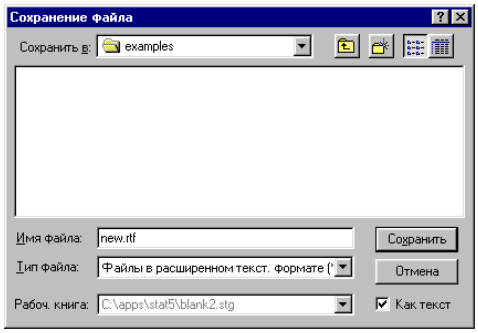
Нажмите в этом поле кнопку с прокручиваемым списком для выбора дополнительных файлов (например, графических файлов, файлов **.ini*, файлов импорта/экспорта, файлов таблиц результатов и т.д.), связанных с текущим файлом данных (список всех файлов, связанных с текущим файлом данных, хранится в *Рабочей книге* системы *STATISTICA*). В целях легкого доступа к нужному файлу, этот список отсортирован по типу файлов (например, файлы **.scr*), который вы хотите открыть или сохранить. Файлы, не относящиеся к текущей операции открытия/сохранения, будут затенены (они включены в список только для справки).

Как текст

Этот режим применим только в *Окне текста/вывода* [включая файлы программ языка *STATISTICA BASIC*, файлы программ *Командного языка* (*SCL*), командные файлы

Copyright © StatSoft, 1995

модуля *Структурное моделирование* и другие приложения *Окна текста/вывода*].



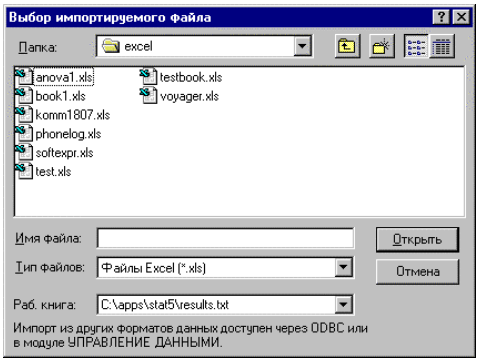
Выберите этот режим для сохранения содержания *Окна текста/вывода* в стандартном формате ASCII текста (расширение имени файла **.txt*). Если этот режим *не* выбран, система *STATISTICA* запишет файл в формате *RTF,* сохраняя все специальное форматирование (шрифт, расстояния, графики и т.д.). Формат по умолчанию (*RTF*) обычно предпочтительнее, но в этом случае файл результатов не может быть изменен с помощью программ, не поддерживающих формат *RTF*. В то же время при сохранении результатов в виде текста, этот файл может быть изменен с помощью любой программы, но при этом вся информация о форматировании и графиках будет утеряна.

Импорт данных – Быстрый импорт

Импорт данных в систему *STATISTICA* путем преобразования различных файлов (например, электронных таблиц или баз данных) является одной из нескольких возможностей, предлагаемых системой *STATISTICA* для доступа к данным из других приложений. Другие методы включают установление динамических связей (*DDE*) между файлами данных и использование буфера обмена (который в системе *STATISTICA* поддерживает соглашения о специальном форматировании в буфере обмена, используемые в других приложениях Windows, таких как Excel и Lotus).

Эта команда позволяет быстро импортировать файлы из электронных таблиц и баз данных, а также из текстовых (ASCII) файлов. Система *STATISTICA* также предоставляет доступ к большому количеству форматов баз данных через средства *ODBC* и дополнительный набор специализированных возможностей импорта данных с помощью модуля *Управления данными*. Этот модуль поддерживает большое количество форматов файлов данных и предоставляет дополнительные возможности для специального перемещения данных (например, возможности разделять/объединять значения и их метки во время перемещения файлов между форматами, которые не поддерживают значения с метками). Команды *Быстрый импорт* и *ODBC* доступны из выпадающего меню *Файл* любого модуля системы *STATISTICA*.

Как только вы выберете пункт *Быстрый импорт* выпадающего меню *Файл* и укажете нужный файл для импорта,



откроется диалоговое окно *Быстрый импорт из…*

Как импортировать текстовые значения длиной более 8 символов. Если текстовые значения, импортированные с помощью команды *Быстрый импорт*, имеют длину более 8 символов, то для сохранения длинных текстовых

StatSoft®

Copyright © StatSoft, 1995

значений полностью будут созданы метки значений. При этом первые 8 символов каждого значения будут использованы для создания соответствующих текстовых значений системы *STATISTICA*. Обратите внимание, что если в результате будут созданы текстовые значения, не соответствующие однозначно длинным меткам (т.е. возникнет ситуация, когда одинаковое текстовое значение в разных случаях будет сопровождаться разными метками), то последние символы текстовых значений будут заменены цифрами для создания уникальных текстовых значений для соответствующих меток.

Например, если значения в исходном (импортируемом) наборе данных таковы:

```
Arizona, the south region
Arizona, the north region
Arizona, the central part,
```

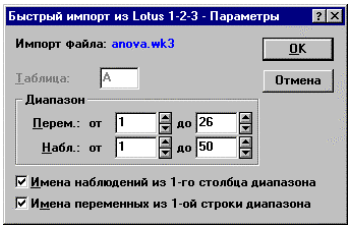
то эти значения будут сохранены как метки значений, а соответствующие (уникальные) 8-символьные текстовые значения будут созданы следующим образом:

```
Arizona1
Arizona2
Arizona3.
```

Те же соглашения действуют при импорте данных в таблицу исходных данных системы *STATISTICA* через буфер обмена [например, при копировании блока значений из другой таблицы (путем выделения блока в таблице и нажатия CTRL+C) и вставке его в таблицу системы *STATISTICA* (путем установки курсора на нужное место и нажатия CTRL+V)]. Если текстовые значения длиннее 8 символов, то они будут восприняты системой *STATISTICA* как метки значений, и из первых 8 символов каждого длинного значения будут созданы соответствующие короткие текстовые значения (как показано выше).

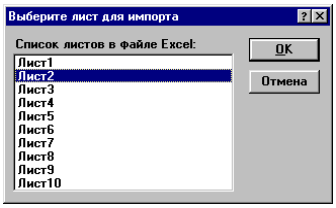
Быстрый импорт - Таблицы

При импорте файлов Lotus, Excel или Quattro в этом диалоговом окне доступны показанные ниже параметры. Обратите внимание, что если выбранный файл содержит несколько таблиц, то система *STATISTICA* выведет список, из которого можно выбрать нужную таблицу.



Импорт файла. Здесь указано имя импортируемого файла.

Таблица. Выберите, какую таблицу импортировать. Этот параметр доступен только при импорте файла программы Lotus 1-2-3 (версии 3 или позднее) или файла программы Quattro. При импорте файла из Excel выбор таблицы будет осуществляться из диалогового окна позднее.



Диапазон. Этот параметр позволяет выбрать диапазон импортируемых переменных или наблюдений.

Перем. Укажите диапазон импортируемых переменных (столбцов) (*от* какой переменной и *до* какой переменной включительно).

Набл. Укажите диапазон импортируемых наблюдений (строк) (*от* какой строки и *до* какой строки включительно).

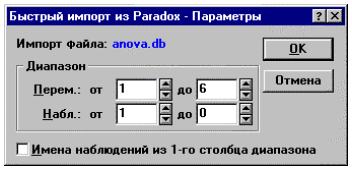
Copyright © StatSoft, 1995

Имена наблюдений из 1-го столбца диапазона. Выберите этот режим, чтобы импортировать имена наблюдений из первого столбца указанного диапазона (см. выше). Если это поле содержит более 20 символов, то только первые 20 будут использоваться как имя наблюдения. Если этот столбец содержит не текстовые, а числовые значения, то именами наблюдений будут эти значения в тестовом виде.

Имена переменных из 1-ой строки диапазона. Выберите этот режим для того, чтобы использовать метки из первой строки указанного диапазона импорта (см. выше). Если какая-либо из меток окажется длиннее восьми символов, она будет урезана и для имени переменной будут использованы только первые восемь символов.

Быстрый импорт – Файлы баз данных

При импорте файлов из dBASE, Paradox или Access в этом диалоговом окне доступны следующие параметры.



Импорт файла. Здесь указано имя импортируемого файла.

Диапазон. Этот параметр позволяет выбрать диапазон импортируемых переменных или наблюдений.

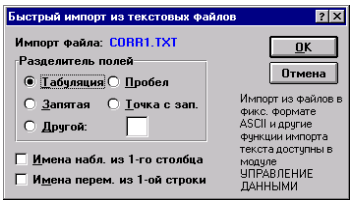
Перем. Укажите диапазон импортируемых переменных (столбцов) (*от* какой переменной и *до* какой переменной включительно).

Набл. Укажите диапазон импортируемых наблюдений (строк) (*от* какой строки и *до* какой строки включительно).

Имена наблюдений из 1-го столбца диапазона. Выберите этот режим, чтобы произвести импорт имен наблюдений из первого столбца указанного диапазона (см. выше). Если это поле содержит более 20 символов, то только первые 20 будут использоваться как имя наблюдения. Если этот столбец содержит не текстовые, а числовые значения, то именами наблюдений будут эти значения в тестовом виде.

Быстрый импорт – Текстовые файлы

При импорте текстового файла в этом диалоговом окне доступны описанные ниже возможности.



Если файл в систему *STATISTICA* импортируется с помощью команды *Быстрый импорт*, то в нем будет создано столько же строк, сколько их было в ASCII файле. Программа также осуществит поиск самой длинной записи в ASCII файле. Длина этой записи будет определять число переменных в файле данных системы *STATISTICA*. Все более короткие записи в ASCII файле будут дополнены значениями пропущенных данных (до размера самой длинной записи).

Импорт файла. Здесь указано имя импортируемого файла.

Разделитель полей. Укажите символ, который используется в импортируемом файле как разделитель. Можно выбрать тип разделителя из четырех заданных (*табуляция*, *пробел*, *запятая* или *точка с запятой*) или задать разделитель с помощью поля ввода *Другой*. Для более подробной информации о

StatSoft®
Copyright © StatSoft, 1995

разделителях полей см. *Импорт файлов в свободном ASCII формате* в модуле *Управление данными*.

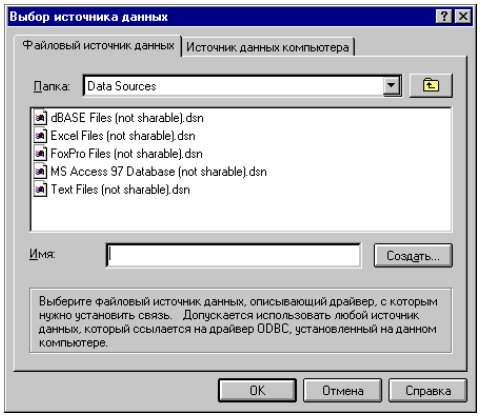
Имена наблюдений из 1-го столбца диапазона. Выберите этот режим, чтобы произвести импорт имен наблюдений из первого столбца указанного диапазона (см. выше). Если это поле содержит более 20 символов, то только первые 20 будут использоваться как имя наблюдения. Если этот столбец содержит не текстовые, а числовые значения, то именами наблюдений будут эти значения в тестовом виде.

Имена переменных из 1-ой строки диапазона. Выберите этот режим для того, чтобы использовать метки из первой строки указанного диапазона импорта (см. выше). Если какая-либо из меток окажется длиннее восьми символов, она будет урезана и для имени переменной будут использованы только первые восемь символов.

Импорт - *ODBC*

Эта команда позволяет легко импортировать данные из большого количества баз данных [включая многие большие системы баз данных (Oracle, Sybase, и т.д.) с помощью соглашений Microsoft Open Data Base Connectivity (*ODBC*)]. В системе *STATISTICA* предусмотрены следующие возможности импорта *ODBC:* поддержка нескольких таблиц баз данных; выделение определенных записей (строк и столбцов) вводом операторов запроса *SQL*. Ниже перечислены шаги, которые необходимо предпринять для импорта данных с помощью *ODBC*.

- При выборе пункта *Импорт данных - ODBC* выпадающего меню *Файл* любого модуля укажите в диалоговом окне *Источники данных SQL* источник данных (имя сервера или папки, в которой расположены данные, например, файл dBASE).



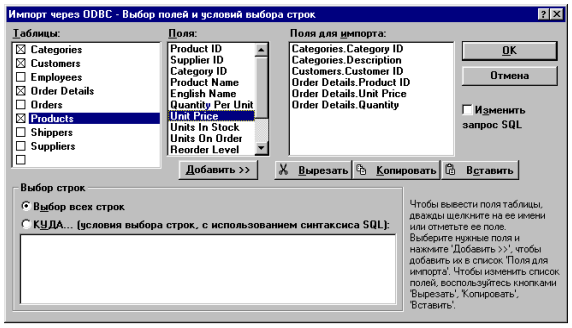
- Далее, укажите имя и расположение файла, содержащего данные, в стандартном диалоговом окне *Выберите папку*.
- Откроется диалоговое окно *Импорт через ODBC – Выбор полей и условий выбора строк*, где можно выбрать таблицы, поля, строки и т.д., которые вы хотите импортировать.

Обратите внимание, что для того, чтобы использовать этот режим, в системе должны быть установлены драйверы *ODBC* (соответствующие формату импортируемых данных). Список установленных драйверов (включая те, которые были установлены системой *STATISTICA*) отображается в диалоговом окне *Источники данных SQL*. Если драйвер для базы данных, которую вы собираетесь импортировать, отсутствует в списке, пожалуйста, свяжитесь с производителем соответствующей базы данных.

Выбор полей и условий для выбора строк

Это диалоговое окно откроется при выборе пункта *Импорт - ODBC* из выпадающего меню *Файл* и указании источника данных. Здесь можно выбрать таблицы, поля, строки и т.д., которые вы хотите импортировать.

Copyright © StatSoft, 1995



Таблицы. Эта часть диалогового окна *Импорт через ODBC – Выбор полей и условий выбора строк* отображает все таблицы в ранее выбранной папке базы данных. Нажмите кнопку ☒ рядом с нужной таблицей, чтобы открыть (активизировать) таблицу и показать названия ее *полей* - (столбцов). Кроме того, для переключения между доступными таблицами можно дважды щелкнуть на названии *таблицы* или нажать клавишу пробела.

Если в папке доступна только одна таблица, то при открытии этого диалогового окна она будет автоматически активизирована, а ее *поля* будут отображены на экране.

Поля. Как только вы выберете нужную *таблицу* в диалоговом окне *Импорт через ODBC – Выбор полей и условий выбора строк*, в этой части диалогового окна будут отображены названия *полей* (столбцов) этой таблицы. Выберите нужные поля, выделяя их [непрерывно или по очереди (удерживая нажатой клавишу CTRL и щелкая на именах)] в этом окне и нажимая кнопку *Добавить>>* для перемещения выделенных имен полей в окно *Поля для импорта*, где их можно обрабатывать.

Поля для импорта. Эта часть диалогового окна отображает импортируемые поля выбранных *таблиц*. Здесь можно обрабатывать ранее выбранные *поля* (столбцы): можно удалять, вырезать, копировать и/или вставлять названия полей, изменяя список. Обратите внимание, что в результате в файле данных системы *STATISTICA* поля будут располагаться в том порядке, в котором они отображаются в данном списке.

Выбор строк. Укажите здесь либо все строки (*Выделить все строки*), либо несколько строк (*КУДА...*) для импорта. Если выбрать пункт *Выделить все строки*, то при импорте данных с помощью *ODBC* система *STATISTICA* создаст оператор *ВЫБОР* для импорта всех указанных пользователем полей из соответствующих таблиц. Если вы хотите импортировать только часть строк, выберите пункт *КУДА...* и правильно напишите оператор *SQL КУДА (WHERE...)*. См. информацию о синтаксисе в руководстве по *SQL*.

Сохранить таблицу исходных данных

С помощью этого пункта выпадающего меню *Файл* можно сохранить таблицу исходных данных в файле. Если файл до этого не был записан, то откроется диалоговое окно *Сохранение файла* (стр. 1144), в котором можно указать имя сохраняемого файла и место его расположения на диске. Если имя файла и место для его сохранения указаны, то при нажатии на кнопку *Сохранить*, таблица исходных данных будет автоматически сохранена, заменив предыдущую копию файла. Если вы хотите изменить имя или место для сохранения файла, выберите пункт *Сохранить как*....

Кнопка на панели инструментов. См. также кнопку 💾 панели инструментов окна таблицы исходных данных.

Комбинация клавиш. CTRL+S

Сохранение таблицы исходных данных под другим именем

При каждом выборе пункта *Сохранить как*... открывается окно *Сохранение файла*... (стр.

StatSoft®

Copyright © StatSoft, 1995

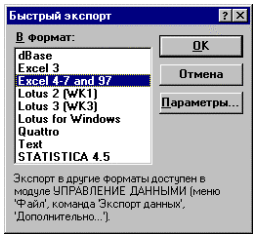
1144), из которого можно сохранить таблицу исходных данных. По умолчанию в этом диалоговом окне будет отображено имя текущего файла и место его расположения на диске. Можно сохранить файл данных под другим именем, изменить имя по умолчанию или задать другое место расположения файла таблицы исходных данных на диске.

Комбинация клавиш. F12

Быстрый экспорт

Эта команда позволит вам экспортировать файлы в форматы Excel, Quattro, Lotus 1-2-3, ASCII (текст) и *STATISTICA* 4.5. Система *STATISTICA* предоставляет набор специальных возможностей экспорта данных в модуле *Управления данными*, который поддерживает большой набор форматов файлов данных и предлагает дополнительные возможности для специального перемещения данных (например, возможность разделять/объединять значения и их метки при перемещении данных между форматами, которые не поддерживают значения с метками). Тем не менее, пункт *Быстрый экспорт* доступен из выпадающего меню *Файл* всех модулей системы *STATISTICA*.

При выборе команды *Быстрый экспорт* из выпадающего меню *Файл* откроется диалоговое окно *Быстрый экспорт*.

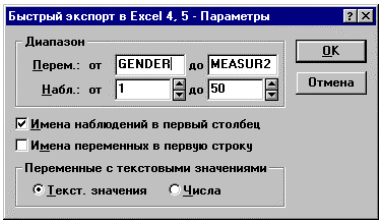


В формат. Файл системы *STATISTICA* будет экспортирован в формат, который вы укажете в этом поле. Более подробную информацию о других форматах можно найти в главе 7 *Управление данными*.

Кнопка *Параметры*. Нажмите эту кнопку, чтобы открыть диалоговое окно параметров экспорта для выбранного формата файла. Пожалуйста, обратите внимание, что при экспорте в формат *STATISTICA* 4.5 доступных параметров экспорта нет и эта кнопка затенена. Выбор пункта *STATISTICA* 4.5 приведет к сохранению текущего файла данных в формате *STATISTICA* 4.5. Поскольку некоторые типы спецификаций переменных (время, научная, валютная, проценты) не доступны в системе *STATISTICA* 4.5, переменные, использующие эти типы спецификаций, будут сохранены как числовые значения. Также, формат системы *STATISTICA* 4.5 поддерживает только блоки данных размером до 300 переменных. Вы не сможете экспортировать блоки данных, содержащие более 300 переменных, в формат *STATISTICA* 4.5.

Быстрый экспорт в форматы Excel/Lotus/Quattro

Нажатие кнопки *Параметры* в диалоговом окне *Быстрый экспорт* откроет диалоговое окно *Быстрый экспорт… - Параметры*.



Диапазон. Этот параметр позволяет выбрать определенный диапазон наблюдений и/или переменных для экспорта.

Перем.: Укажите диапазон переменных для экспорта (*от* какой переменной и *до* какой переменной включительно).

Copyright © StatSoft, 1995

StatSoft®

Набл. Укажите диапазон наблюдений для экспорта (*от* какого наблюдения и *до* какого наблюдения включительно).

Имена наблюдений в первый столбец. Выберите этот режим для того, чтобы поместить имена наблюдений в первый столбец экспортируемого файла.

Имена переменных в первую строку. Выберите этот режим для того, чтобы поместить имена переменных в первую строку экспортируемого файла.

Переменные с текстовыми значениями. Эти параметры позволяют экспортировать текстовые или числовые значения.

Текст. значения. Выберите этот вариант, если вы хотите экспортировать текстовые значения переменных.

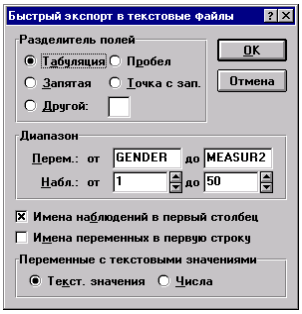
Числа. Выберите этот вариант, если вы хотите экспортировать числовые эквиваленты текстовых значений переменных.

Для того, чтобы экспортировать и текстовые значения, и числовые и/или экспортировать их метки, воспользуйтесь экспортом из модуля *Управление данными*, где предусмотрена возможность сохранения меток значений и/или пары двойной нотации в виде значений в тех файлах, которые не поддерживают этот формат записи.

Метки значений также могут быть перемещены в приложения через буфер обмена (см. раздел *Просмотр текстовых значений таблицы исходных данных*).

Быстрый экспорт текстовых (ASCII) файлов

Нажатие кнопки *Параметры* в диалоговом окне *Быстрый экспорт* откроет диалоговое окно *Быстрый экспорт… - Параметры*.



Разделитель полей. Выберите тип разделителя, который будет использоваться в экспортируемом файле. Можно выбрать тип разделителя из четырех заданных (*Табуляция*, *Пробел*, *Запятая* или *Точка с запятой*) или задать разделитель с помощью поля ввода *Другой*. Более широкие возможности использования разделителей полей при экспорте предоставлены в модуле *Управление данными*.

Диапазон. Этот параметр позволяет выбрать определенный диапазон наблюдений и/или переменных для экспорта.

Перем.: Укажите диапазон переменных для экспорта (*от* какой переменной и *до* какой переменной включительно).

Набл. Укажите диапазон наблюдений для экспорта (*от* какого наблюдения и *до* какого наблюдения включительно).

Имена наблюдений в первый столбец. Выберите этот режим для того, чтобы поместить имена наблюдений в первое поле каждой записи экспортируемого файла.

Имена переменных в первую строку. Выберите этот режим для того, чтобы поместить имена переменных в первую запись экспортируемого файла.

Переменные с текстовыми значениями. Эти параметры позволяют экспортировать текстовые или числовые значения.

StatSoft®

Copyright © StatSoft, 1995

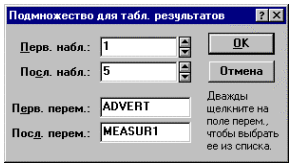
Текст. значения. Выберите этот вариант для экспорта текстовых значений переменных.

Числа. Выберите этот вариант для экспорта числовых эквивалентов текстовых значений переменных.

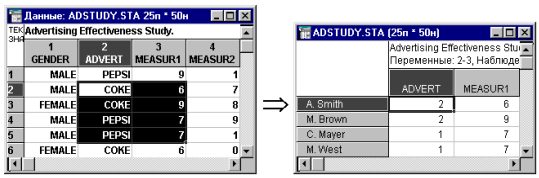
Для того чтобы экспортировать и текстовые значения, и числовые, и/или экспортировать их метки, воспользуйтесь экспортом из модуля *Управление данными*, где предусмотрена возможность сохранения меток значений и/или пары двойной нотации в виде значений в тех файлах, которые не поддерживают метки в двойной записи. Метки значений также могут быть перемещены в приложения через буфер обмена.

Преобразование в таблицу результатов

С помощью этой команды из текущей таблицы (файла данных) создается таблица результатов . При выборе пункта *Преобразовать в таблицу результатов* открывается диалоговое окно *Подмножество для табл. результатов*.



Здесь можно указать диапазон значений файла данных (по умолчанию это либо все данные, либо выделенный блок), который вы хотите включить в таблицу результатов.



Имена переменных и наблюдений станут соответственно столбцами и строками таблицы результатов; имя файла, заголовок файла и диапазон (подмножество) выделенных данных станут заголовками таблицы результатов.

Комбинация клавиш. F11

Первое наблюдение

Укажите номер первого (включительно) наблюдения из диапазона включаемых в таблицу результатов.

Последнее наблюдение

Укажите номер последнего (включительно) наблюдения из диапазона включаемых в таблицу результатов.

Первая переменная

Укажите имя первой переменной или ее номер из диапазона включаемых в таблицу результатов. Для выбора переменной из списка текущих переменных таблицы щелкните дважды на поле ввода переменной.

Последняя переменная

Укажите имя последней переменной или ее номер из диапазона включаемых в таблицу результатов. Для выбора переменной из списка текущих переменных таблицы щелкните дважды на поле ввода переменной.

Печать/выдача текущих страниц

Если запущена процедура анализа данных, а в диалоговом окне *Параметры страницы/вывода* (см. стр. 1157) включен *вывод* на *принтер* или в *файл* , то система *STATISTICA* будет действовать следующим образом. При выборе режима *Авт. выдача каждой заполненной страницы* в окне *Параметры страницы/вывода* программа будет сохранять выводимые результаты перед тем, как посылать их на принтер или в файл, до тех пор, пока они не заполнят страницу. Если этот режим

Copyright © StatSoft, 1995

не выбран, то результаты будут накапливаться, пока не будет закончен анализ. Команда *Печать/выдача текущих страниц* отменяет это внутреннее сохранение, и получаемые результаты сразу будут напечатаны (или выведены в файл) (иначе говоря, будет производиться "выброс страницы").

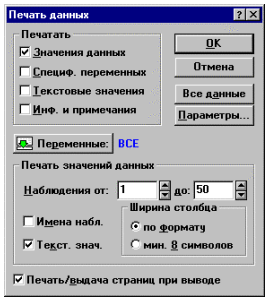
Обратите внимание, что такая возможность не предусмотрена в *Окне текста/вывода* (см. стр. 1211).

Печать данных

С помощью этой команды можно распечатать активную таблицу в файл (для печати в дальнейшем), послать на принтер (для непосредственной печати) или в *Окно текста/вывода* в зависимости от установок, сделанных в диалоговом окне *Параметры страницы/вывода* (стр. 1157). При выборе команды *Печать*, откроется диалоговое окно *Печать данных*.

Доступ через панель инструментов. Нажмите кнопку печати панели инструментов для печати таблицы без вывода диалогового окна *Печать данных* (при этом для печати будут использованы ранее выбранные параметры или параметры по умолчанию). Обратите внимание, что если в таблице выделен блок данных, то напечатан будет только этот блок; в противном случае будет напечатана вся таблица.

Комбинация клавиш. CTRL+P или F4



Печатать

Здесь можно выбрать элементы таблицы для печати, щелкая на соответствующем поле для выбора или отмены. Обратите внимание, что сделанный здесь выбор частей таблицы для печати отменит выбор стиля вывода в диалоговом окне *Параметры страницы/вывода* (стр.1157), если произойдет конфликт этих установок. Например, если в диалоговом окне *Параметры страницы/вывода* вы выбрали *Миним.* вывод и теперь выберете все варианты печати, описанные ниже, то при печати таблицы будут использованы параметры из диалогового окна *Печать данных*.

Значения данных. Выберите этот вариант для того, чтобы включить в вывод содержание таблицы (значений данных). Обычно этот режим установлен по умолчанию.

Специф. переменных. Выберите этот вариант для того, чтобы включить в вывод таблицу всех спецификаций переменных (имя, код пропущенных данных, формат и длинное имя переменной, как указано в диалоговом окне *Изменение (текущих) спецификаций таблицы*.

Текстовые значения. Выберите этот вариант для того, чтобы включить в вывод отдельную таблицу присвоения текстовых значений (текстовые значения, числовые эквиваленты и метки значений как указано в диалоговом окне *Диспетчер текстовых значений*, см. стр. 1178).

Инф. и примечания. Выберите этот вариант для того, чтобы включить в вывод заголовок файла, информацию о файле и примечания (см. раздел *Заголовок файла*, стр. 1192).

Переменные

Выберите переменные таблицы, которые вы хотите включить в вывод (по умолчанию: *Все* или переменные текущего блока, если он выделен). Нажав на эту кнопку, вы откроете

StatSoft®

Copyright © StatSoft, 1995

диалоговое окно *Выбор переменных для печати*, в котором можно выбрать нужные переменные.

Все данные

Нажмите эту кнопку для того, чтобы автоматически выбрать все переменные и наблюдения файла данных. Эта кнопка предоставляет возможность быстрого изменения выбора блока данных в этом диалоговом окне для всех переменных и наблюдений. Она действует, если в таблице выделен блок данных.

Печать значений данных

Наблюдения. Можно указать в этих полях набор наблюдений, которые хотите включить в вывод (*от* какого наблюдения и *до* какого наблюдения включительно).

Имена набл. Выберите этот режим для того, чтобы включить в вывод имена наблюдений.

Текст. знач. Выберите этот режим для печати текстовых значений вместо числовых значений переменных таблицы.

Ширина столбца. Выберите вид ширины столбца таблицы: форматированная ширина (*По формату*) или минимальная ширина 8 символов (*Мин. 8 символов*); последнее позволит вам печатать полные имена переменных (независимо от текущих форматов значений, которые могут быть короче, чем полные имена).

Печать/выдача страниц при выводе

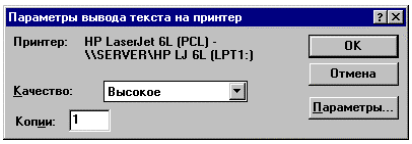
Выбор этого режима позволит вам автоматически отправить распечатку на принтер или в файл как только вы нажмете кнопку *OK* этого диалогового окна или после нажатия кнопки *Печать* на панели инструментов (текст, оставшийся во внутреннем буфере, также будет распечатан). Поэтому каждая распечатка будет начинаться с новой страницы. Если этот режим не выбран , то вывод будет сохраняться во внутреннем буфере и отправляться на принтер в зависимости от установки параметра *Авт. выдача каждой заполненной страницы* (т.е. либо сразу после заполнения каждой страницы, либо позже, например, при выходе из модуля или очистке буфера с помощью команды *Печать/выдача текущих страниц*). Обратите внимание, что этот режим доступен только локально (временно) для текущего модуля системы *STATISTICA* и текущего сеанса, тогда как режим *Печать/выдача страниц при выводе* из диалогового окна *Параметры страницы/вывода* (стр. 1157) действует глобально (постоянно).

Параметры

Нажмите эту кнопку, чтобы открыть диалоговое окно *Параметры вывода текста на принтер*, в котором можно выбрать качество (разрешение) распечатки и число копий (см. ниже). В этом окне также показан выбранный драйвер принтера, который можно изменить, нажав кнопку *Параметры*, чтобы открыть диалоговое окно *Параметры печати* (см. ниже), в котором можно выбрать принтер, изменить ориентацию или размер бумаги и ее источник.

Параметры вывода текста на принтер

Это диалоговое окно, доступное из диалогового окна *Печать данных* (см. выше) через кнопку *Параметры*, позволяет вам выбрать качество (разрешение) распечатки и число копий при выводе текста.



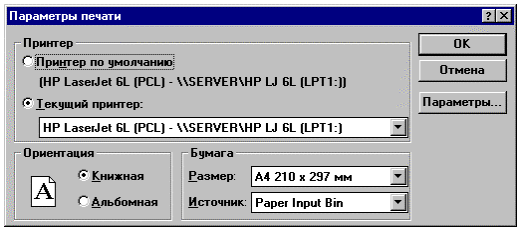
В этом окне также показан выбранный драйвер принтера, который можно изменить, нажав

Copyright © StatSoft, 1995

StatSoft®

кнопку *Параметры*, чтобы открыть диалоговое окно *Параметры печати* (см. стр. 1165).

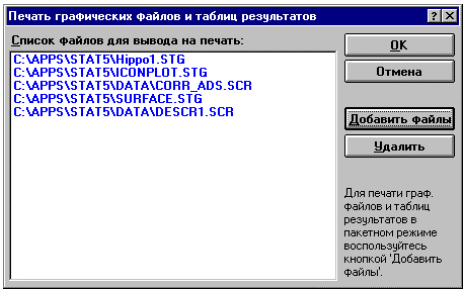
В диалоговом окне *Параметры печати* можно выбрать принтер из списка тех, которые установлены на вашем компьютере, задать ориентацию печатной страницы и указать размер бумаги и ее источник для печати графических изображений и текста.



Обратите внимание, что границы вывода текста можно изменить в диалоговом окне *Поля вывода текста* (открывающемся с помощью кнопки *Поля/Принтер* диалогового окна *Параметры страницы/вывода*, см. ниже).

Печать графиков и таблиц результатов

Эта команда (из выпадающего меню *Файл*) позволяет напечатать ранее сохраненные графики и таблицы результатов совместно. Она может пригодиться для того, чтобы позднее, в конце работы напечатать совместно графики и таблицы результатов. Вы просто сохраняете графики или таблицы (с помощью команд *Сохранить таблицу результатов* или *Сохранить график*, см. стр. 1144) после того, как получили их, а потом открываете это диалоговое окно,



выделяете нужные графики и таблицы результатов (см. выше) и нажимаете кнопку *OK* для начала печати.

Когда вы нажмете кнопку *OK* в этом диалоговом окне, сначала будут напечатаны все выделенные графики (на отдельных страницах, в том порядке, в котором они выделены), а потом будут напечатаны совместно все таблицы результатов (в том порядке, в котором они выделены).

Список файлов для вывода на печать

В этой части диалогового окна отображается список файлов для печати. Можно добавлять или удалять файлы из списка с помощью описанных ниже кнопок. Как только график или таблица результатов будет отправлена на принтер, имя соответствующего файла будет удалено из поля *Список файлов для вывода на печать*.

Добавить файлы

Как только вы нажмете эту кнопку, откроется диалоговое окно *Выбор файлов для печати* (см. стр. 1144), в котором можно выбрать график или таблицу результатов, которую вы хотите напечатать. Нажмите клавишу CTRL для того, чтобы выбрать несколько файлов из списка (непрерывно или по очереди). Имена выделенных файлов будут помещены в поле *Список файлов для вывода на печать* (см. выше).

StatSoft®
Copyright © StatSoft, 1995

Удалить

Нажмите кнопку для того, чтобы удалить выделенные файлы из поля *Список файлов для вывода на печать*.

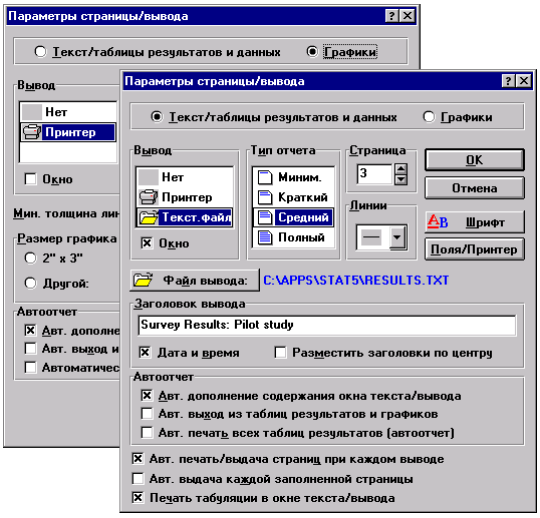
Отмена

Как только вы нажмете кнопку *OK* в этом диалоговом окне, начнется печать, но диалоговое окно будет по-прежнему отображаться на экране. Будут затенены все кнопки, кроме кнопки *Отмена*, которую можно использовать для прерывания процесса печати. Когда вы нажмете кнопку *Отмена*, печать прекратится после того, как будет допечатан последний график или таблица результатов, содержащаяся в буфере принтера.

Можно также прервать печать отдельного графика или таблицы результатов (не прерывая печать всей группы), нажав кнопку *Отмена* в диалоговом окне Windows *Печать*.

Параметры страницы/вывода (текст/таблицы результатов и данных)

Это диалоговое окно позволяет вам задать, куда будет производится вывод файлов, объем дополнительной информации, поля и другие параметры. Его можно вызвать двойным щелчком на поле *Вывод* строки состояния, с помощью пункта *Параметры страницы/вывода* выпадающего меню *Файл* или выбрав пункт *Принтер* выпадающего меню *Сервис*. Все параметры, которые вы здесь измените, будут установлены как параметры по умолчанию, и будут доступны глобально каждый раз, когда вы открываете какой-либо модуль системы *STATISTICA*.



Перед тем, как печатать текст, таблицу исходных данных или таблицу результатов, выберите, куда будет производиться печать (принтер, файл и/или *Окно текста/вывода* на закладке *Текст/табл. рез./табл. данных*, или принтер и/или *Окно текста/вывода* на закладке *Графики*), а также задайте различные другие параметры на соответствующей закладке этого диалогового окна.

Следующие варианты доступны при выборе закладки *Текст/табл. рез./табл. данных*.

Комбинация клавиш. SHIFT+F4

Вывод

Выберите, куда будет производиться вывод.

Нет. Если выбран этот вариант, то текст не будет отправлен ни на принтер, ни в файл.

Текст. файл. Выберите этот вариант для того, чтобы отправить содержимое таблицы результатов или таблицы исходных данных в указанный пользователем текстовый файл. Когда вы выберете этот пункт и нажмете *OK*, откроется диалоговое окно *Укажите файл вывода* (см. стр. 1144), в котором можно указать

Copyright © StatSoft, 1995

файл, в который будет отправлено содержание таблицы. Кроме того, после выбора этого режима можно нажать кнопку *Файл вывода*, чтобы указать имя файла, под которым будет сохранен вывод.

Принтер. При выборе этого варианта содержание таблицы исходных данных или таблицы результатов будет отправлено на принтер.

Окно. Если выбран этот вариант, то независимо от установок вывода, выбранных ранее (т.е. *Файл* или *Принтер*), вывод таблицы исходных данных или таблицы результатов будет отправлен в *Окно текста/вывода* (см. стр. 1211). Примечание: при выборе этого режима откроется *Окно текста/вывода*, которое было сохранено последним. Чтобы сохранить вывод из этого окна, воспользуйтесь командой *Сохранить* выпадающего меню *Файл* или нажмите клавиши CTRL+S.

Тип отчета

Здесь можно задать объем дополнительной информации для вывода на печать.

Миним. Если вы хотите напечатать только содержание выбранных таблиц результатов или таблиц исходных данных (без заголовков страниц, заголовков заданий и другой информации о переменных или условиях анализа), выберите этот наиболее "экономный" стиль.

Краткий. Этот стиль включает заголовки страницы [название модуля, дату/время (если требуется, см. ниже), номер страницы], текущее имя файла данных, информацию об *условиях выбора и весах наблюдений* (если они были заданы, см. стр. 1223 и 1225 соответственно), заголовок вывода (см. ниже), содержание таблицы результатов или таблицы исходных данных, список (коротких) имен всех переменных, выбранных для каждого анализа, и значения пропущенных данных.

Средний. Этот формат отчета занимает больше места, включая помимо таблицы результатов или таблицы исходных данных всю информацию *краткого* формата, и, кроме того, длинные имена переменных (например, формулы). При этом резервируется одна строка для каждой переменной, которая выбрана для текущего анализа.

Полный. *Полный* стиль отчета предоставляет наиболее полную информацию о каждой переменной, которая выбрана для анализа. В дополнение к информации, содержащейся в отчете *среднего* формата, он также включает полный список для каждой выбранной переменной о ее "двойных" значениях, то есть, о значениях, у которых есть и числовые, и текстовые компоненты и/или метки.

Обратите внимание, что при печати таблиц исходных данных, ваш выбор печати частей таблиц исходных данных, выбранных в диалоговом окне *Печать данных* (стр. 1154), отменит ваш выбор стиля вывода, если произойдет конфликт двух установок. Например, если здесь вы выберете *минимальный* стиль вывода, а потом выберете все параметры для печати в диалоговом окне *Печать данных*, то для печати таблицы будут использованы параметры диалогового окна *Печать данных*.

Файл вывода

Эта кнопка доступна, если выбран вывод в *файл* и затенена в противном случае. Нажатие этой кнопки приведет к открытию диалогового окна *Укажите файл вывода* (см. стр. 1144). Здесь можно указать имя файла (с расширением **.txt*), под которым будет сохранен текстовый (ASCII) вывод из таблиц результатов или таблиц исходных данных. Эти файлы вывода могут быть позднее открыты в *Окне текста/вывода* для просмотра, печати или внесения изменений.

StatSoft®

Copyright © StatSoft, 1995

Обратите внимание, что, в отличие от файлов отчета, созданных в результате непосредственного вывода в *Окно текста/вывода*, эти текстовые (ASCII) файлы не содержат графиков и информации о форматировании (шрифт, границы, расположение и т.д.).

Страница

Укажите здесь номер страницы, с которой начнется текстовый вывод. Дальнейшие страницы будут пронумерованы последовательно, начиная с этого номера (это удобно, если вы продолжаете уже существующий отчет). Обратите внимание, что страницы не могут быть пронумерованы последовательно, если у вас одновременно запущено более одного модуля системы *STATISTICA* и все эти модули производят вывод (т.е. работа происходит не в режиме одного приложения, см. раздел *Переключатель модулей* системы *STATISTICA*, стр. 1217).

Линии

Здесь можно выбрать толщину линий (или их отсутствие) в выводимой таблице. Если вы не хотите, чтобы в таблицах были линии или границы, выберите здесь пустое поле. Теперь таблицы результатов или таблицы исходных данных будут печататься без граничных линий, и печать будет производиться гораздо быстрее.

Заголовок вывода

Здесь можно задать заголовок вывода. Этот заголовок будет печататься на каждой странице (см. выше *краткий* стиль).

Дата и время

Выберите этот вариант, если вы хотите включить в заголовок страницы дату и время (см. выше параметры *краткого* вывода).

Разместить заголовки по центру

Выберите этот вариант для расположения *заголовка вывода* по центру (вместо смещения влево) .

Шрифт

Нажатие этой кнопки приведет к открытию диалогового окна *Шрифт* (стр. 1208), в котором можно выбрать шрифт и размер шрифта для вывода на печать.

Поля/Принтер

Нажмите эту кнопку, чтобы открыть диалоговое окно *Поля вывода текста* (см. ниже), в котором можно задать поля для печати страниц с таблицами результатов, таблицами исходных данных или отчета из *Окна текста/вывода*. Можно задать границы для печати графиков в диалоговых окнах *Предварительный просмотр* или *Параметры страницы/вывода*, доступные из графического выпадающего меню *Файл* (см. том II настоящего руководства).

Автоматическая печать/выдача страниц при каждом выводе

Если выбран этот режим, то распечатка будет автоматически посылаться на принтер или в файл, как только вы нажмете на кнопку *OK* в диалоговом окне *Печать данных* или *Печать табл. результатов* (стр. 1154 и глава 4, соответственно) или на кнопку *Печать* на панели инструментов (при этом также будет напечатан весь текст, оставшийся во внутреннем буфере). Поэтому каждая распечатка (таблица) будет начинаться с новой страницы. Если этот режим не выбран, то вывод будет сохранен во внутреннем буфере и отправлен на принтер в зависимости от текущего состояния поля *Авт. выдача каждой заполненной страницы* (либо каждая страница, когда она заполнена, либо все сразу позже, например, при выходе из модуля или очистке буфера с помощью команды

Copyright © StatSoft, 1995

Печать/выдача текущих страниц, см. стр. 1153).

Автоматическая выдача каждой заполненной страницы

Если выбран это режим, система *STATISTICA* будет сохранять результаты анализа во внутреннем буфере только до тех пор, пока не будет заполнена страница. Сразу после заполнения она будет отправлена на принтер или в файл (в зависимости от установки в поле *Вывод*). Система *STATISTICA* будет продолжать сохранять результаты и посылать вывод на принтер или в файл по одной странице за один раз. Можно воспользоваться командой *Печать/выдача текущих страниц* из выпадающего меню *Файл* (см. стр. 1153) в любой момент анализа, чтобы немедленно послать на принтер или в файл то, что находится в буфере.

Если этот режим не используется, печать будет занимать меньше времени: система *STATISTICA* будет сохранять все результаты анализа и не будет печатать или посылать их в файл до тех пор, пока вы не закроете активный модуль или не выберете пункт *Печать/выброс текущих страниц* из выпадающего меню *Файл* (см. стр. 1153). Можно воспользоваться командой *Печать/выброс текущих страниц* из выпадающего меню *Файл* (см. стр. 1153) в любой момент анализа, чтобы немедленно послать на принтер или в файл то, что находится в буфере.

Режимы автоотчета

Следующие параметры позволяют автоматически создать отчет (например, отчет, содержащий текст/графики) о всех процедурах анализа (см. стр. 1162 для информации о параметрах автоотчета для графиков).

Авт. дополнение содержания Окна текста/вывода. Можно выбрать этот режим, чтобы автоматически открывать последний сохраненный текстовый файл (расширение **.txt*) или файл в формате *R*ich *T*ext *F*ormat (расширение **.rtf*) в новом *Окне текста/вывода* (например, для того, чтобы добавить текущий вывод к результатам предыдущего анализа). Если этот режим не выбран, то из диалогового окна *Параметры страницы/вывода* будет открыто пустое *Окно текста/вывода*.

Авт. выход из таблиц результатов и графиков. Если таблицы результатов (и/или графики, см. ниже) созданы совместно как результат анализа (например, таблицы частот для серий переменных), можно выбрать этот режим, чтобы предотвратить появление кнопки *Далее* в таблицах результатов, если очередь переполнена. Если теперь выбрать режим *Авт. печать всех таблиц результатов*, таблицы результатов будут напечатаны в пакетном режиме без возможности прерывания печати.

Авт. печать всех таблиц результатов (автоотчет). Если выбран этот режим печати, то кроме дополнительной информации (объем которой зависит от текущих установок формата отчета, заданных в поле *Тип отчета*), будет автоматически напечатан на принтере или отправлен в файл весь вывод анализа (и/или *Окно текста/вывода*). Этот режим печати может быть использован, если вы хотите сохранять весь вывод, отображаемый во время проведения анализа.

Если вы отмените выбор режима *Автоматическая печать всех таблиц результатов*, то таблицы результатов (или их части) будут распечатаны только после выбора пункта *Печать* выпадающего меню *Файл* (или нажатия кнопки).

Автоматическая печать/выдача страниц при каждом выводе

Если включен этот режим, страницы будут отправлены на принтер (или в файл) после каждого произведенного программой вывода текста (например, через команды *Печать данных*

StatSoft®

Copyright © StatSoft, 1995

или *Печать таблицы результатов* или кнопку панели инструментов). Поэтому, каждая распечатка (таблица) будет начинаться с новой страницы. При создании большого числа достаточно коротких порций выходных данных такой режим может замедлить вашу работу, так как процесс печати будет начинаться очень часто (для каждой маленькой части вывода). Тем не менее, преимущество этого режима заключается в возможности вывода на печать сразу после поступления запроса.

Автоматическая выдача каждой заполненной страницы

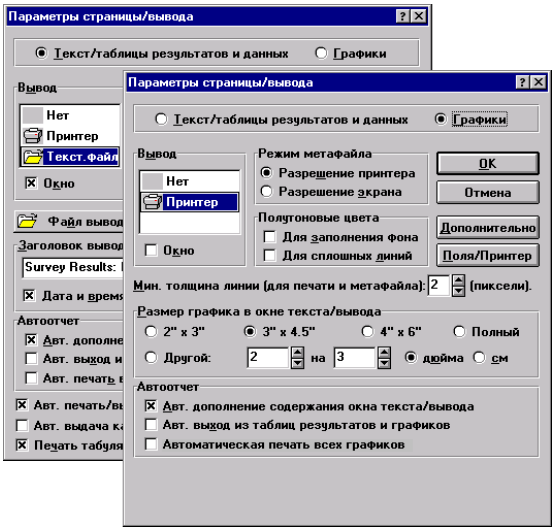
Если включен этот режим, то каждая страница распечатки будет выдаваться сразу после заполнения. Этот режим немного медленнее, так как процесс печати начинается отдельно для каждой заполненной страницы. Однако, преимущество его состоит в том, что он автоматически производит распечатку вывода по мере прохождения анализа. Кроме того, если этот режим, а также режим *Авт. печать/выдача* (см. предыдущий параграф) выключены, то весь вывод будет сохраняться во внутреннем буфере и посылаться на принтер сразу весь в конце сеанса (или когда вы выберете пункт *Печать/выброс текущих страниц* выпадающего меню *Файл*, который позволяет вам очистить буфер вывода в любое время).

Печать табуляции в окне текста/вывода

Если выбран этот вариант, столбцы таблицы результатов при печати в *Окно текста/вывода* будут разделены табуляцией (это установлено по умолчанию). Текстовые таблицы, форматированные таким образом, могут быть быстро преобразованы в таблицы текстовых редакторов (например, MS Word).

Параметры страницы/вывода (графики)

Установки параметров на этой закладке диалогового окна *Параметры страницы/вывода* определяют печать графиков.



Вывод

В этом поле можно направить вывод на *принтер*, и/или в *Окно* (для вывода отчета на экран). Информацию об *Окне текста/вывода* можно найти в главе 5.

Режим метафайла

Все графические метафайлы, которые создаются системой *STATISTICA*, включая представление буфера обмена, сохраненные метафайлы и графики, выводимые в *Окно текста/вывода*, могут быть представлены либо в разрешении принтера (которое зависит от текущего драйвера принтера и позволяет делать вывод самого высокого качества, но требует больше памяти или дискового пространства), или разрешении экрана (которое обеспечивает лучшее отображение на экране).

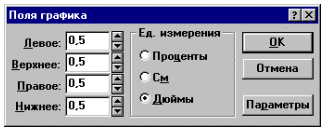
Copyright © StatSoft, 1995

Полутоновые цвета

Эти параметры контролируют использование полутонов для заполнения фона графиков и текста пользователя, условных обозначений, внедренных объектов и др. (по умолчанию они не выбрана, чтобы при печати отображение графика было оптимальным), а также для печати сплошных линий (это может быть полезным, если линии достаточно толстые, иначе они могут оказаться “разорванными”).

Поля/Принтер

Нажатие кнопки *Поля/Принтер* откроет диалоговое окно, контролирующее границы вывода графиков.



Для выбора ориентации бумаги (книжной или альбомной), разрешения принтера и других параметров вывода графиков, нажмите кнопку *Параметры* (это аналогично выбору пункта *Принтер* выпадающего меню *Файл*).

Обратите внимание, что границы (а также ориентация страницы и значения других параметров) для текстовых распечаток устанавливаются независимо (см. выше и стр. 1164).

Минимальная толщина линии

Здесь можно указать минимальную толщину линии для распечаток и метафайлов (в точках принтера или другого устройства). Этот параметр часто используется для устройств вывода с высоким разрешением (например, 1200 точек на дюйм; при этом линии толщиной в одну точку оказываются слишком тонкими), а также, если вы по умолчанию задаете маленький размер графика в *Окне текста/вывода* (например, 2x3, см. ниже параметр *Размер графика в Окне текста/вывода*).

Если вы не измените значение этого параметра, установленное по умолчанию (2 пикселя принтера), то ширина всех линий, которые будут печататься системой *STATISTICA*, будет равна большему из двух значений: индивидуально заданной ширине линий графических компонент (по умолчанию 0,5 пункта) или 2 пикселям устройства вывода. Этот параметр позволяет задать минимальную толщину печатаемых линий; например, если вы зададите 3, то все печатаемые линии будут как минимум 3 пикселя шириной, даже если в диалоговых окнах *Шаблоны точек/линий/областей* для них установлены меньшие значения.

Если значение параметра *Размер графика в Окне текста/вывода* выбрано меньше, чем *Полный*, то линии будут пропорционально тоньше. Для компенсации этого вам может понадобиться изменить минимальную толщину линий.

Размер графика в окне текста/вывода

В этом поле можно установить размер графика по умолчанию для вывода в *Окно текста/вывода* (после вывода размер графиков можно изменить).

Автоотчет

Следующие параметры позволяют автоматически создавать отчет (например, содержащий текст и графические изображения) обо всех производимых процедурах анализа (см. выше режимы автоотчета для таблиц результатов и таблиц исходных данных).

Авт. дополнение содержания окна текста/вывода. Можно выбрать этот режим для того, чтобы автоматически открывать последний сохраненный текстовый файл (расширение **.txt*) или файл в формате *R*ich *T*ext *F*ormat (с расширением **.rtf*) в новом *Окне*

StatSoft®

Copyright © StatSoft, 1995

текста/вывода (например, чтобы добавить текущий вывод к результатам предыдущего анализа). Если этот режим не выбран, то из диалогового окна *Параметры страницы/вывода* будет открыто пустое *Окно текста/вывода*.

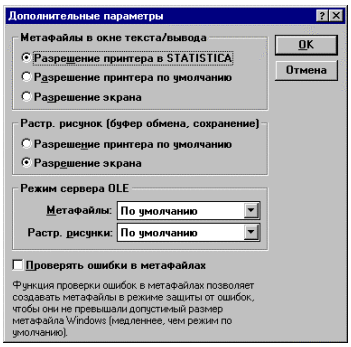
Авт. выход из таблиц результатов и графиков. Если графики (и/или таблицы результатов, см. выше) созданы совместно как результат анализа (например, пункт *Категоризованные диаграммы рассеяния* в диалоговом окне *Описательные статистики*), можно выбрать этот режим, чтобы предотвратить появление кнопки *Далее* в таблицах результатов, если очередь переполнена. В этом случае при выборе режима *Автоматическая печать всех графиков* графики будут напечатаны в пакетном режиме без возможности прерывания печати.

Автоматическая печать всех графиков. Эта кнопка позволяет автоматически получать распечатки всех графиков, отображаемых на экране. Вывод будет производиться в текущий канал вывода (принтер и/или *Окно текста/вывода*) с использованием текущих параметров печати (например, типа принтера, границ, ориентации бумаги и др., установленных в диалоговом окне *Параметры принтера*). Обратите внимание, что существует также возможность печати в пакетном режиме, которая может быть использована для печати сохраненных ранее графиков (в файлах с расширением **.stg*). Она вызывается с помощью пункта *Печать файлов...* выпадающего меню *Файл*, см. стр. 1156.

При печати в *Окно текста/вывода* с включенным режимом *Авт. печать всех таблиц результатов* (см. выше) будет создан комбинированный отчет обо всех процедурах анализа, содержащий как текст, так и графические изображения.

Дополнительные параметры

Это диалоговое окно открывается с помощью кнопки *Дополнительно* в диалоговом окне *Параметры страницы/вывода - Графики*.



Метафайлы в Окне текста/вывода

С помощью этих параметров можно изменять разрешение и представление (цвета) всех графических метафайлов, создаваемых системой *STATISTICA,* которые отображаются в *Окне текста/вывода*.

Разрешение принтера в STATISTICA. Если выбран этот вариант и в *Окно текста/вывода* посылается график, то представление метафайла, созданного в *Окне текста/вывода*, будет зависеть от установок принтера для текущего сеанса системы *STATISTICA*.

Разрешение принтера по умолчанию. Если выбран этот вариант и в *Окно текста/вывода* посылается график, то представление метафайла (рисунка), созданного в *Окне текста/вывода*, будет зависеть от текущих установок (по умолчанию) драйвера принтера. (Они не изменятся при изменении установок драйвера во время текущего сеанса системы *STATISTICA*.)

Разрешение экрана. Если выбран вариант *Разрешение экрана* и в *Окно текста/вывода* посылается график, то представление метафайла (рисунка), созданного в *Окне текста/вывода*, будет зависеть от текущего драйвера экрана.

Copyright © StatSoft, 1995

Растровый рисунок (буфер обмена, сохранение)

С помощью этих параметров можно контролировать разрешение и представление (цвета) всех графических растровых изображений, создаваемых системой *STATISTICA* при копировании в буфер обмена или сохранении в виде растрового изображения.

Разрешение принтера по умолчанию. Если выбрана этот вариант, то представление растрового изображения будет зависеть от текущего драйвера принтера.

Разрешение экрана. Если выбран вариант *Разрешение экрана*, то представление растрового изображения будет зависеть от текущего драйвера экрана.

Режим сервера *OLE*

Эти параметры позволяют контролировать разрешение и представление (цвета) графиков, создаваемых системой *STATISTICA,* которые будут в дальнейшем включены в другие приложения с помощью метода *OLE*.

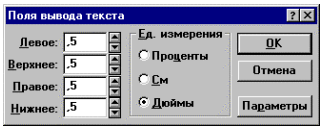
Метафайлы/Растр. рисунки. В каждом выпадающем списке доступны те же три варианта: по умолчанию, разрешение принтера и разрешение экрана. Режим по умолчанию использует разрешение, запрашиваемое клиентом. Если выбрано разрешение принтера, то представление будет зависеть от текущего драйвера принтера. Если выбрано разрешение экрана, то представление будет зависеть от текущего драйвера экрана.

Проверять ошибки в метафайлах

Это поле позволяет включить проверку ошибок в метафайлах, чтобы они не превышали допустимый в Windows объем; заметим, что режим, установленный по умолчанию обеспечивает более высокую скорость работы с этими файлами.

Поля вывода текста

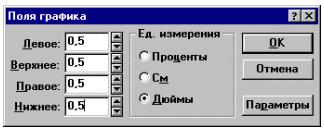
В этом диалоговом окне, доступном из закладки *Текст/Табл. результатов/Табл. исходных данных* диалогового окна *Параметры страницы/вывода* (см. выше), можно задавать расположение нужного текста (например, из таблиц исходных данных, таблиц результатов или *Окна текста/вывода*) на печатаемой странице. Для этого нужно изменить значения полей (задав *левое*, *верхнее*, *правое* и *нижнее* поля) в соответствии с используемыми единицами измерения (*проценты*, *сантиметры* или *дюймы*). Обратите внимание, что эти значения не зависят от значений полей для вывода графиков (которые можно изменить в диалоговых окнах *Предварительный просмотр* или *Поля графика*, см том II настоящего руководства).



Для изменения размера и ориентации бумаги (книжная или альбомная) может быть использован пункт *Принтер* выпадающего меню *Файл*.

Поля графика

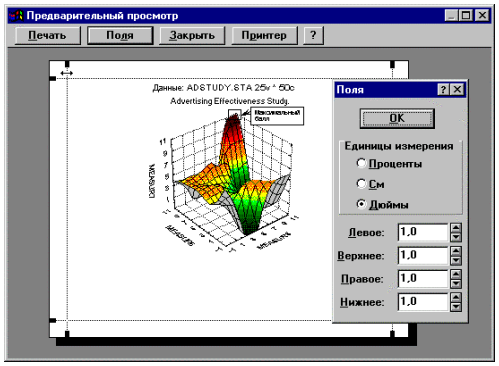
В этом диалоговом окне можно задавать расположение графического изображения на печатаемой странице, изменяя значения полей (задав *левое*, *верхнее*, *правое* и *нижнее* поля) в соответствии с используемыми единицами измерения (*проценты*, *сантиметры* или *дюймы*).



StatSoft®

Copyright © StatSoft, 1995

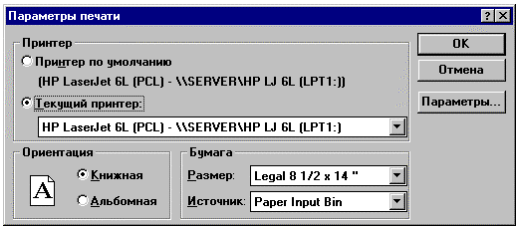
После нажатия кнопки *Параметры* в диалоговом окне *Поля графика* откроется диалоговое окно параметров печати для принтера, выбранного в диалоговом окне *Выбор печатающих устройств*. В нем можно задать параметры принтера, которые будут использованы по умолчанию при печати графических изображений на принтере. Эти параметры принтера изменяются в зависимости от типа выбранного принтера; поэтому выбор принтера соответствующим образом отразится в диалоговом окне, вызываемом кнопкой *Параметры*. Границы также можно изменить интерактивно в диалоговом окне *Предварительный просмотр*. Это окно вызывается с помощью кнопки на панели инструментов графиков (см. главу 2).



Для изменения размера и ориентации бумаги (книжная или альбомная) можно использовать пункт *Принтер* выпадающего меню *Файл*.

Принтер…

После выбора этого пункта из выпадающего меню *Файл* откроется диалоговое окно *Параметры печати*, в котором можно выбрать принтер, ориентацию и размер бумаги и источник печати графиков и текста.



Обратите внимание, что можно задавать различные параметры для вывода текста (а также таблиц результатов и таблиц исходных данных) и графических изображений.

Например, можно выбрать *альбомную* ориентацию страницы для печати текста в диалоговом окне *Параметры печати*, доступном из выпадающего меню *Файл* таблицы исходных данных или таблицы результатов, и *книжную* ориентацию страницы для печати графических изображений в диалоговом окне *Параметры печати*, доступном из выпадающего меню *Файл* графического окна.

Принтер

Выберите принтер из списка принтеров, установленных в вашей системе, или задайте принтер по умолчанию.

Ориентация

Выберите *альбомную* или *книжную* ориентацию страницы для печати графических изображений или текста.

Бумага

Здесь можно выбрать размер и источник бумаги.

Параметры

После нажатия этой кнопки появится диалоговое окно (характерное для выбранного принтера, см. выше), в котором можно выбрать различные параметры печати (например, область печати, границы, изменение размера и т.д.).

Copyright © StatSoft, 1995
StatSoft®

Выход

Выбор этого пункта приведет к окончанию сеанса системы *STATISTICA*. Если в таблицы исходных данных, таблицы результатов, текст или графики были внесены изменения и пункт *Выход* выбран до их сохранения, то система *STATISTICA* выведет запрос о сохранении измененных объектов.

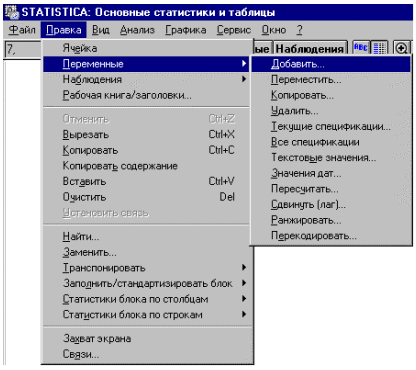
Комбинация клавиш. ALT+F4

StatSoft®

Copyright © StatSoft, 1995

ВЫПАДАЮЩЕЕ МЕНЮ *ПРАВКА* ТАБЛИЦЫ ИСХОДНЫХ ДАННЫХ

В выпадающем меню *Правка* доступны следующие диалоговые окна и команды.



Далее приведено описание показанных выше команд.

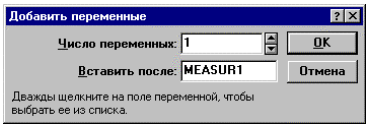
Изменение ячейки

Пункт *Ячейка* выпадающего меню *Правка* позволяет изменить текущую выделенную ячейку, не удаляя ее содержимое. Исправление можно начать, выбрав этот пункт, нажав F2, вводя значения при выделенной ячейке или вводя значения вместо текущего.

Комбинация клавиш: F2

Добавить переменные

Добавить переменные в таблицу исходных данных можно, задав число добавляемых переменных, а также место, куда их добавить (после какой переменной), в диалоговом окне *Добавить переменные*. Оно доступно из выпадающего меню *Правка - Переменные* таблицы исходных данных, контекстного меню таблицы исходных данных, или через кнопку **Переменные** панели инструментов.



После двойного щелчка на поле ввода переменной диалогового окна *Добавить переменные* будет выведен список переменных текущей таблицы исходных данных, в котором можно выбрать, после какой переменной будут добавлены новые.

Число переменных

Укажите число переменных, которые нужно добавить в текущую таблицу исходных данных. В таблицу исходных данных системы *STATISTICA* можно включить до 4092 переменных, а в таблицу *Менеджера мегафайлов* - до 32000 переменных (до 255 символов на переменную) (см. главу 7 *Управление данными*).

Вставить после

Укажите переменную таблицы исходных данных, справа от которой будут добавлены новые переменные. Можно щелкнуть на поле ввода переменных диалогового окна *Добавить переменные* для вывода списка переменных текущей таблицы исходных данных, в котором можно выбрать, куда будут добавлены новые переменные.

Как добавить блок перед первой переменной.. Первый способ вставить значение переменной или блока переменных перед первой переменной - использовать возможности метода *Перетащить и отпустить* системы *STATISTICA*. Для этого надо сначала выделить нужный блок, затем переместить курсор к одной из границ блока, где он изменится на стрелку. Теперь для того, чтобы переместить и вставить блок, надо нажать

Copyright © StatSoft, 1995

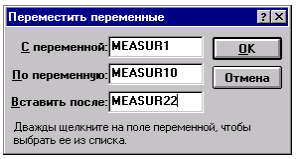
StatSoft

клавишу SHIFT и перетащить его. Далее надо установить полосу вставки между столбцом имен наблюдений и первым столбцом переменных и отпустить кнопку мыши. Обратите внимание, что этот метод перемещает не выделенные переменные (т.е. их длинные метки значений, форматы, присвоения текстовых значений и т.д.), а только их значения из выделенного блока.

Другой способ переместить и вставить переменную или блок переменных до первой переменной заключается в том, чтобы вставить переменные после первой переменной, а потом переместить первую переменную направо от перемещенных ранее переменных (также можно воспользоваться командой *Копировать переменные*, чтобы сначала скопировать переменные, а потом переместить их, см. стр. 1169).

Переместить переменные

В отличие от операций вырезания и вставки выделенного диапазона или блока значений из буфера обмена в таблице исходных данных, эта операция удалит и содержание столбцов, и сами столбцы (независимо от того, выделены ли все наблюдения для соответствующих переменных, или только их часть) и вставит столбцы на указанное место таблицы исходных данных; поэтому размер файла не изменится. Можно перемещать одну или несколько переменных таблицы исходных данных, указав диапазон (включительно) перемещаемых переменных и место (переменную, после которой произвести вставку) в диалоговом окне *Переместить переменные* (доступного из выпадающего меню *Правка - Переменные* таблицы исходных данных или через кнопку Переменные панели инструментов).



Можно быстро заменить одну выделенную переменную другой, дважды щелкнув на поле ввода переменной в окне *Переместить переменные* для вывода списка переменных текущей таблицы исходных данных, в котором можно выбрать нужную переменную.

С переменной

Здесь указывается первая переменная (включительно) в перемещаемом блоке переменных. (Обратите внимание, что если вы перемещаете только одну переменную, то в списках *С переменной* и *По переменную* надо указать одно и то же значение.)

По переменную

Здесь указывается последняя переменная (включительно) в перемещаемом блоке переменных.

Вставить после

Здесь указывается место (переменная) в таблице исходных данных, после которого будут вставлены переменные (нельзя вставить блок переменных *после* переменной, которая включена в диапазон перемещаемых переменных). Блок переменных будет “вырезан” из текущего места в таблице исходных данных и перемещен направо от указанной здесь переменной.

Как поместить блок до первой переменной. Первый способ вставить значение переменной или блока переменных перед первой переменной – использовать возможности метода *Перетащить и отпустить* системы *STATISTICA*. Для этого надо сначала

StatSoft

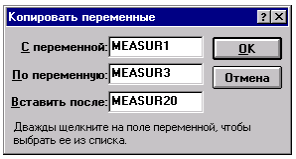
Copyright © StatSoft, 1995

выделить нужный блок. Затем переместить курсор к одной из границ блока, где он изменится на стрелку. Теперь для того, чтобы переместить и вставить блок, надо нажать клавишу SHIFT и перенести его. Далее надо установить полосу вставки между столбцом имен наблюдений и первым столбцом переменных и отпустить кнопку мыши. Обратите внимание, что этот метод перемещает не выделенные переменные (т.е. их длинные метки значений, форматы, присвоения текстовых значений и т.д.), а только их значения из выделенного блока.

Другой способ переместить и вставить переменную или блок переменных до первой переменной заключается в том, чтобы вставить переменные после первой переменной, а потом переместить первую переменную направо от перемещенных ранее переменных (также можно воспользоваться командой *Копировать переменные*, чтобы сначала скопировать переменные, а потом переместить их, см. ниже).

Копировать переменные

В отличие от операций копирования и вставки выделенного диапазона или блока значений из буфера обмена в таблице исходных данных, эта операция копирует и содержание столбцов, и сами столбцы (независимо от того, выделены ли все наблюдения для соответствующих переменных, или только их часть) и вставляет столбцы на указанное место, поэтому размер файла при этой операции увеличивается. Можно копировать указанные переменные, указав диапазон (включительно) копируемых переменных и место в таблице исходных данных (переменную, после которой произвести вставку) в диалоговом окне *Копировать переменные* (доступно из выпадающего меню *Правка - Переменные*).



Операция копирования вставит не только данные, но и формат, длинное имя, формулы, присвоенное числовое или текстовое значение и т.д. Можно быстро заменить выделение одного имени переменной другим, дважды щелкнув на поле ввода переменной в окне *Копировать переменные* для вывода списка переменных текущей таблицы исходных данных, в котором можно выбрать нужную переменную.

С переменной

Здесь указывается первая переменная (включительно) в копируемом блоке переменных. (Если вы перемещаете только одну переменную, то в списках *С переменной* и *По переменную* надо указать одно и то же значение.)

По переменную

Здесь указывается последняя переменная (включительно) в копируемом блоке переменных.

Вставить после

Здесь указывается место (переменная) в таблице исходных данных, после которого будут вставлены переменные (нельзя вставить блок переменных *после* переменной, которая включена в диапазон копируемых переменных). Копируемые данные будут помещены справа от указанной здесь переменной.

Как поместить блок до первой переменной. Чтобы скопировать и вставить значения переменных или блока переменных до первой переменной, можно воспользоваться возможностями метода *Перетащить и отпустить* системы *STATISTICA*. Для этого

Copyright © StatSoft, 1995

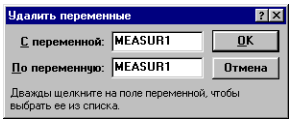
StatSoft

надо сначала выделить нужный блок, затем переместить курсор к одной из границ блока, где он изменится на стрелку. Теперь для того, чтобы скопировать и вставить блок, надо нажать клавишу SHIFT и перенести его. Далее надо установить полосу вставки между столбцом имен наблюдений и первым столбцом переменных и отпустить кнопку мыши. Обратите внимание, что этот метод копирует не выделенные переменные (т.е. их длинные метки значений, форматы, присвоения текстовых значений и т.д.), а только их значения из выделенного блока.

Другой способ скопировать и вставить переменную или блок переменных до первой переменной заключается в том, чтобы вставить переменные после первой переменной, а потом переместить первую переменную направо от перемещенных ранее переменных с помощью команды *Переместить переменные* (см. выше).

Удалить переменные

Одна или более переменных могут быть удалены из текущей таблицы исходных данных, если задать диапазон (включительно) удаляемых переменных в диалоговом окне *Удалить переменные* (доступно из выпадающего меню *Правка - Переменные* или через кнопку **Переменные** панели инструментов). Можно быстро заменить одну выбранную переменную другой, щелкнув на поле ввода переменной в диалоговом окне *Удалить переменные* и выбрав нужную переменную в списке переменных текущей таблицы исходных данных.



В отличие от удаления (очистки) содержания выделенного диапазона или блока значений в таблице исходных данных, эта операция удаляет и содержимое столбцов, и сами столбцы; поэтому последующие переменные будут сдвинуты налево, и размер файла уменьшится.

С переменной

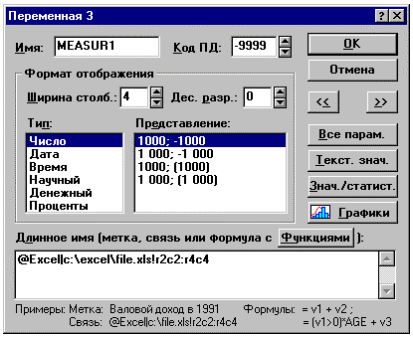
Здесь указывается первая переменная (включительно) в копируемом блоке переменных. (Если вы копируете только одну переменную, то в списках *С переменной* и *По переменную* надо указать одно и то же имя).

По переменную

Здесь указывается последняя переменная (включительно) в копируемом блоке переменных.

Изменение (текущих) спецификаций

Спецификации текущей выделенной переменной (т.е. ее имя и формат, значение пропущенных данных и метка, формула или связь) могут быть изменены в диалоговом окне *Переменная...*(*Изменение (текущих) спецификаций*) таблицы исходных данных.



Это диалоговое окно доступно из выпадающего меню *Правка - Переменные* таблицы исходных данных или через команду *Текущие спецификации*, вызываемую с помощью кнопки **Переменные** панели инструментов; кроме того, чтобы вызвать это окно, можно дважды

StatSoft

Copyright © StatSoft, 1995

щелкнуть на имени переменной в таблице исходных данных.

Имя

Здесь можно изменить имя переменной. Имена переменных могут иметь длину до 8 символов, не могут начинаться с символов: точка (.), минус (-), плюс (+) или с цифр (т.е. *0-9*) и не могут содержать пробелов, непечатаемых символов или символов, используемых для преобразования данных и при программировании на языках системы *STATISTICA* (**" # & ' ( ) + * , - . / ; : < = > [ ] ^ \ { } ~**). Если имя переменной будет содержать какие-либо из этих символов, то система *STATISTICA* заменит их символом подчеркивания [за исключением наблюдений, когда точка (.), минус (-), или плюс (+) использованы в качестве первого символа; в этих наблюдениях имя будет считаться неправильным, и будет выдано сообщение об ошибке]. Программа не различает верхний и нижний регистр букв текстовых значений (все буквы будут переведены в верхний регистр).

Код ПД

Здесь можно ввести значение кода пропущенных данных (значение переменной, которое указывается для точек, не содержащих данные) или принять код по умолчанию, который равен -9999. Значения пропущенных данных отображаются в ячейках таблицы исходных данных как пробелы и игнорируются при проведении любого анализа. Эти значения можно посмотреть окошке *Поле просмотра* панели инструментов таблицы исходных данных (см. главу 2), когда выделена пустая ячейка.

Формат отображения

Ширина столбца. Здесь можно изменить ширину столбца (от 1 до 30). По умолчанию ширина столбца составляет 8 символов. Формат отображения влияет только на то, как переменные отображаются в таблице исходных данных и других списках, а не на внутреннюю точность хранения числа.

Десятичные разряды. Здесь указывается число десятичных разрядов (меньшее, чем ширина столбца), которые будут отображаться в таблице исходных данных и других списках (применяется для типов *Число*, *Научный*, *Денежный* и *Проценты*, см. ниже). По умолчанию используется 3 десятичных разряда. Формат отображения влияет только на то, как переменные отображаются в таблице исходных данных и других списках, а не на внутреннее представление чисел. Для переменных в формате *Дата* (см. раздел *Операции с датами*, стр. 1180) выбор типа *Число* приведет к отображению дат в юлианском формате (например, *24858* вместо *1/21/1968*). Для переменных в формате *Время* выбор типа *Число* приведет к отображению времени в десятичном виде (например, *24858.9375* вместо *1/21/1968 10:30 PM*, или *.78125* вместо *18:45*)

Тип/Представление. Для каждого выбранного *Типа* также можно выбрать *представление*, т.е. формат отображения значений текущей переменной.

Число. Этот формат используется для переменных, значения которых числовые или числовые с текстовым эквивалентом. Выбрав этот тип, можно задать число десятичных разрядов, которые будут отображаться в таблице исходных данных, с помощью параметра *Дес. разр.* (см. выше).

Представление. Эти значения будут представлены в десятичном виде с запятыми или без них (например, *1000* или *1,000*). Отрицательное значение переменной может представляться либо знаком "минус" (например, *-1000*), либо заключением числа в скобки (например, *(1000)*). В этой категории доступны четыре вида представления:

Copyright © StatSoft, 1995

1000.00	-1000.00
1,000.00	-1,000.00
1000.00	(1000.00)
1,000.00	(1,000.00)

Обратите внимание, что система *STATISTICA* поддерживает европейское представление чисел (например, 1,20), если оно выбрано в окне *Панель управления* Windows. Если в окне *Панель управления* этот стиль не выбран и будет набрана десятичная запятая, то система *STATISTICA* интерпретирует это значение как текстовое.

Дата. Этот формат используется для интерпретации переменных как дат (например, переменные, содержащие даты в юлианском формате, импортированные или вставленные из программы Excel). Обратите внимание, что для того, чтобы отображение даты было правильным, все значения переменной должны содержать правильные даты или точно определенные текстовые значения (см. также *Операции с датами*, стр. 1180).

Представление. Здесь можно выбрать нужное представление дат из списка форматов дат. В окне *Спецификации переменных* (см. стр. 1178), различные представления *дат* обозначены следующими ключевыми словами (которые также могут быть введены непосредственно в столбце *Формат* этого окна):

3/6/87	DATE1
03/06/1987	DATE2
6-Mar-87	DATE3
6-Mar	DATE4
Mar-1987	DATE5
06/03/87	DATE6
Текущий формат Windows	DATE7

Обратите внимание, что если выбран установленный формат Windows, система *STATISTICA* представит дату, используя спецификации диалогового окна *Международные* (Windows 3.1) или *Язык и стандарты* (Windows 95) в окне *Панель управления*.

Время. Значения этого типа переменных отображаются в одном из форматов времени. Значения времени переменных сохраняются в данных как (необязательные) десятичные значения, представляющие часть дня, прошедшую с полуночи; например, *6:00AM* хранится как *0,25*. Значения времени, хранимые таким образом, могут быть использованы в последующем анализе и преобразованы с помощью арифметических операций; в то же время они могут быть отображены как время в отчетах или графиках (например, для обозначений на осях). Система *STATISTICA* произведет форматирование отображения значений времени в соответствии с текущими установками параметра *Международный формат времени* в окне *Панель управления* Windows.

Представление. Время может быть представлено в нескольких форматах. Можно выбрать комбинацию даты и времени [в 24-часовом или 12-часовом (AM/PM) формате] или просто время [в 24-часовом или 12-часовом (AM/PM) формате] в часах и минутах или часах, минутах и секундах. Также можно выбрать формат Windows для того, чтобы использовать текущие параметры в диалоговом окне *Международные* (Windows 3.1) или *Язык и стандарты* (Windows 95) в окне *Панель управления* Windows. В окне *Спецификации переменных* для представления времени используются следующие ключевые слова (которые также могут быть введены непосредственно в столбец *Формат* этого окна):

3/6/87 10:30 PM	TIME1
87/3/6 22:30	TIME2
10:30 PM	TIME3
22:30	TIME4
10:30:25 PM	TIME5
22:30:25	TIME6

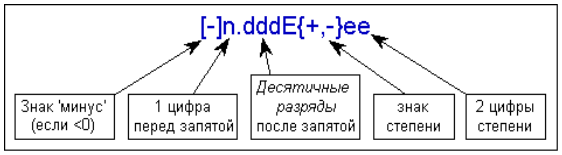
StatSoft®

Copyright © StatSoft, 1995

Текущий формат	
Windows	TIME7

Научный. Значения этого типа переменных отображаются в научной записи. Выбрав этот тип, можно указать число десятичных знаков, отображаемых в таблице исходных данных, с помощью параметра *Дес. разр.* (см. выше).

Представление. В первом представлении (стандартная научная запись) все числа столбца будут отображаться в научной записи в соответствии со следующими правилами:



Например, 1.2345E-02 или -3.2100E+08.

Второе представление будет касаться только тех (маленьких) значений переменных, которые в формате *Число* были бы отображены как 0 (ноль) [т.е., числа из интервала (-1,1), значащие знаки которых начинаются после позиции, указанной в поле ввода *Дес. разр.*]. Например, если *Дес. разр.* = 4, а значение переменной 0.0000123, то в формате числа значение будет отображаться как 0.0000; тогда как во втором научном представлении число будет отображено как 1.230E-5 (отображаются 4 знака перед E).

Денежный. Значения этого типа переменных отображаются в денежном формате. Выбрав эту категорию, можно указать число десятичных разрядов, отображаемых в таблице исходных данных, с помощью параметра *Дес. разр.* (см. выше).

Представление. Денежные значения отображаются с денежным символом до или после числа (в зависимости от конфигурации, установленной на компьютере пользователя; например, $, DM и т.п., см. *Панель управления* Windows). Можно выбрать представление отрицательных значений переменных либо со знаком “минус” (например, $*-1,000*), либо с заключением числа в скобки (например *($1,000)*).

Процентный. Значения этого типа переменных отображаются как проценты. Выбрав этот тип, можно указать число десятичных знаков, отображаемых в таблице исходных данных, с помощью параметра *Дес. разр.* (см. выше). Процент считается как доля от единицы, которую представляет значение переменной. Например, значение 0,23 будет переведено в 23%, тогда как 23,4 будет переведено в 2340%.

Представление. Проценты отображаются с символом % после числа.

Длинное имя - Метка

Каждой переменной можно присвоить длинное имя, которое может содержать до 128 символов. Метка может содержать примечания, включая все печатаемые символы.

Длинное имя - Формула

Формулы должны начинаться со знака равенства. Если будет введена метка, начинающаяся со знака равенства, система *STATISTICA* будет считать, что это формула, и проверит ее на формальную правильность. Если формула формально правильна, будет предоставлена возможность пересчитать переменную сейчас или позже (с помощью диалогового окна *Пересчитать*, стр. 1184). Формулы можно использовать для проверки данных, изменения переменной, перекодирования переменной или создания значений переменных, основанных на локальных условиях (например, формула *=(v0<=100)*1 + (v0>100)*2* присваивает значение *1* наблюдениям от *1* до *100* и *2* наблюдениям больше *100*). Ссылки на переменные могут производиться по их именам (например, *Test1*, *Income*) или номерам

Copyright © StatSoft, 1995

(например, *v1*, *v2*, *v3*, ...); *v0* – номер наблюдения. К формулам могут быть добавлены комментарии, начинающиеся с точки с запятой. Для подробной информации о синтаксисе формул см. раздел *Обзор синтаксиса формул системы STATISTICA* (стр. 1176).

Примеры:

Эта формула вычисляет среднее для первых трех переменных.

```
=(v1+v2+v3)/3
```

Эта формула перекодирует наблюдения 1-10 в *1*, а остальные - в *2* (после нее написан комментарий).

```
= (v0<=10)*1+(v0>10)*2; this is a comment
```

Обратите внимание, что можно нажать кнопку Функциями в этом диалоговом окне, чтобы открыть диалоговое окно *Мастер функций*. В этом диалоговом окне можно выбрать нужную часть формулы (например, оператор, распределение, математическую функцию). При нажатии в этом диалоговом окне кнопки *OK*, выделенная часть будет вставлена в поле ввода, начиная с текущей позиции курсора.

Длинное имя - Связь

Кроме *примечаний* и *формул*, длинные имена переменных также могут содержать спецификации связи. В соответствии с соглашениями MS Windows о динамическом обмене данными (*DDE*), можно связать переменную (или блок данных) системы *STATISTICA* с переменной (или блоком) в электронной таблице другой программы, что позволяет автоматически обновлять таблицу при изменении файла-“источника” (сервера). Проще всего установить связь *DDE* между системой *STATISTICA* и другим приложением через стандартные для Windows команды *Установить связь* (стр. 1197) и *Связи* (стр. 1206) выпадающего меню *Правка*. Кроме того, оператор связи может также быть введен непосредственно в поле *Длинное имя* этого диалогового окна [см. раздел *Создание связей (динамический обмен данными)* в главе 7 для подробной информации о шагах установления связи *DDE*, а также соглашения о синтаксисе и описание окна *Диспетчер связей* для информации об управлении связями *DDE*].

Пример установки связи (присоединение блока данных системы *STATISTICA* к блоку данных программы Excel): *@Excel|c:\file.xls!r2c2:r4c4*. В этом примере устанавливается динамическая связь с блоком “от строки 2, столбца 2 до строки 4, столбца 4” из файла электронной таблицы *File.xls* программы Microsoft Excel (убедитесь, что вы указали полный путь, задавая имя файла в операторе связи, т.е. *c:\Excel\File.xls*).

Все параметры

Нажатие этой кнопки приведет к закрытию этого диалогового окна (с сохранением всех изменений) и открытию диалогового окна *Переменные…*, стр. 1178), которое отображает спецификации всех переменных в файле данных.

Текстовые значения

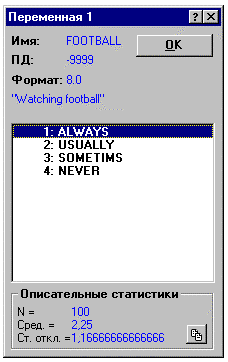
Нажатие этой кнопки приведет к закрытию этого диалогового окна (с сохранением всех изменений) и открытию диалогового окна *Диспетчер текстовых значений* (стр. 1178).

Значения/статистики

Нажатие этой кнопки приведет к открытию окна *Переменная…*,

StatSoft

Copyright © StatSoft, 1995



в котором можно посмотреть значения переменной, а также некоторые ее описательные статистики (число значений *N*, среднее и стандартное отклонение).

Графики

Нажатие этой кнопки приведет к закрытию этого диалогового окна (с сохранением всех изменений) и появлению меню *Быстрые статистические графики* (см. стр. 1219 этой главы для краткого обзора; подробную информацию можно найти в томе II настоящего руководства).

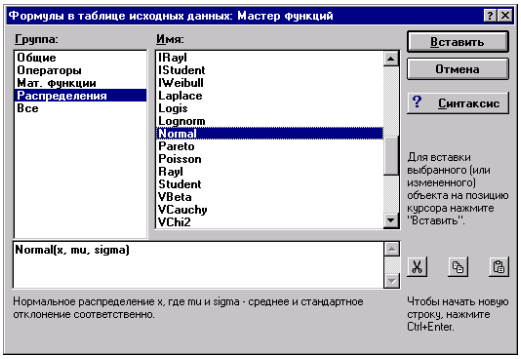
Кнопки << , >>

Нажатие этих кнопок приведет к открытию диалогового окна *Переменная…*(*текущие спецификации*) для предыдущей (<<) или последующей (>>) переменной. Обратите внимание, что когда вы нажимаете одну из этих кнопок, система *STATISTICA* применит сделанные для текущей переменной изменения перед тем как перейти к диалоговому окну следующей или предыдущей переменной. Если вы не хотите сохранять изменения, нажмите кнопку *Отмена* в текущем диалоговом окне, а потом откройте диалоговое окно *Переменная…* (*текущие спецификации*) для нужной переменной.

Формулы в таблице исходных данных: Мастер функций

После нажатия кнопки Функциями в диалоговом окне *Переменная…* (*текущие спецификации*, см. стр. 1170), откроется диалоговое окно *Формулы в таблице исходных данных: Мастер функций*. Здесь в первом списке можно выделить *группу,* т.е. список объектов которой будет выведен в поле *Имя*.

Если во втором списке выделить *имя*, в нижней части этого диалогового окна будет отображен объект и его краткое описание (см. ниже).



Нажатие кнопки *Вставить* в этом диалоговом окне приведет к тому, что выбранный объект будет вставлен в поле ввода диалогового окна *Переменная* (*текущие спецификации*) на текущую позицию курсора.

В этом диалоговом окне доступны следующие группы формул (для более подробной информации об элементах формул, доступных в этом диалоговом окне, см. ниже раздел *Обзор синтаксиса*):

Общие: =, *Vxxx*, *V0*

Операторы: <, >, <>, <=, >=, +, -, *, /, **, ^, *and*, *not*, *or*

Мат. функции: *Abs*, *ArcSin*, *ArcCos*, *ArcTan*, *Cos*, *CosH*, *Exp*, *Euler*, *Hypot*, *Log*, *Log10*, *Log2*, *Max*, *Min*, *Normal*, *Pi*,

Copyright © StatSoft, 1995

StatSoft

Poisson, *Rnd*, *Round*, *Sign*, *Sin*, *SinH*, *Sqrt*, *Tan*, *TanH*, *Trunc*, *Uniform*

Распределения: *Beta*, *iBeta*, *vBeta*, *Binom*, *iBinom*, *Cauchy*, *iCauchy*, *vCauchy*, *Chi2*, *iChi2*, *vChi2*, *Expon*, *iExpon*, *vExpon*, *Extreme*, *iExtreme*, *vExtreme*, *F*, *iF*, *vF*, *Gamma*, *iGamma*, *vGamma*, *Geom*, *iGeom*, *Laplace*, *iLaplace*, *vLaplace*, *Logis*, *iLogis*, *vLogis*, *Lognorm*, *iLognorm*, *vLognorm*, *Normal*, *iNormal*, *vNormal*, *Pareto*, *iPareto*, *vPareto*, *Poisson*, *iPoisson*, *Rayleigh*, *iRayleigh*, *vRayleigh*, *Student*, *iStudent*, *vStudent*, *Weibull*, *iWeibull*, *vWeibull*

Все: выводит весь список приведенных выше формул.

Обзор синтаксиса формул системы *STATISTICA*

Ниже приведен обзор соглашений о синтаксисе формул для таблиц исходных данных, которые могут быть введены в диалоговом окне таблицы исходных данных *Переменная…*(*Изменение (текущих) спецификаций*).

Имена переменных в формулах. Ссылаться на переменные можно по их номерам или именам (*v0* – номер наблюдения). Например:

```
= v1 - v2
```

```
= Retail - Cost
```

Комментарии в формулах. Комментарии в формуле начинаются с точки с запятой. Например:

```
= v1 + v2 ;
```
это комментарий

Операторы условий. В формулы таблиц исходных данных можно вводить логические выражения. Логическое выражение равняется *1,* если оно верно, и *0*, если оно ложно. Поэтому операторы условий могут состоять из логических выражений, заключенных в скобки. Например, выражение

```
=(v0<=10)*1 + (v0>10)*2
```

кодирует наблюдения 1-10 как *1*, а остальные как *2*. В операторах условий могут быть использованы следующие логические операторы:

`AND, &`	логическое И
`OR, |`	логическое ИЛИ
`NOT, ~`	логическое НЕТ

Например, выражение

```
=((v1=1) AND (v2=5))*5
```

возвращает значение *5,* если v1=1 и v2=5, и *0* в противном случае.

```
=(((v1=1) OR (v2=5))<>0)*5
```

возвращает значение *5,* если одно *или* оба условия соблюдены, и *0* в противном случае.

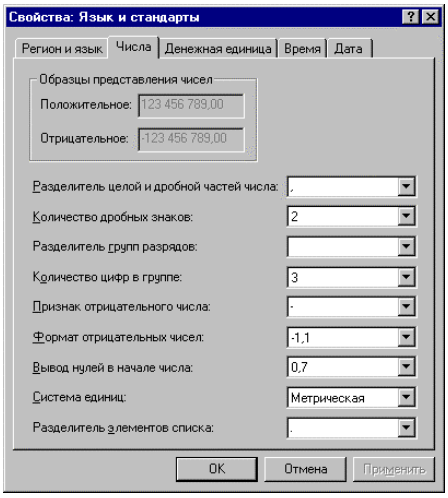
Разделители списка. Некоторые функции формул таблиц исходных данных требуют более одного параметра (например, *Normal(x,0,1)*), которые отделяются друг от друга *разделителями списка.* Система *STATISTICA* использует разделитель списка, указанный в диалоговом окне *Международные* из блока *Панель управления* Windows. Например, для разделения списка в США чаще всего используют запятую (,). Поэтому списки чисел, имен и т.д. будут разделены запятыми. Например, в следующей формуле параметры нормального распределения разделены запятыми:

```
=normal(v2,0,1)
```

Если разделитель списка изменился (например, на используемый в другой стране):

StatSoft®

Copyright © StatSoft, 1995



то предыдущая формула будет записана так:

```
=normal(v2;0;1)
```

При изменении разделителя списка убедитесь, что выбранный вами разделитель отличается от разделителя десятичных разрядов.

Операторы. Операторы, которые можно использовать в формулах таблиц исходных данных, приведены ниже в восходящем порядке (порядке выполнения), от низшего к высшему (для изменения порядка выполнения могут быть использованы скобки):

`=, <, >, <>, <=, >=`	- операторы отношений
+	- сложение
-	- вычитание
*	- умножение
/	- деление
**, ^	- возведение в степень

Константы. В формулах таблиц исходных данных поддерживаются следующие константы:

`Pi`	- π (3.14159265358979)
`E`	- *e* (2.71828182845905)

Функции. В формулах таблиц исходных данных поддерживаются следующие функции:

`Abs(x)`	- Модуль *x*
`Arcsin(x)`	- Арксинус *x*
`Cos(x)`	- Косинус *x*
`Exp(x)`	- *e* в степени *x*
`Log(x)`	- Натуральный логарифм *x*
`Log2(x)`	- Логарифм *x* по основанию 2
`Log10(x)`	- Десятичный логарифм *x*
`Rnd(x)`	- Произвольное число, равномерно распределенное на интервале от *0* до *x*
`Sign(x)`	- Знак *x*: если *x*>0 то *+1*, если *x*<0 то *-1*, если *x*=0 то *0*
`Sin(x)`	- Синус *x*
`Sinh(x)`	- Гиперболический синус *x*
`Sqrt(x)`	- Квадратный корень из *x*
`Tan(x)`	- Тангенс *x*
`Trunc(x)`	- Целая часть *x*

Функции распределений, их интегралы и обратные функции распределений. Система *STATISTICA* предоставляет широкий выбор заданных функций распределений, их интегралов и обратных функций распределений, которые могут быть использованы в формулах таблиц исходных данных, как и все прочие функции. Полный перечень всех доступных распределений и формул для них см. в главе *Приложение I* этого тома руководства.

Распределение	Плотность или функция вероятности	Функция распределения	Обратная функция распределения
Бета	beta(x,ν,ω)	ibeta(x,ν,ω)	vbeta(x,ν,ω)
Биномиал.	binom(x,p,n)	ibinom(x,p,n)	
Коши	cauchy(x,η,θ)	icauchy(x,η,θ)	vcauchy(x,η,θ)

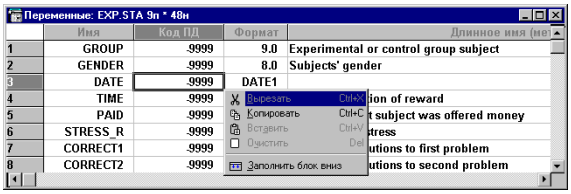
Copyright © StatSoft, 1995

StatSoft

Хи-квадр.	chi2(x,ν)	ichi2(x,ν)	vchi2(x,ν)
Экспонен.	expon(x,λ)	iexpon(x,λ)	vexpon(x,λ)
Экстрем.	extreme(x,a,b)	iextreme(x,a,b)	vextreme(x,a,b)
F	F(x,ν,ω)	iF(x,ν,ω)	vF(x,ν,ω)
Гамма	gamma(x,c)	igamma(x,c)	vgamma(x,c)
Геометрич.	geom(x,p)	igeom(x,p)	
Лапласа	laplace(x,a,b)	ilaplace(x,a,b)	vlaplace(x,a,b)
Логистич.	logis(x,a,b)	ilogis(x,a,b)	vlogis(x,a,b)
Логнорм.	lognorm(x,μ,σ)	ilognorm(x,μ,σ)	vlognorm(x,μ,σ)
Нормальное	normal(x,μ,σ)	inormal(x,μ,σ)	vnormal(x,μ,σ)
Парето	pareto(x,c)	ipareto(x,c)	vpareto(x,c)
Пуассона	poisson(x,λ)	ipoisson(x,λ)	
Релея	rayleigh(x,b)	irayleigh(x,b)	vrayleigh(x,b)
Стьюдента	student(x,df)	istudent(x,df)	vstudent(x,df)
Вейбулла	weibull(x,b,c,θ)	iweibull(x,b,c,θ)	vweibull(x,b,c,θ)

Все спецификации

Выбрав эту команду, можно просмотреть спецификации всех переменных таблицы исходных данных (доступна из выпадающего меню *Правка - Переменные* или через кнопку **Переменные** панели инструментов).



В таблицу спецификаций для каждой переменной включаются ее номер и имя, код пропущенных данных, формат значения и длинное имя или формула. Спецификации переменной могут быть изменены здесь или в диалоговом окне *Переменная…*таблицы исходных данных (стр. 1170).

Кнопка на панели инструментов. Это окно также может быть вызвано нажатием кнопки панели инструментов таблицы исходных данных.

Изменение текстовых значений

Для изменения текстовых значений (и эквивалентного им числового представления) переменной, выделенной в таблице исходных данных, используется диалоговое окно *Диспетчер текстовых значений*. Его можно вызвать из выпадающего меню *Правка – Переменные*, кнопкой **Переменные** панели инструментов или кнопкой *Текстовые значения* в диалоговом окне таблицы исходных данных *Переменная…*(*Изменение (текущих) спецификаций*, см. стр. 1170). В диалоговом окне *Диспетчер текстовых значений* можно управлять соотношениями между числовыми данными, текстовыми данными и метками значений. Значения могут быть введены, отсортированы, изменены, скопированы из переменной в переменную и т.д.

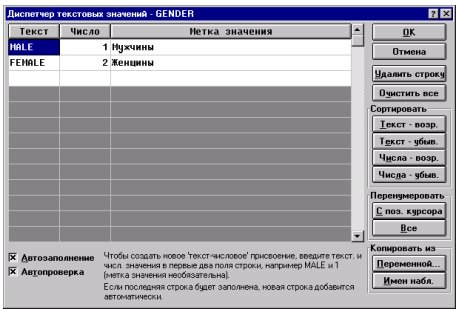
Вызов. Дважды щелкните на имени переменной (столбца), затем в появившемся диалоговом окне нажмите кнопку *Текст. знач.*. Для вызова этого диалогового окна можно использовать и другие способы: щелкните правой кнопкой и выберите пункт *Текстовые значения* из контекстного меню или нажмите кнопку (*Диспетчер текстовых значений*) панели инструментов таблицы исходных данных.

Пример. В этом примере в переменной *Gender* будут изменены текстовые значения (*Male* и *Female*).

Шаг1. Щелкните правой кнопкой мыши на любом значении переменной *Gender* в таблице исходных данных и выберите пункт *Текстовые значения…* из контекстного меню, чтобы открыть диалоговое окно *Диспетчер текстовых значений*.

StatSoft

Copyright © StatSoft, 1995



Шаг 2. Чтобы изменить текстовое значение *Male* на *Boys*, щелкните на значении *Male* в переменной *Gender* в диалоговом окне *Диспетчер текстовых значений* и впишите значение *Boys*.

Шаг 3. Нажмите клавишу TAB, чтобы перейти в следующее поле (числовое значение, связанное с *Boys*), нажмите TAB еще раз, чтобы перейти в поле *Метка значения*. Теперь можно ввести короткую метку, описывающую значение *Boys* (например, введите текстовую строку *Boys in single parent homes*).

Шаг 4. Теперь, когда вы изменили текстовое значение и ввели метку значения, нажмите кнопку *OK* , чтобы закрыть это диалоговое окно и применить изменения переменной.

Новое “текст-числовое” присвоение

Введите два соответствующих значения (текстовое, числовое) в первые два поля строки, например, *MALE* и *1* (*метка значения* необязательна).

Изменение “текст-числового” присвоения

Нажмите клавишу TAB, чтобы перейти в следующее поле (или строку); содержание последней строки будет скопировано автоматически (его можно изменить или набирать поверх него). Чтобы изменить поле, можно набрать поверх него; чтобы исправить поле, можно нажать F2 или дважды щелкнуть на нем. Клавишу INSERT можно использовать для вставки новой строки после текущей, а клавишу DELETE - для удаления текущей строки.

Текст

Здесь можно управлять текстовыми значениями. Текстовые значения могут быть длиной не более 8 символов и могут содержать любые символы. Обратите внимание, что пробелы в начале и конце текстовых значений будут удалены программой. Система *STATISTICA* различает верхний и нижний регистр текстовых значений (например, “Smith” не эквивалентно “SMITH”).

Переключаться между просмотром текстовых значений и их числовых эквивалентов таблицы исходных данных можно, нажимая кнопку на панели инструментов.

Число

Здесь можно управлять числовыми значениями, которые присвоены текстовым. Двойная запись переменных (когда каждое значение может одновременно иметь и текстовое, и числовое представление) позволяет автоматически использовать числовые значения, если для анализа выбрана переменная с текстовыми значениями.

Метка значения

Каждое текстовое значение переменной может иметь присвоенную ему метку. В этом поле можно вводить метки и управлять ими. Метки значений могут быть длиной до 40 символов и содержать примечания, включая любые печатаемые символы. Метки значений автоматически используются в выводе результатов анализа, где упоминаются

Copyright © StatSoft, 1995

индивидуальные значения (например, частота, кросстабуляция, распад и т.д.).

Чтобы вывести метки значений в таблице исходных данных, можно нажать кнопку , удерживая при этом нажатой клавишу CTRL, или выбрать пункт *Текстовые значения – Метки значений* выпадающего меню *Вид* таблицы исходных данных.

Чтобы напечатать метки значений вместе с данными таблицы исходных данных, надо выбрать пункт *Текстовые значения* в диалоговом окне *Печать данных* или выбрать *полный* тип отчета в диалоговом окне *Параметры страницы/вывода* (см. стр. 1157).

Удалить строку

Эта кнопка позволяет удалить целую строку из списка текстовых значений.

Очистить все

С помощью этой кнопки можно очистить все записи во всех полях диалогового окна *Диспетчер текстовых значений*.

Сортировать: Текст - возрастание

Эта кнопка предназначена для сортировки полей *Текст*, *Число* и *Метка значения* в диалоговом окне *Диспетчер текстовых значений* по алфавиту в возрастающем порядке с использованием для сортировки поля *Текст*.

Сортировать: Текст - убывание

Эта кнопка предназначена для сортировки полей *Текст*, *Число* и *Метка значения* в диалоговом окне *Диспетчер текстовых значений* по алфавиту в убывающем порядке с использованием для сортировки поля *Текст*.

Сортировать: Числа - возрастание

Эта кнопка предназначена для сортировки полей *Текст*, *Число* и *Метка значения* в диалоговом окне *Диспетчер текстовых значений* - по алфавиту в возрастающем порядке с использованием для сортировки поля *Число.*

Сортировать: Числа - убывание

Эта кнопка предназначена для сортировки полей *Текст*, *Число* и *Метка значения* в диалоговом окне *Диспетчер текстовых значений* по алфавиту в убывающем порядке с использованием для сортировки поля *Число*.

Копировать из

Переменной. Эта кнопка открывает стандартное диалоговое окно выбора переменной, в котором можно выбрать переменную, текстовые или числовые значения которой нужно скопировать.

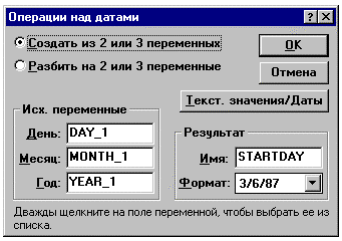
Имен наблюдений Эта кнопка доступна (активна) только если текущий файл данных содержит имена наблюдений. Она позволяет копировать эти имена наблюдений в столбец *Текст* диалогового окна *Диспетчер текстовых значений* и присваивает соответствующие целые значения в столбце *Число*.

Операции над датами

Можно создать переменную, содержащую дату, из двух или трех переменных (содержащих месяц и год, день и месяц, или месяц, день и год) или разделить одну переменную на две или три переменные (по одной на месяц и год, день и месяц или месяц, день и год) с помощью диалогового окна *Операции над датами*.

StatSoft

Copyright © StatSoft, 1995



Это диалоговое окно вызывается с помощью пункта *Значения дат* выпадающего меню *Правка - Переменные* таблицы исходных данных или кнопки Переменные панели инструментов (см. примеры ниже).

Создать из 2 или 3 переменных

В этом режиме можно создавать значения дат (в новой переменной) из значений дня, месяца и года трех существующих переменных или из значений месяца и года или дня и месяца двух существующих переменных.

Разбить на 2 или 3 переменные

Этот режим позволяет разделить одно значение даты (из существующей переменной) на отдельные значения дня, месяца и года по трем (ранее созданным) новым переменным или на отдельные значения месяца и года, или дня и месяца по двум (ранее созданным) новым переменным.

Текстовые значения/Даты

Нажатие этой кнопки приведет к открытию диалогового окна *Текстовые значения/даты* (см. ниже). В этом диалоговом окне можно поменять даты на текстовые метки и наоборот (например, для того, чтобы использовать переменные, содержащие дату, в качестве категориальных переменных) (см. стр. 1182 для подробной информации).

Исходные переменные

Переменные, которые можно выбрать здесь, зависят от того, собираетесь ли вы создать или разделить переменную.

Создать из 2 или 3 переменных. Выбрав этот вариант, вы должны будете указать в полях *Исх. переменные* три переменные, содержащие значения *День, Месяц* и *Год* (см. *Пример 1*, ниже), или две переменные, содержащие *День* и *Месяц* (в качестве года система *STATISTICA* введет текущий год) или *Месяц* и *Год* (в качестве дня система *STATISTICA* введет первый день месяца).

Разбить на 2 или 3 переменные. Выбрав этот вариант, вы должны будете указать в поле *Исх. переменная* переменную, содержащую полное значение даты.

Результаты

Переменные, которые можно выбрать здесь, зависят от того, собираетесь ли вы создать или разделить переменную.

Создать из 2 или 3 переменных. Выбрав этот вариант, вы должны будете указать в поле *Результат* имя переменной, в которую будет помещено новое значение даты (см. *Пример 1*, ниже). Вы также должны будете указать формат даты, выбрав его из выпадающего списка *Формат*.

Разбить на 2 или 3 переменные. Выбрав этот вариант, вы должны будете указать в поле *Результат* ранее созданные переменные, в которые будут помещены значения *День*, *Месяц* и *Год* (для 3-х переменных, см *Пример 2*, ниже), *День* и *Месяц* или *Месяц* и *Год* (для 2-х переменных) или *День, Месяц* или *Год* (для 1-ой переменной).

Copyright © StatSoft, 1995

Пример 1 (создание даты)

В этом примере из трех переменных с помощью файла данных *Heart.sta* системы *STATISTICA* будет создана переменная, содержащая дату. Диалоговое окно *Операции над датами* можно вызвать из пункта *Переменные – Значения дат* выпадающего меню *Правка* или нажатием кнопки **Переменные** панели инструментов и выбором пункта *Значения дат*.

Шаг 1. Сначала нужно создать новую переменную, которая будет содержать дату. Для этого надо выбрать пункт *Переменные - Добавить* выпадающего меню *Правка* и указать одну новую переменную, добавляемую к файлу данных.

Для подробной информации о том, как добавить переменные, см. раздел *Добавить переменные*, стр. 1167; Для подробной информации о том, как изменить имя переменной и другие спецификации, см. пункт *Изменение (текущих) спецификаций*, стр. 1170.

Шаг 2. Теперь надо открыть диалоговое окно *Операции над датами* (см. выше) и выбрать вариант *Создать из 2 или 3 переменных*.

Шаг 3. В диалоговом окне *Операции над датами* надо указать 3 переменные с датами (поле *Исх. переменные*), дважды щелкая на полях ввода *День*, *Месяц* и *Год* и выбирая нужную переменную из списка переменных файла данных (для подробной информации о выборе переменных, см. раздел *Выбор переменной из списка* в главе 1).

Шаг 4. Надо указать имя переменной, в которую будет помещено значение даты (поле *Результат*), а также формат для отдельных значений. Сначала нужно дважды щелкнуть на поле ввода переменной и выбрать новую переменную, созданную ранее для размещения дат. Далее надо выбрать формат значения из выпадающего списка *Формат*. Позднее формат значения даты может быть изменен в диалоговом окне *Переменная…*(*Изменение (текущих) спецификаций,* стр. 1170).

Шаг 5. Нажмите *OK* в диалоговом окне *Операции над датами* для завершения создания из трех переменных одной переменной, содержащей дату.

Пример 2 (разделение переменной, содержащей дату)

В этом примере из переменной, содержащей дату, с помощью файла данных *Heart.sta* системы *STATISTICA* и новой переменной, созданной в *Примере 1* будут созданы три переменные, содержащие месяц, день и год.

Шаг 1. Нужно создать три новые переменные, которые будут содержать день, месяц и год соответственно (см. *Пример 1, Шаг 1* для информации о создании новых переменных).

Шаг 2. Надо открыть диалоговое окно *Операции над датами* (см. выше) и выбрать вариант *Разбить на 2 или 3 переменные*.

Шаг 3. В диалоговом окне *Операции над датами* в поле *Исх. переменная* теперь указана одна переменная (а не три, как в предыдущем примере). Чтобы выбрать переменную, содержащую дату, из списка переменных файла данных, нужно дважды щелкнуть на поле ввода переменной (см. *Пример 1*, *Шаг 3*).

Шаг 4. Нужно указать переменную в поле *Результат*, дважды щелкая на полях

StatSoft®

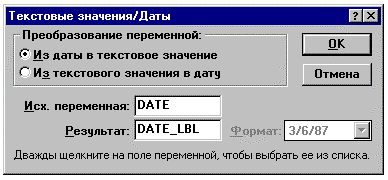
Copyright © StatSoft, 1995

ввода *День*, *Месяц* и *Год* и выбирая ранее созданные переменные (см. *Шаг 1*).

Шаг 5. Нажмите *OK* в диалоговом окне *Операции над датами* для завершения разделения переменной, содержащей дату, на три переменные.

Текстовые значения-/Даты

В некоторых случаях бывает полезным создавать текстовые значения, содержащие информацию о дате (например, когда переменная, содержащая дату, используется в качестве кодирующей переменной).



В диалоговом окне *Текстовые значения/Даты* можно преобразовать переменную, содержащую дату, в переменную, содержащую текстовые значения с числовыми эквивалентами в диапазоне, позволяющем системе *STATISTICA* использовать их как группирующие коды (т.е., числовые значения менее 32000). Также можно преобразовать переменную, содержащую текстовые значения дат в переменную, содержащую дату в нужном формате.

Преобразование переменной

Здесь можно выбрать тип преобразования переменной.

Из даты в текстовое значение . В этом случае переменная, содержащая дату, преобразуется в переменную, содержащую дату в виде текстовых значений (см. выше раздел *Операции над датами*). Итоговая дата в виде текстового значения будет в формате ММММДД_ГГ. Например, дата *6-Янв-68* (в юлианском формате *24843*) будет преобразована в текстовое значение *ЯНВ06_68* (с числовым эквивалентом *4843*).

Из текстового значения в дату . Этот метод позволяет преобразовывать переменные с датами в виде текстовых значений в переменные, содержащие дату. Например, текстовое значение даты *МАЙ25_80* будет преобразовано в дату заданного формата *05/25/80*. Чтобы система *STATISTICA* распознала текстовое значение даты, оно должно быть введено в формате ММММДД_ГГ. При преобразовании этого текстового значения в дату можно выбрать нужный формат даты (см. ниже).

Исходная переменная

Выбираемая здесь переменная будет зависеть от типа производимого преобразования.

Если производится преобразование *из даты в текстовое значение*, то здесь выбирается переменная с форматированным значением даты.

Если производится преобразование *из текстового значения в дату*, то здесь выбирается переменная, содержащая дату в виде текстового значения.

Результат

В качестве конечной переменной (переменной, в которую будут помещены преобразованные значения) может быть выбрана либо исходная переменная (т.е., преобразованные значения заменят начальные) либо другая переменная, в которую будут помещены преобразованные значения.

Формат

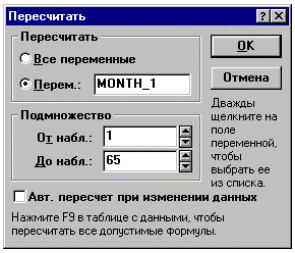
Если производится преобразование переменной с датой в виде текстового значения в переменную,

Copyright © StatSoft, 1995

содержащую дату, то из этого выпадающего списка можно выбрать формат конечной даты.

Пересчитать

Переменная (или группа переменных), основанная на формуле, введенной в диалоговом окне таблицы исходных данных *Переменная…*(*Изменение (текущих) спецификаций*, стр. 1170), может быть пересчитана. Это диалоговое окно доступно из выпадающего меню *Правка – Переменные* или с помощью кнопки **Переменные** панели инструментов.



Кнопка на панели инструментов. Кнопка панели инструментов таблицы исходных данных.

Комбинация клавиш. Нажатие кнопки F9 приведет к пересчету всех формул текущего набора данных без вывода диалогового окна (это стандартная кнопка *пересчета* в таблице исходных данных).

Все переменные

Этот режим предназначен для одновременного пересчета всех переменных, содержащих формулы.

Переменная

Чтобы пересчитать одну переменную, нужно указать ее имя в этом поле. После двойного щелчка на поле ввода переменной будет выведен список текущих переменных таблицы исходных данных.

Подмножество

По умолчанию будут пересчитываться все наблюдения (в показанные ниже поля будут автоматически введены первый и последний номера наблюдения в наборе данных) или, если в таблице исходных данных выделен блок, будут пересчитаны только наблюдения, включенные в этот блок (в показанные ниже поля будут автоматически введены первый и последний номера наблюдений в блоке).

Однако, если необходимо пересчитать наблюдения, отличные от введенных по умолчанию, то необходимо ввести номера соответствующих наблюдений в полях *От набл.* и *До набл.*.

От наблюдения. Здесь указывается первое наблюдение (включительно) в блоке пересчитываемых наблюдений.

До наблюдения. Здесь указывается последнее наблюдение (включительно) в блоке пересчитываемых наблюдений.

Автоматический пересчет при изменении данных

Этот режим предназначен для того, чтобы автоматически пересчитывать все формулы таблицы исходных данных при изменении данных в таблице.

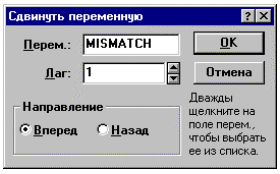
Сдвинуть (лаг)

Эта операция (доступная из выпадающего меню *Правка – Переменные* или через кнопку **Переменные** панели инструментов) предназначена для сдвига выбранных переменных *назад* (“вверх”) или *вперед* (“вниз”)

StatSoft®

Copyright © StatSoft, 1995

относительно остальных переменных в файле на заданное число наблюдений (*лаг*).



Система *STATISTICA* добавит пустые наблюдения к оставшимся переменным в файле либо до первого наблюдения (если переменная сдвигается *назад*), либо после последнего наблюдения файла данных (если переменная сдвигается *вперед*).

Переменная

Здесь указывается сдвигаемая переменная. После двойного щелчка на поле ввода переменной будет выведен список текущих переменных таблицы исходных данных.

Лаг

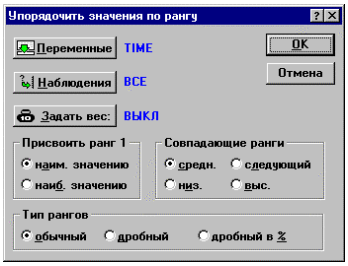
Укажите величину сдвига (т.е., число наблюдений, на которое сдвигается переменная).

Направление

Выбор направления сдвига переменной осуществляется выбором соответствующего положения переключателя *вперед* (“вниз”) или *назад* (“вверх”).

Ранжировать

В этом диалоговом окне можно ранжировать значения одной или более переменных (вызывается из выпадающего меню *Правка* – *Переменные* или через кнопку **Переменные** панели инструментов).



Ранжирование переменных может производиться несколькими различными способами. Значения переменных таблицы исходных данных будут замещены рангами этих значений.

Если начальные значения переменной необходимо сохранить, надо сделать копию переменной (с помощью команды *Копировать переменные*, см. выше, или с помощью метода *Перетащить и отпустить*, удерживая нажатой клавишу CTRL) и произвести ранжирование скопированной переменной.

Переменные

Нажатие этой кнопки откроет диалоговое окно *Выберите переменные*, в котором можно выбрать одну или более переменных для ранжирования.

Наблюдения

Нажатие этой кнопки откроет диалоговое окно *Условия выбора наблюдений* (стр. 1223). Здесь можно указать подмножество наблюдений, которые нужно ранжировать. Заданное здесь условие будет применяться для каждой выбранной выше переменной.

Задать вес

Перед тем, как ранжировать переменные, можно приписать вес выделенным ранее переменным по отношению к другой переменной таблицы исходных данных. Нажатие этой кнопки откроет диалоговое окно *Задание веса* (стр. 1225), в

Copyright © StatSoft, 1995

котором можно выбрать переменную, которая содержит значения весов.

Присвоить ранг 1

Можно выбрать ранжирование значений в возрастающем (наименьшему значению присваивается ранг 1) или убывающем порядке (наибольшему значению переменной присваивается ранг 1).

Ранги для совпадающих значений

Для присвоения рангов совпадающим значениям доступны четыре варианта.

- *Средн.* Совпадающим значениям в качестве ранга присваивается среднее значение их рангов.
- *Следующий*. Каждое совпадающее значение ранжируется последовательно (совпадения игнорируются).
- *Низ.* Каждому из совпадающих значений присваивается низший из их рангов.
- *Выс.* Каждому из совпадающих значений присваивается высший из их рангов.

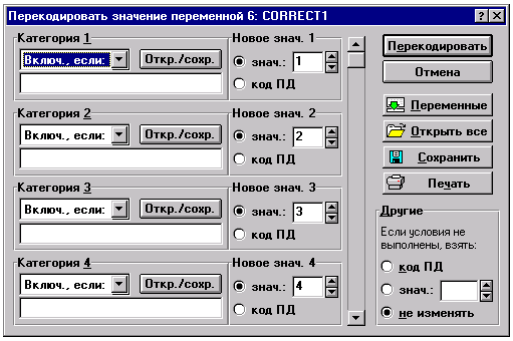
Тип рангов

С помощью данного переключателя производится выбор типа рангов.

- *Обычный*. Диапазон ранга от *1* до *n*.
- *Дробный*. Диапазон ранга от *0* до *1*.
- *Дробный в %* . Ранги считаются в процентах от дробных рангов значений переменных.

Перекодировать

Исходные значения выделенной переменной могут быть перекодированы или “переведены” в другие значения (в зависимости от значений любых переменных в текущем наборе данных) с помощью диалогового окна *Перекодировать значения переменной* (доступно из выпадающего меню *Правка – Переменные* или через кнопку Переменные панели инструментов).



Эта функция позволяет перекодировать значения переменной в 17 “категорий” значений (16 заданных категорий и “неопределенные” значения). Такое “быстрое перекодирование”, доступное в таблице исходных данных на любой стадии анализа, может быть использовано для присвоения переменным кодов (или “значений категории”, обычно маленьких целых чисел). Обычно эта функция используется для создания значений новой “группирующей” переменной с целью выделения наборов данных как групп для сравнения при дальнейшем анализе. Например, “присвоить значение *1* людям старше 17, значение *2* - людям младше 18” или “присвоить значение *1* мужчинам от 18 до 25 лет с уровнем холестерина до 200, значение *2* - мужчинам от 18 до 25 лет с уровнем холестерина более 200 и значение пропущенных данных - всем остальным лицам.” Если необходимо, чтобы после перекодировки значения конечной переменной были не фиксированными, а вычислялись по определенным формулам (например, как среднее значений других переменных), можно воспользоваться *формулами* таблицы исходных данных (стр. 1170) или заданными пользователем перекодирующими функциями на языке *STATISTICA BASIC* (см. *Электронное руководство*).

StatSoft

Copyright © StatSoft, 1995

Пример:

Категория 1 Включ. если: **`v1>=0`** *Новое значение:* **`1`**

Категория 2 Включ. если : **`v1<0`** *Новое значение:* **`-1`**

Этот набор (из двух) перекодирующих условий преобразует ("переведет") все наблюдения (наблюдение) выделенной переменной в *1* (если значение переменной номер 1 для этих наблюдений положительное или равно нулю) или в *-1* (если значение переменной номер 1 для этих наблюдений отрицательное).

Обратите внимание, что условия перекодирования могут быть определены таким образом, что новые значения текущей переменной не будут зависеть от старых значений этой переменной, а только от значений других переменных набора данных. Другими словами, выделенные переменные могут не включаться в текст условий перекодирования. В приведенном примере выделенной переменной может быть переменная номер 3, и ее новые ("перекодированные") значения будут зависеть только от значений переменной номер 1. Если была выделена переменная номер 1, то ее значения будут "переведены" в новые (положительные значения "станут" *1,* а отрицательные "станут" *-1*).

Категория

Для перекодирования переменной можно использовать до 16 категорий включительно.

Включ./Искл., если. Если выбран режим *Включ., если* и условие перекодирования истинно, начальные значения перекодированной переменной для наблюдений (наблюдений), удовлетворяющих условию, будут заменены на заданное *новое значение* (см. ниже). Операция *Искл., если* присвоит *новое значение* начальным значениям наблюдений, которые не удовлетворяют заданному условию перекодирования.

Условие перекодирования. Условие перекодирования задается в поле ввода *Категория* согласно стандартным *условиям выбора наблюдений* системы *STATISTICA* (см. стр. 1223).

- Операторы: =, <>, <, >, <=, >=, *NOT*, *AND*, *OR*
- Имена переменных: Указываются номера переменных (например, *v1*, *v2*, *v3*, ...) или их имена (например, *GENDER*, *Date*, *time*, . . .)
- Номера наблюдений*: v0*
- Примеры: *v1=0 OR v2>=0*

  (v1<1 OR v9='YES') AND v4<>0
- Примечание: Для ссылки на текстовые значения переменной используются апострофы, например, *'Yes'*

Кнопка Откр./Сохр. Эта кнопка используется для того, чтобы открыть (и использовать для перекодирования) существующий стандартный файл *условий выбора наблюдений* системы *STATISTICA* (эти условия используются для выбора или фильтрации наблюдений во всех процедурах анализа), а также сохранить текущее условие для каждой категории. Нажатие на кнопку *Откр./Сохр.* откроет диалоговое окно *Условия выбора наблюдений* (стр. 1223), в котором можно изменять, сохранять или открывать условия выбора наблюдений (это те же условия, которые используются в программе для обработки наборов наблюдений). Условия выбора наблюдений, сохраненные из этого диалогового окна, записываются в файл с расширением **.sel*. Позднее они могут как обычно использоваться для анализа данных, когда будет нажата кнопка *OK* в диалоговом окне *Условия выбора наблюдений* (стр. 1223).

Copyright © StatSoft, 1995

Новое значение Здесь указывается новое значение, на которое будет заменено начальное. В качестве нового значения может быть выбрана либо определенная величина (поле *Знач.*), либо значение пропущенных данных (кнопка *Код ПД*).

Переменные

Нажатие этой кнопки откроет диалоговое окно *Переменные*, в котором можно просмотреть переменные таблицы исходных данных.

Открыть все

Эта кнопка открывает диалоговое окно *Открытие файла с условиями перекодирования*, в котором можно выбрать имя файла, чтобы открыть его в диалоговом окне *Перекодировать значения переменных*. Файлы, открываемые из этого диалогового окна, являются текстовыми файлами с расширением **.txt*. Условия, задаваемые здесь, совместимы с условиями, вводимыми для проверки данных (см. *Проверка данных*, глава 7 *Управление данными*), категоризации частот и построения графиков (см. *Основные статистики и таблицы*, глава 9 этого тома, а также том II настоящего руководства соответственно), а также для проведения других операций в системе *STATISTICA,* которые требуют задания множественных условий выбора наблюдений. Эту операцию можно использовать для того, чтобы открыть и использовать здесь ранее сохраненные условия выбора.

Сохранить

Эта кнопка открывает диалоговое окно *Сохранение условий перекодирования*, в котором можно указать файл, где будут сохранены все условия перекодирования. Этот файл будет записан в текстовом формате (с расширением **.txt*). Эти условия выбора совместимы с условиями, используемыми для проверки данных (см. *Проверка данных*, глава 7, *Управление данными*), категоризации частот и построения графиков (см. *Основные статистики и таблицы*, глава 9 этого тома и том II настоящего руководства, соответственно) и других операций в системе *STATISTICA,* которые требуют задания множественных условий выбора наблюдений. Условия перекодирования, сохраненные с помощью этой операции, могут быть использованы позже в качестве условий выбора для других приложений системы *STATISTICA*.

Печать

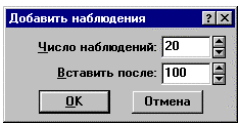
Нажатие этой кнопки приведет к выводу всех текущих условий перекодирования. Содержание этого диалогового окна может быть отправлено в файл, на принтер и/или в *Окно текста/вывода* в зависимости от установок в диалоговом окне *Параметры страницы/вывода* (стр. 1157).

Другие

Если наблюдение не удовлетворяет ни одному условию при перекодировке переменной, можно выбрать два варианта его замены: на код пропущенных данных (положение переключателя *Код ПД*) или на альтернативное значение (установив переключатель в положение *Знач.* и введя значение), или оставить начальные значения (положение переключателя *Не изменять*).

Добавить наблюдения

Для того, чтобы добавить наблюдения в таблицу исходных данных, надо в диалоговом окне *Добавить наблюдения* указать число добавляемых наблюдений, а также куда они добавляются (после какого наблюдения).



StatSoft
Copyright © StatSoft, 1995

Это диалоговое окно доступно из выпадающего меню *Правка – Наблюдения* таблицы исходных данных, контекстного меню, вызываемого правой кнопкой мыши, или меню кнопки Наблюдения панели инструментов.

Число наблюдений

Здесь указывается число наблюдений, добавляемых в текущую таблицу исходных данных. Число добавляемых наблюдений не ограничено.

Вставить после

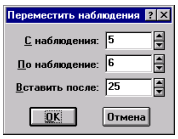
Здесь указывается номер наблюдения, после которого будут добавлены новые наблюдения.

Как вставить наблюдения перед первым наблюдением. Первый метод скопировать и вставить значения наблюдения или блока наблюдений перед наблюдением номер 1 состоит в том, чтобы использовать возможности метода *Перетащить и отпустить* системы *STATISTICA*. Для этого сначала необходимо выделить нужный блок, затем переместить курсор к одной из границ блока, где он изменится на стрелку. Теперь, чтобы переместить блок, надо нажать клавишу SHIFT, перетаскивая при этом блок. Далее надо поместить полосу вставки между именами переменных и первой строкой таблицы исходных данных и отпустить кнопку мыши. Обратите внимание, что этот метод копирует не выбранные наблюдения, а только значения выделенного блока (т.е., имена наблюдений скопированы не будут).

Чтобы скопировать наблюдение или блок наблюдений полностью до наблюдения номер 1, можно воспользоваться операцией копирования (доступной через кнопку Наблюдения панели инструментов) и ввести наблюдение *0* в поле ввода *Вставить после*. Будут перемещены не только значения выделенного блока, но и имена наблюдений.

Переместить наблюдения

Чтобы переместить одно или более наблюдений (строк) в таблице исходных данных, надо в диалоговом окне *Переместить наблюдения* указать диапазон (включительно) перемещаемых наблюдений и место, куда их надо переместить (после какого наблюдения вставить).



Это диалоговое окно доступно из выпадающего меню *Правка – Наблюдения* таблицы исходных данных, контекстного меню, вызываемого правой кнопкой мыши, или через кнопку Наблюдения панели инструментов.

С наблюдения

Здесь указывается первое наблюдение (включительно) из перемещаемого блока наблюдений. Обратите внимание, что при перемещении одного наблюдения в полях *С наблюдения* и *По наблюдение* нужно указать одно и то же значение.

По наблюдение

Здесь указывается последнее наблюдение (включительно) из перемещаемого блока наблюдений.

Вставить после

Здесь указывается место (наблюдение) в таблице исходных данных, куда будут перемещены наблюдения. Перемещенные наблюдения будут вставлены после указанного здесь наблюдения.

Как вставить наблюдения до наблюдения номер 1 . Первый способ переместить и вставить значения наблюдения или блока наблюдений до наблюдения номер 1 –

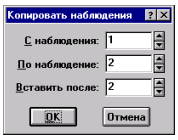
Copyright © StatSoft, 1995

воспользоваться возможностями метода *Перетащить и отпустить* системы *STATISTICA*. Для этого сначала необходимо выделить нужный блок. Затем переместить курсор мыши к одной из границ блока, где он изменится на стрелку. Теперь, чтобы переместить и вставить блок, нужно нажать клавишу SHIFT, перетаскивая при этом блок. Далее надо установить полосу вставки между именами переменных и первой строкой таблицы исходных данных и отпустить кнопку мыши. Обратите внимание, что этот метод перемещает не выбранные наблюдения (т.е. имена наблюдений не будут перемещены), а только значения выделенного блока.

Чтобы переместить все наблюдение или блок наблюдений до наблюдения номер один, можно воспользоваться операцией перемещения (доступной через кнопку **Наблюдения** панели инструментов) и ввести наблюдение *0* в поле ввода *Вставить после*. При этом будут перемещены не только значения, но и имена наблюдений.

Копировать наблюдения

В отличие от операции копирования выделенного диапазона или блока значений таблицы исходных данных через буфер обмена, эта команда копирует и содержание строк, и сами строки (независимо от того, все переменные или только часть переменных выделены для соответствующих наблюдений). Последующие наблюдения будут вставлены (см. ниже параграф *Вставить после*) в таблицу исходных данных, что приведет к увеличению размера файла. Копия указанных наблюдений будет помещена в нужное место, если в диалоговом окне *Копировать наблюдения* задан диапазон (включительно) копируемых наблюдений и место для их вставки в таблице исходных данных.



Это диалоговое окно доступно из выпадающего меню *Правка – Наблюдения* таблицы исходных данных, контекстного меню, вызываемого правой кнопкой мыши, или через кнопку *Наблюдения* панели инструментов.

С наблюдения

Здесь указывается первое наблюдение (включительно) из копируемого блока наблюдений. Обратите внимание, что при копировании одного наблюдения в полях *С наблюдения* и *По наблюдение* нужно указать одно и то же число.

По наблюдение

Здесь указывается последнее наблюдение (включительно) из перемещаемого блока наблюдений.

Вставить после

Здесь указывается место (наблюдение) в таблице, куда наблюдения будут копироваться. Копируемые наблюдения будут помещены после наблюдения, указанного здесь.

Как скопировать наблюдения до наблюдения номер 1. Первый способ скопировать и вставить значения наблюдения или блока наблюдений до наблюдения номер 1 – воспользоваться возможностями метода *Перетащить и отпустить* системы *STATISTICA*. Для этого сначала необходимо выделить нужный блок, затем переместить курсор к одной из границ блока, где он изменится на стрелку. Теперь, чтобы скопировать и вставить блок, надо нажать клавишу SHIFT, перетаскивая при этом блок. Далее надо установить полосу вставки между

StatSoft®

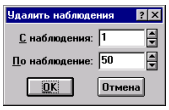
Copyright © StatSoft, 1995

именами переменных и первой строкой таблицы исходных данных и отпустить кнопку мыши. Обратите внимание, что этот метод копирует не выбранные наблюдения (т.е. имена наблюдений не будут скопированы), а только значения выделенного блока.

Чтобы скопировать наблюдение или блок наблюдений на место до наблюдения номер 1, можно воспользоваться операцией копирования (доступной через кнопку **Наблюдения** панели инструментов) и ввести наблюдение *0* в поле ввода *Вставить после*. При этом будут скопированы и вставлены не только значения, но и имена наблюдений.

Удалить наблюдения

Одно или более наблюдений (строк) можно удалить из текущей таблицы исходных данных, если в диалоговом окне *Удалить наблюдения* (доступном из выпадающего меню *Правка – Наблюдения*, контекстного меню, вызываемого правой кнопкой мыши, или через кнопку *Наблюдения* панели инструментов)указать диапазон (включительно) удаляемых наблюдений.



В отличие от операции удаления (очистки) содержания выделенного диапазона или блока значений таблицы исходных данных, эта операция удаляет и содержание строк, и сами строки (независимо от того, все переменные или только часть переменных выделены для соответствующих наблюдений); поэтому последующие наблюдения будут сдвинуты вверх и размер файла уменьшится.

С наблюдения

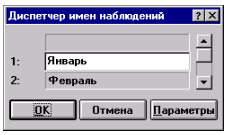
Здесь указывается первое наблюдение (включительно) из блока удаляемых наблюдений. Обратите внимание, что для удаления одного наблюдения надо в полях *С наблюдения* и *По наблюдение* указать одно и тоже значение.

По наблюдение

Здесь указывается последнее наблюдение (включительно) в блоке удаляемых наблюдений.

Изменение имен наблюдений

Диалоговое окно *Диспетчер имен наблюдений* отображает имена наблюдений в прокручиваемой таблице, где можно просмотреть и изменить их. Оно доступно из выпадающего меню *Правка – Наблюдения*, контекстного меню, вызываемого правой кнопкой мыши, через кнопку **Наблюдения** панели инструментов или после двойного щелчка на имени наблюдения, которое надо изменить.



Имена наблюдений используются только как метки (а не как текстовые значения); поэтому они могут содержать любые печатаемые символы (также как и метки значений, см. *Диспетчер текстовых значений*, стр. 1178).

Имена наблюдений

Имена наблюдений могут быть изменены в прокручиваемой таблице в диалоговом окне *Диспетчер имен наблюдений*. Изменения могут быть внесены в среднее из трех полей. Если диалоговое окно открывается в первый раз, то в среднем поле появится имя выделенного наблюдения таблицы исходных данных с

Copyright © StatSoft, 1995

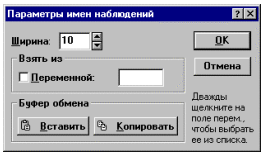
указанным слева от него номером. После того, как это имя будет изменено, можно прокрутить таблицу вверх или вниз (с помощью полосы прокрутки или кнопок со стрелками), чтобы отобразить другое имя, при этом номера наблюдений будут прокручиваться вместе с именами.

Параметры

Нажатие этой кнопки откроет диалоговое окно *Параметры имен наблюдений* (см. ниже), в котором можно изменить ширину столбца наблюдения, получить имена наблюдений из значений другой переменной или произвести с именами наблюдений операции, использующие буфер обмена.

Параметры имен наблюдений

Для изменения имен наблюдений в диалоговом окне *Диспетчер имен наблюдений* (см. выше), можно открыть диалоговое окно *Параметры имен наблюдений*, содержащее еще несколько изменяемых параметров.



Ширина

Здесь можно изменить ширину имени наблюдения (до 20 символов). Обратите внимание, что в отличие от *Ширины переменной* (“формат отображения”, см. стр. 1170), которое влияет только на отображение данных, а не на сами данные, текущие параметры ширины имен наблюдений влияют на размер резервируемого для них пространства в файле данных. Поэтому, уменьшение ширины “обрежет” (удалит) те части имен наблюдений, которые будут иметь большее количество символов. Если в качестве ширины будет указан *0*, то имена наблюдений будут удалены.

Взять из

Можно взять имена наблюдений из значений переменной из текущего набора данных. Для этого надо отметить пункт *Переменной* и указать номер переменной, из которой нужно взять имена наблюдений.

Буфер обмена

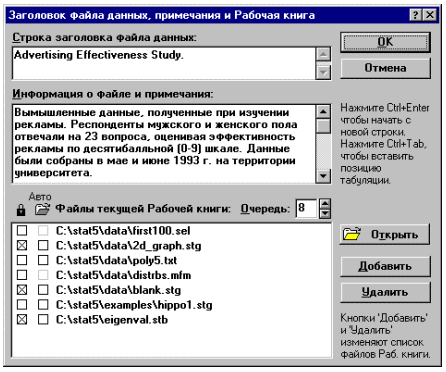
Для управления именами наблюдений могут быть выбраны следующие операции с буфером обмена.

Вставить. После нажатия на эту кнопку содержимое буфера обмена будет вставлено в столбец имен наблюдений (вниз от выделенного наблюдения).

Копировать. После нажатия на эту кнопку имена наблюдений (вниз от выделенного наблюдения) будут скопированы в буфер обмена.

Заголовок файла данных, примечания и Рабочая книга

В диалоговом окне *Заголовок файла данных, примечания и Рабочая книга* можно создавать примечания и комментарии о текущем файле данных.



StatSoft

Copyright © StatSoft, 1995

Это диалоговое окно позволяет ввести одну *Строку заголовка файла данных*, который размещается в верхней части таблицы исходных данных, подробную информацию о файле или примечания о содержании набора данных, их источнике и т.д., а также управлять списком файлов Рабочей книги, то есть списком всех дополнительных файлов (например, графиков, отчетов, программ), используемых с текущим набором данных.

Строка заголовка файла данных

Здесь можно ввести одну строку комментария к файлу (до 77 символов). *Строка заголовка* сохраняется как часть файла и печатается вместе с файлом данных, если выбран режим *Инф. и примечания* в диалоговом окне *Печать данных* (стр. 1154) или вместе с выводом анализа, если выбран *краткий*, *средний* или *полный* тип отчета в диалоговом окне *Параметры страницы/вывода* (стр. 1157). Заголовок также включается в название всех *пользовательских графиков*, построенных для текущего набора данных (см. том II).

Информация о файле и примечания

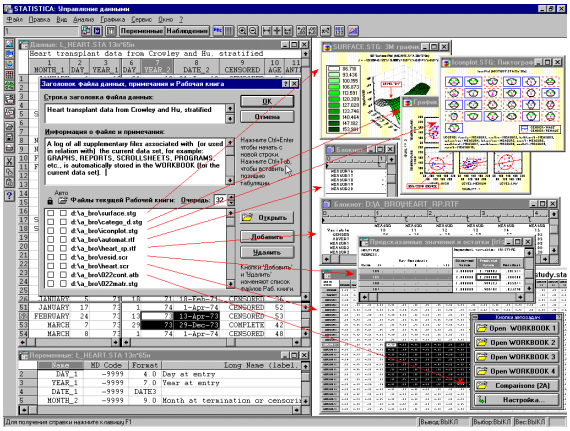
Здесь можно создать комментарий о файле данных таблицы исходных данных. Эти примечания могут быть до 512 символов длиной и содержать любые печатаемые символы. Для начала новой строки нужно нажать клавиши CTRL+ENTER, для ввода позиции табуляции – клавиши CTRL+TAB.

Информация о файле и примечания печатаются вместе с файлом данных таблицы исходных данных, если выбран пункт *Инф. и примечания* в диалоговом окне *Печать данных* (стр. 1154).

Рабочая книга

Файлы данных системы *STATISTICA* могут быть рассмотрены как “Рабочие книги” файлов, потому что они содержат списки всех дополнительных файлов, относящихся к набору данных.

Эта область диалогового окна *Заголовок файла данных, примечания и Рабочая книга* содержит параметры управления Рабочими книгами. Рабочие книги помогают организовывать наборы дополнительных файлов (например, таблицы результатов, графики, отчеты, содержащие текст и графики, настройки пользователя, условия перекодирования и т.д.), которые были созданы или использованы (например, просмотрены) во время анализа набора данных.



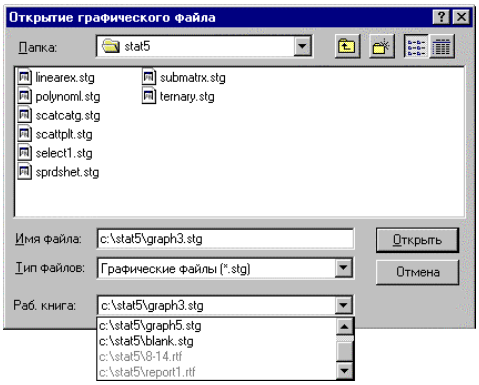
Каждый раз, когда файл (например, график) открывается или закрывается, его имя автоматически добавляется в список файлов Рабочей книги. Длина этого списка (по умолчанию 16) может быть изменена. Если список заполнен, то при добавлении новых файлов старые файлы удаляются из списка (если только они не заблокированы, см. ниже). Обновленный список этих файлов автоматически записывается с файлом данных (этот режим можно изменить в диалоговом окне *Параметры по умолчанию: общие настройки* системы *STATISTICA*; см. также ниже пункт *Если Рабочую книгу нужно отключить*).

Список файлов Рабочей книги также доступен из каждого диалогового окна

Copyright © StatSoft, 1995

StatSoft

Открытие/Сохранение системы *STATISTICA*, так что он может быть выбран (при открытии файла) или заменен (при сохранении файла).



Существует несколько способов открыть дополнительные файлы из списка Рабочей книги.

(1) Дважды щелкнуть на имени нужного файла в списке дополнительных файлов. Выбранный файл откроется позади диалогового окна *Заголовок файла данных, примечания и Рабочая книга*.

(2) Выбрать (выделить) один или несколько открываемых файлов в списке и нажать кнопку *Открыть* в диалоговом окне *Заголовок файла данных, примечания и Рабочая книга*. Это диалоговое окно закроется, и выбранные файлы будут поочередно открыты.

(3) Воспользоваться режимом Авто для выбора файлов, которые будут автоматически открываться вместе с текущим файлом данных (см. ниже).

Рабочая книга вызывается двойным щелчком на заголовке таблицы исходных данных, нажатием кнопки Рабочей книги на панели инструментов или при выборе пункта *Рабочая книга/Заголовки...* выпадающего меню *Правка*.

Ниже описаны параметры Рабочей книги.

Заблокировать. Дополнительные файлы могут быть заблокированы путем выбора кнопки рядом с именем файла под значком *Заблокировать* , чтобы они не удалялись из списка при его обновлении и изменении.

Авто. Здесь можно выбрать до 9 дополнительных файлов, которые будут автоматически открываться вместе с файлом данных. При этом станет доступна вся дополнительная информация (например, графики или отчеты с предыдущей стадии анализа). Если эта кнопка затенена, то файлы не могут быть открыты автоматически или открыты двойным щелчком на их именах в списке (например, это относится к файлам данных других программ, а также к файлам, не являющимся документами, например, содержащим параметры осей, спецификации поверхности/контура и т.п.).

Обратите внимание, что для перемещения между кнопками списка Рабочей книги можно использовать клавишу пробела, выделяя файлы, которые будут заблокированы и/или будут открываться автоматически (или отменяя выделение).

Файлы текущей Рабочей книги. Список всех дополнительных файлов, использовавшихся или созданных из текущего файла данных, будет отображен в этой части диалогового окна. Список управляется на основе метода “первый пришел, первый ушел”, то есть, когда список заполнен, самый старый файл будет удален из списка при добавлении нового (если только файл не заблокирован, см. выше). Типы отображенных здесь дополнительных файлов включают ранее сохраненные графики, таблицы результатов, отчеты, программы (например, на языке *STATISTICA BASIC* или *SCL*), файлы импорта/экспорта, условий выбора наблюдений и т.д.

Очередь. Здесь указывается число дополнительных файлов, содержащихся в списке. Максимальное число файлов, которое может содержаться в этом списке: 32. Список управляется на основе метода “первый пришел,

StatSoft
Copyright © StatSoft, 1995

первый ушел", то есть, когда список заполнен, самый старый файл будет удален из списка при добавлении нового (если только файл не заблокирован, см. выше).

Добавить. Эта кнопка открывает диалоговое окно выбора файлов, в котором можно выбрать один или несколько ранее сохраненных файлов, чтобы поместить их в список дополнительных. В этом диалоговом окне можно выбрать несколько файлов в виде непрерывного (просто проводя курсором) или произвольного (с помощью клавиши CTRL) списка.

Удалить файлы. Нажатие этой кнопки приведет к тому, что выделенные файлы будут удалены из списка *Файлы текущей Рабочей книги*.

Открыть. Нажатие этой кнопки приведет к тому, что выделенные файлы в списке *Файлы текущей Рабочей книги* будут последовательно открыты после того, как будет закрыто диалоговое окно *Заголовок файла данных, примечания и Рабочая книга*. Чтобы открыть файл, оставив это окно открытым, нужно дважды щелкнуть на имени файла в списке.

Если Рабочую книгу нужно отключить. Обычно Рабочая книга работает в автоматическом режиме, который прозрачен для пользователя. Иногда при выходе из программы (или при изменении файла) файлы данных сохраняются, даже если набор данных не был изменен (исправлен), для того чтобы информация о Рабочей книге в файле была обновлена. Однако, если только размер набора данных не очень большой, это не приводит к заметной задержке. Единственный случай, когда рекомендуется отключать Рабочую книгу (с помощью диалогового окна *Параметры по умолчанию: общие настройки* системы *STATISTICA*, см. стр. 1225), это когда набор данных очень большой и его сохранение приводит к заметной задержке. В этом случае отключение Рабочей книги предотвратит автоматическое сохранение файла данных и обновление списка Рабочей книги.

Вырезать

Операция *Вырезать* удаляет содержание выделенных ячеек, текст или графические объекты и копирует их в буфер обмена. Эта операция удаляет значения данных в таблицах результатов или таблицах исходных данных и заменяет их пропущенными данными, сохраняющимися до тех пор, пока не будут введены или вставлены новые значения. Выделенный текст просто удалятся из *Окна текста/вывода*. Операция *Вырезать* заменяет предыдущее содержание буфера обмена удаленными значениями, объектами или текстом.

Кнопка панели инструментов. См. кнопку панели инструментов в окнах таблицы исходных данных и таблицы результатов.

Комбинация клавиш. CTRL+X или SHIFT+DEL

Копировать

Эта операция копирует содержание выделенных ячеек или текст из таблицы результатов, таблицы исходных данных или *Окна текста/вывода* или выделенный графический объект в буфер обмена. В таблице результатов или таблице исходных данных эта операция скопирует не только содержание выделенного блока ячеек таблицы результатов или таблицы исходных данных, но и имена строк и столбцов, связанных с выделенными ячейками [для того, чтобы скопировать выделенный блок в таблице результатов или таблице исходных данных без имен строк и столбцов, можно воспользоваться операцией *Копировать содержание* (см. ниже)]. Команда *Копировать* заменяет предыдущее содержание буфера обмена содержанием

Copyright © StatSoft, 1995

выделенных ячеек, текстом или графическим объектом.

Кнопка панели инструментов. См. кнопку панели инструментов в окне таблицы исходных данных или таблицы результатов.

Комбинация клавиш. CTRL+C или CTRL+INS

Копировать содержание

Эта команда выпадающего меню *Правка* таблицы исходных данных предназначена для того, чтобы копировать в буфер обмена только данные, которые выделены в таблице результатов или таблице исходных данных. В отличие от команды *Копировать* (см. выше), эта операция не копирует имена строк или столбцов, связанных с выделенными данными. Команда *Копировать содержание* заменяет предыдущее содержание буфера обмена выделенным содержанием таблицы исходных данных или таблицы результатов. См. главу 2 *Панели инструментов*, пример сравнения операций *Копировать* и *Копировать содержание*.

Вставить

Эта операция вставляет текущее содержание буфера обмена в таблицу исходных данных, таблицу результатов, график или *Окно текста/вывода* системы *STATISTICA*.

Если эта операция используется в таблице исходных данных, таблице результатов или *Окнах текста/вывода* системы *STATISTICA*, текущая позиция курсора обозначает место (верхний левый угол), куда будет произведена вставка. Операция *Вставить* будет заполнять область до тех пор, пока не будет вставлено все содержимое буфера обмена. Эта операция недоступна (затенена), если буфер обмена пуст (или содержит несовместимый формат, например, при попытке вставить рисунок в таблицу результатов).

Кнопка панели инструментов. См. кнопку панели инструментов в окне таблицы исходных данных или таблицы результатов.

Комбинация клавиш. CTRL+V или SHIFT+INS

Вставить в графики

Если нужно вставить графический объект в график системы *STATISTICA,* можно воспользоваться операцией *Вставить*, в этом случае объект будет вставлен в верхний левый угол графика. После этого положение и размеры объекта могут быть изменены (см. *Изменение размера объекта*, том II настоящего руководства).

Буфер обмена и графические объекты (техническое замечание). Если в буфер обмена из системы *STATISTICA* копируется объект, графические представления (форматы) объекта будут созданы в буфере обмена в следующем порядке (иерархии): *График системы STATISTICA* (т.е. ее собственный формат), *Метафайл* (т.е. рисунок), а затем *Растровое изображение*; для текстовых объектов порядок следующий: *График системы STATISTICA*, *Метафайл*, *Только текст*, затем *Растровое изображение*. Если объект (или текст) копируется из другого приложения (например, программы Excel), порядок может отличаться от приведенного выше из-за того, что выбор форматов буфера обмена и приоритет этих форматов определяются приложением-источником. Поэтому при использовании операции *Вставить* используемый формат может не являться лучшим или выбранным для этого объекта. В этом случае лучше всего воспользоваться командой *Специальная вставка* (см. том II, главу *Общие свойства и параметры графиков*), которая позволит выбрать наиболее подходящий графический формат для объекта буфера обмена.

StatSoft®

Copyright © StatSoft, 1995

Например, для того, чтобы вставить большое изображение (например, растровое изображение в 256 цветах) из другого приложения, сначала нужно скопировать это изображение (из этого приложения) в буфер обмена (иерархия форматов буфера обмена будет определена приложением-источником; в этом примере предполагается следующая иерархия: растровое изображение, затем метафайл). Теперь, если выбрать операцию *Вставить* системы *STATISTICA*, изображение будет вставлено как растровое (поскольку этот формат имеет высший приоритет в буфере обмена); однако, формат метафайла может оказаться более удобным и более эффективным (по занимаемому пространству). Если же использовать команду *Специальная вставка*, то формат метафайла можно выбрать при вставке изображения в программу *STATISTICA*.

Вставка с помощью метода *OLE.* Информацию о возможностях метода Windows *OLE* (Связывание и внедрение объектов - *O*bject *L*inking and *E*mbedding), поддерживаемых системой *STATISTICA,* можно получить в томе II (глава *Общие свойства и параметры графиков*).

Очистить

Эта операция очищает текст или содержание выделенных ячеек. В таблицах результатов или таблицах исходных данных удаленные значения заменяются пропущенными значениями до тех пор, пока не будут введены или вставлены новые значения. В отличие от операции *Вырезать* (см. выше), операция *Очистить* не копирует текст или содержание ячеек в буфер обмена и поэтому не изменяет его содержание. Обратите внимание, что ячейки могут быть также очищены с помощью метода *Перетащить и отпустить* (см. стр. 1134).

Комбинация клавиш. DEL

Установить связь

Эта команда используется для установления связи *DDE* (динамического обмена данными) между системой *STATISTICA* (клиентом) и другим приложением (источником или сервером) с помощью стандартной операции Windows *Установить связь.* Она позволяет вставлять ячейки, скопированные из приложений Windows в блок ячеек таблицы исходных данных системы *STATISTICA* (вниз и вправо от выделенной ячейки), устанавливая при этом постоянную связь этих данных в таблице с их источником (сервером).

При использовании этой операции связь *DDE* создается и хранится в *длинном имени* текущей переменной (см. диалоговое окно таблицы исходных данных *Изменение (текущих) спецификаций*, стр. 1170). Связь *DDE* также можно изменить в диалоговом окне *Диспетчер связей* (см. стр. 1206). Для подробной информации о связях *DDE* см. раздел *Создание связей (динамический обмен данными)* в главе 7 *Управление данными*.

Чтобы установить связь *DDE* с помощью операции *Установить связь*, нужно проделать следующие шаги.

Шаг 1. Открыть исходный файл и скопировать в буфер обмена блок данных (или текст), который нужно связать с файлом-клиентом (файлом системы *STATISTICA*). Затем надо переключиться в нужный модуль системы *STATISTICA*, оставив файл-сервер открытым.

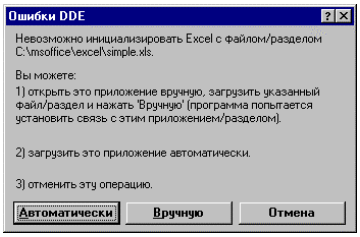
Шаг 2. В таблице исходных данных системы *STATISTICA* надо указать (выделить) ячейку, к которой будут “привязаны” данные, вставляемые в таблицу исходных данных (данные будут вставлены вниз и вправо от “связывающей” ячейки).

Copyright © StatSoft, 1995

Шаг 3. Чтобы установить связь, нужно выбрать пункт *Установить связь*. Система *STATISTICA* автоматически запишет путь *DDE* (см. раздел *Создание связей (динамический обмен данными)* в главе 7 *Управление данными*) и сохранит его в *длинном имени переменной* (см. стр. 1170) и в окне *Диспетчер связей* (стр. 1206).

Ошибки *DDE*

Это диалоговое окно открывается, если системе *STATISTICA* требуется повторно установить удаленную (через динамический обмен данными) связь – то есть если сервер закрыт и неактивен или не открыт раздел *DDE*.



Автоматически

Нажатие этой кнопки приведет к повторному установлению удаленной связи и автоматическому открытию файла/раздела приложения-источника, отображенных в верхней части этого диалогового окна (см. раздел *Создание связей (динамический обмен данными)* главы 7 *Управление данными* для получения подробной информации о *DDE*). Обратите внимание, что имя приложения должно быть прописано в пути DOS и файл/раздел должен загружаться с командной строки.

Вручную

Чтобы вручную загрузить указанный файл/раздел, можно воспользоваться этой кнопкой. Для этого нужно проделать следующие шаги.

Шаг 1. Выбрав файл/раздел для загрузки вручную (но до нажатия на кнопку *Вручную*), надо переключиться в окно *Диспетчер программ* Windows (или эквивалентную программу) и открыть приложение-источник (например Excel).

Шаг 2. Открыть нужный файл в приложении-источнике.

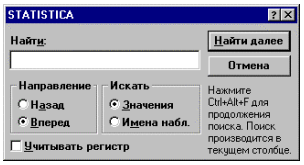
Шаг 3. Переключиться обратно в модуль системы *STATISTICA* и нажать кнопку *Вручную*, чтобы загрузить вручную файл/раздел.

Отмена

Нажатие этой кнопки приведет к временному отключению связи.

Найти

Диалоговое окно *Найти* (доступное из выпадающего меню *Правка* таблицы исходных данных) позволяет производить поиск определенных значений (в выделенной переменной) или имен наблюдений в таблице исходных данных, а также определенного текста или числа в *Окне текста/вывода*.



Найти

Здесь указывается значение переменной (текстовое или числовое) или имя наблюдения, которое необходимо найти в таблице исходных данных, или текст или число, которое необходимо найти в *Окне текста/вывода*.

StatSoft®
Copyright © StatSoft, 1995

Направление

Операция *Найти* будет производить поиск вверх (кнопка *Назад*) или вниз (кнопка *Вперед*) от выделенной ячейки, до тех пор, пока не будет достигнуто начало или конец файла.

Искать

Здесь можно выбрать, что нужно искать: *значения* текущей переменной или *имена наблюдений* (этот режим применяется только в таблице исходных данных).

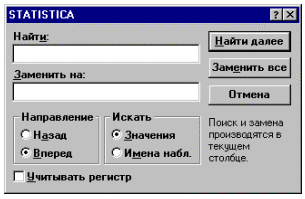
Найти далее

После выбора вышеуказанных вариантов надо нажать эту кнопку для начала поиска. Если обнаружено более одного совпадения, операция *Найти* сделает паузу, чтобы можно было закрыть диалоговое окно (кнопкой *Отмена*) или нажать кнопку *Найти далее* для поиска следующего совпадения.

Комбинация клавиш. Повторить поиск можно, нажав клавиши CTRL+F.

Заменить

Диалоговое окно *Заменить* доступно из выпадающего меню *Правка* таблицы исходных данных.



Это диалоговое окно позволяет производить поиск заданных значений данных в выделенной переменной (или заданное имя наблюдения) и заменять его другим значением, или производить поиск заданного текста или числа в *Окне текста/вывода*, и заменять его другим текстом или числом.

Найти

Здесь указывается значение переменной (текст или число) или имя наблюдения, которое нужно найти и заменить.

Заменить на

Здесь указывается значение (текст или число), которое будет использоваться для замены искомого значения, текста или имени наблюдения.

Направление

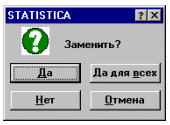
Операция *Заменить* будет производить поиск вверх (кнопка *Назад*) или вниз (кнопка *Вперед*) от выделенной ячейки или текста до тех пор, пока не будет достигнуто начало или конец файла.

Поиск

Здесь можно выбрать, что нужно искать: *значения* в текущей переменной или *имена наблюдений* (этот режим применяется только в таблице исходных данных).

Найти далее

После выбора вышеуказанных режимов надо нажать эту кнопку для начала поиска. Если системой *STATISTICA* был найден указанный текст, значение или имя наблюдения, будут предложены следующие варианты его замены:



Да. Эта кнопка предназначена для замены найденного текста или значения указанным ранее.

Copyright © StatSoft, 1995

StatSoft

Нет. Эта кнопка предназначена для того, чтобы пропустить это значение (оно не будет заменено).

Да для всех. Эта кнопка предназначена для замены всего совпадающего текста или значений без вывода этого запроса.

Отмена. Эта кнопка предназначена для прекращения поиска и закрытия диалогового окна.

Заменить все

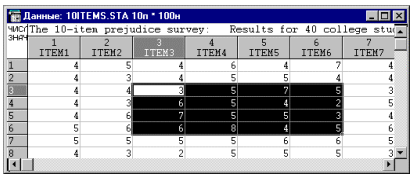
Эту кнопку можно нажать вместо кнопки *Найти далее* для замены всех совпадающих значений или текста. При этом не будет выводиться подтверждение каждой замены или ее отмены.

Комбинация клавиш. Можно повторить поиск, нажав клавиши CTRL+H.

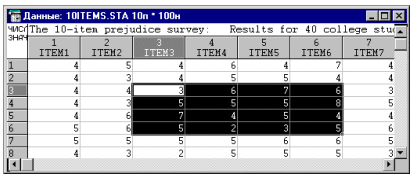
Транспонировать блок

Можно транспонировать квадратный блок данных в таблице исходных данных, выделив блок и выбрав пункт *Транспонировать – Блок* выпадающего меню *Правка*. Эта операция транспонирует только значения выделенного блока, а не файла данных (см. ниже *Транспонирование файла данных*).

До транспонирования выделенного блока:



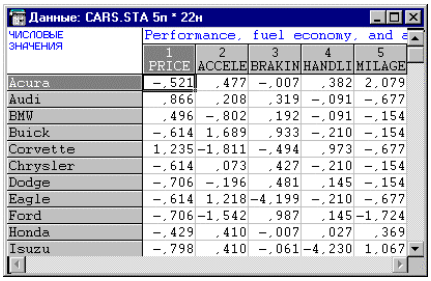
После транспонирования выделенного блока:



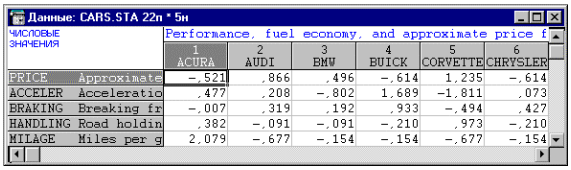
Транспонирование файла данных

Выбрав пункт *Транспонировать – Файл данных* выпадающего меню *Правка* , можно превратить наблюдения в переменные, а переменные в наблюдения, транспонировав файл данных. (Чтобы транспонировать квадратный блок данных, нужно выбрать команду *Транспонировать - Блок*, см. выше)

До транспонирования файла данных:



После транспонирования файла данных:



При транспонировании файла данных эквиваленты текстовых значений будут утеряны, поскольку текстовые/числовые присвоения в файлах данных системы *STATISTICA* являются спецификациями переменных (т.е. столбцов) а не наблюдений (т.е. строк).

Имена наблюдений (если они есть) из оригинального файла станут *длинными именами переменных*, созданными из бывших (8-

StatSoft®

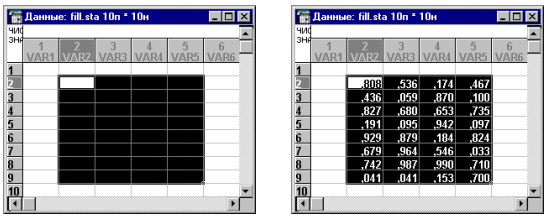
Copyright © StatSoft, 1995

символьных) имен переменных и до 12 символов бывших длинных имен переменных (всего 20 символов).

Заполнить/стандартизировать блок - Заполнить случайными значениями

Эта операция заполнит выделенный блок равномерно распределенными случайными значениями (в диапазоне от 0 до 1).

До: После:

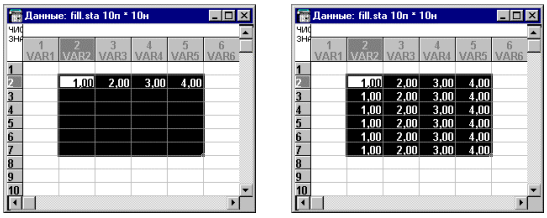


О копировании строк вниз см. *Заполнить/копировать вниз* (ниже); о копировании столбцов в сторону см. *Заполнить вправо* (ниже). Обратите внимание, что для копирования блока значений или экстраполяции значений в сторону можно воспользоваться возможностями метода *Перетащить и отпустить* (см. стр. 1132).

Заполнить/стандартизировать блок - Заполнить/копировать вниз

Эта операция (доступная из выпадающего меню *Правка* таблицы исходных данных или из контекстного меню, вызываемого правой кнопкой мыши) скопирует содержание первой строки выделенного блока ячеек таблицы исходных данных до конца выделенного блока (т.е. в оставшиеся ячейки блока).

До: После:

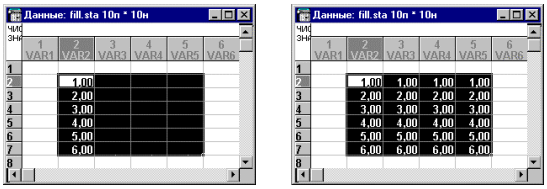


Если нужно скопировать значение только в последующие наблюдения одной переменной, надо выделить блок из одного столбца; его первое (“верхнее”) значение будет скопировано до конца выделенного блока. (О копировании значений в сторону см. ниже пункт *Заполнить вправо*.) Для копирования блока значений или экстраполяции значений блока можно воспользоваться возможностями метода *Перетащить и отпустить* (см .стр. 1132). Для заполнения ячеек блока произвольными значениями можно воспользоваться командой *Заполнить случайными значениями* (см. выше) (при этом значения будут распределены равномерно).

Заполнить/стандартизировать блок - Заполнить/копировать вправо

Эта операция (доступная из выпадающего меню *Правка* таблицы исходных данных) скопирует содержание первого столбца выделенного блока ячеек таблицы исходных данных, до конца выделенного блока (направо от первого столбца), в оставшиеся ячейки блока.

До: После:



Copyright © StatSoft, 1995

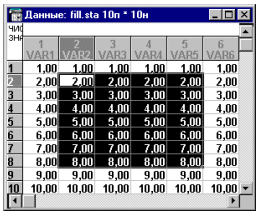
StatSoft®

Если нужно скопировать значение в последующие ячейки (переменные) только одного наблюдения, надо выделить блок из одной строки; его первое (самое левое) наблюдение будет скопирован во все последующие ячейки наблюдения. (О копировании строк вниз см. выше пункт *Заполнить вниз*.) Для копирования блока данных или экстраполяции значений блока можно воспользоваться возможностями метода *Перетащить и отпустить* (см. стр. 1132). Для заполнения ячеек блока произвольными значениями можно воспользоваться командой *Заполнить случайными значениями* (см. выше) (при этом значения будут распределены равномерно).

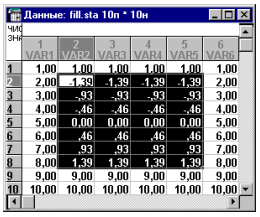
Заполнить/стандартизировать блок – Стандартизировать столбцы

Эта операция (доступная из выпадающего меню *Правка* таблицы исходных данных) стандартизирует значения в каждом столбце выделенного блока.

До:



После:



Стандартизованные значения вычисляются следующим образом:

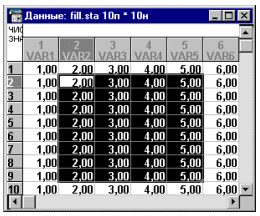
Станд. значение = (значение ячейки – среднее выделенного столбца) / станд. отклонение

Стандартизировать значения выделенных строк можно с помощью команды *Стандартизировать строки* (см. ниже). Выделенные переменные также могут быть стандартизованы с помощью команды *Стандартизировать переменные* модуля *Управление данными* (эта процедура действует независимо от выделенного блока, но принимает в расчет текущие условия выбора наблюдений и вес).

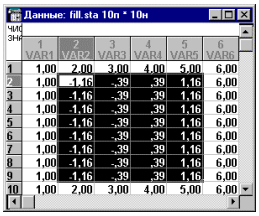
Заполнить/стандартизировать блок – Стандартизировать строки

Эта операция (доступная из выпадающего меню *Правка* таблицы исходных данных) стандартизирует значения в каждой строке выделенного блока.

До:



После:



Стандартизованные значения вычисляются следующим образом:

Станд. значение = (значение ячейки – среднее выделенной строки) / станд. отклонение

Стандартизировать значения выделенных столбцов можно с помощью команды *Стандартизировать столбцы* (см. выше). Выделенные переменные также можно стандартизировать с помощью команды *Стандартизировать переменные* модуля *Управление данными* (эта процедура действует независимо от выделенного блока, но принимает в расчет текущие условия выбора наблюдений и вес).

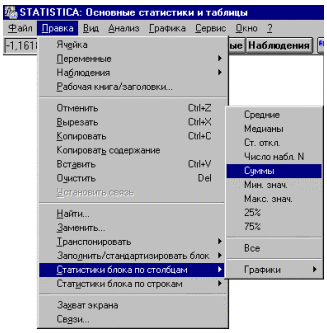
Статистики блока по столбцам

Выбрав эту команду в выпадающем меню *Правка* таблицы исходных данных или таблицы результатов (или в контекстном меню), можно

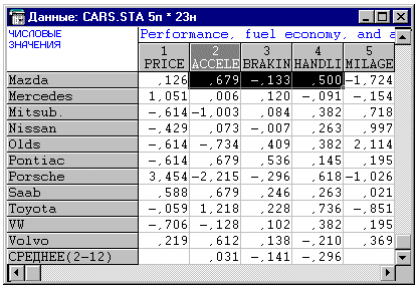
StatSoft

Copyright © StatSoft, 1995

вычислить *средние*, *медианы*, *стандартные отклонения*, *число наблюдений N*, *суммы значений*, *мин*имальное значение, *макс*имальное значение, *25* процентную точку или *75* процентную точку переменных (столбцов) выделенного блока или *все* перечисленные выше статистики.



Вычисленные значения будут добавлены в новые строки после последней строки таблицы исходных данных или таблицы результатов (система *STATISTICA* создаст в выделенном блоке имена наблюдений с именами статистик и номерами наблюдений).

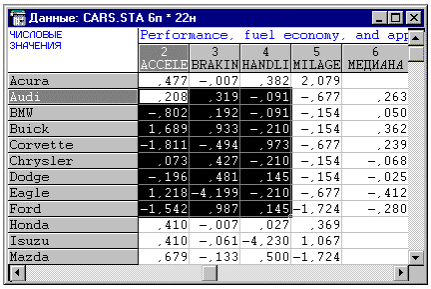


Для данных в выделенном блоке также можно построить несколько типов статистических графиков (см. ниже *Блоковые статистические графики*); данные из ячеек можно отобразить на графике с помощью *пользовательских графиков*.

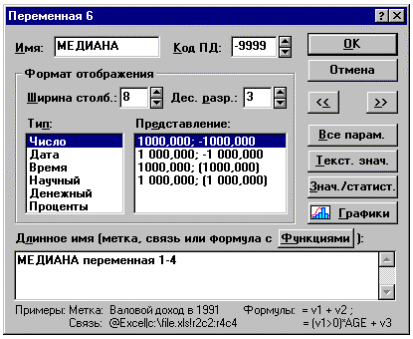
Примечание: Если файл данных очень большой (например, более 10 мегабайт), в диалоговом окне *Параметры по умолчанию: общие настройки* можно включить режим *Запрос при вставке строк или столбцов*, тогда система *STATISTICA* будет выводить запрос на подтверждение, перед тем как увеличить размер файла данных.

Статистики блока по строкам

Выберите этот пункт выпадающего меню *Правка* таблицы исходных данных или таблицы результатов (или из контекстного меню) для вычисления *средних*, *медиан*, *стандартных отклонений*, *числа значений*, *сумм значений*, *мин*имального значения, *макс*имального значения, *25* процентной точки или *75* процентной точки наблюдений (строк) выделенного блока или *всех* перечисленных выше статистик. Вычисленные значения будут добавлены в новые столбцы после последнего столбца таблицы исходных данных или таблицы результатов.



Имена переменных будут содержать имена выбранных статистик.



Copyright © StatSoft, 1995

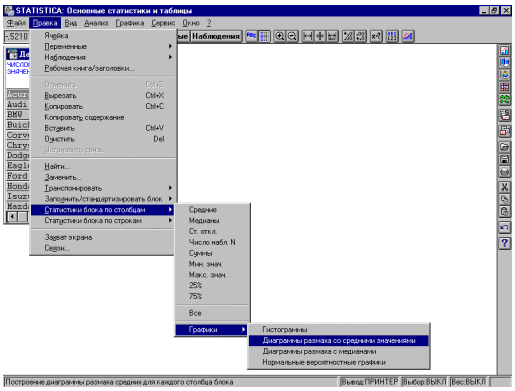
StatSoft®

Для данных в выделенном блоке также можно построить несколько типов статистических графиков (см. *Блоковые статистические графики* ниже); данные из ячеек можно отобразить на графике с помощью *пользовательских графиков.*

Примечание: Если файл данных очень большой (например, более 10 мегабайт), в диалоговом окне *Параметры по умолчанию: общие настройки* можно включить режим *Запрос при вставке строк или столбцов*, тогда система *STATISTICA* будет выводить запрос на подтверждение, перед тем как увеличить размер файла данных.

Блоковые статистические графики

Блоковые статистические графики, как и *пользовательские графики*, отображают преобразованные значения выделенной части (блока) таблицы результатов или таблицы исходных данных в графики. Однако в отличие от *пользовательских графиков*, которые позволяют визуализировать любую комбинацию значений из ячеек таблицы результатов или таблицы исходных данных, *блоковые статистические графики* представляют собой *итоговые статистические графики*; они позволяют сравнивать распределения значений в столбцах (или строках) выделенного блока, путем построения представления всех столбцов (или строк) на одном графике. Также по этим графикам можно эффективно исследовать (и подводить итоги) численные результаты анализа, отображенные в таблицах результатов (например, гистограммы расчетов *Монте-Карло* в модуле *Структурное моделирование* или диаграммы общего среднего из таблиц классификации многих переменных в модуле *Дисперсионный анализ*.



Меню *Блоковые статистические графики* также можно вызвать из контекстного меню, выпадающего меню *Графика* или диалогового окна *Галерея графиков*.

Если меню *Блоковые статистические графики* вызвано из окна *Галерея графиков* и активной является таблица результатов, то будет обрабатываться выделенный блок таблицы результатов. Во всех остальных случаях (например, когда активным окном является график или *Окно текста/вывода*) будет обрабатываться выделенный блок таблицы исходных данных, которая будет помещена впереди всех окон.

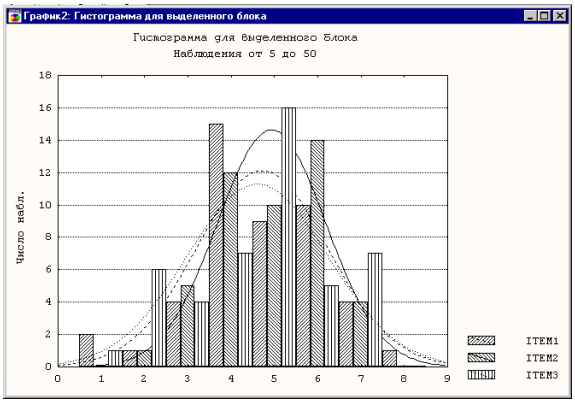
Если *блоковые статистические графики* создаются из таблицы исходных данных, то они будут обновляться автоматически при обновлении данных в таблице (см. окно *Параметры по умолчанию: общие настройки* системы *STATISTICA*).

Гистограммы

При выборе операции *Статистики блока по столбцам* для каждого столбца блока будет создана двумерная гистограмма с подогнанной нормальной кривой, которые будут построены на одном графике (множественная гистограмма) для сравнения распределения частот значений в столбцах блока.

StatSoft®

Copyright © StatSoft, 1995



Для создания гистограммы одной переменной можно выбрать блок, содержащий только один столбец.

Таким же образом, если выбрать пункт *Статистики блока по строкам*, то для каждого блока будет построена двумерная гистограмма; все гистограммы будут построены на одном графике (смешанная гистограмма) для сравнения распределения частот значений в строках блока. Для создания гистограммы одной переменной можно выбрать блок, содержащий только одну строку.

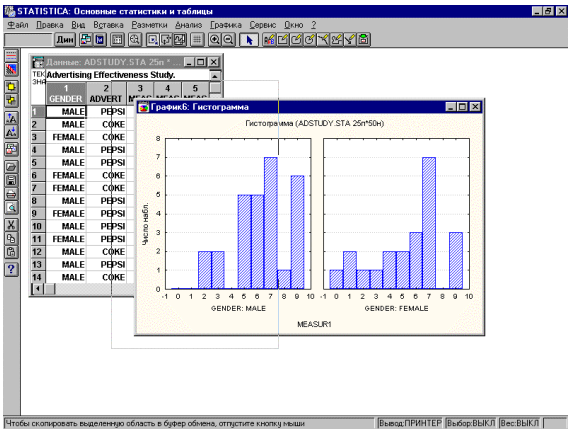
Примечания. Если меню *Блоковые статистические графики* вызвано из окна *Галерея графиков* и активной является таблица результатов, то будет обрабатываться выделенный блок таблицы результатов. Во всех остальных случаях (например, когда активным окном является график или *Окно текста/вывода*), то будет обрабатываться выделенный блок таблицы исходных данных, которая будет помещена впереди всех окон.

Если *блоковые статистические графики* создаются из таблицы исходных данных, то они будут обновляться автоматически при обновлении данных в таблице исходных данных (см. окно *Параметры по умолчанию: общие настройки* системы *STATISTICA*).

Захват экрана

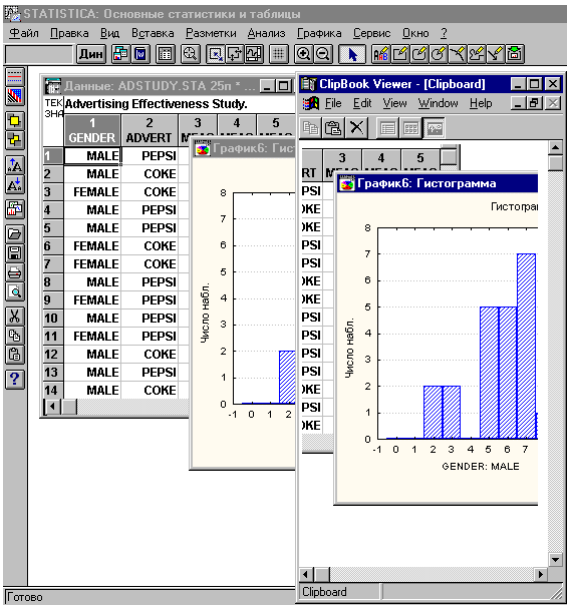
Эта команда (доступная из выпадающего меню *Правка* таблицы исходных данных, а также в любом месте программы по нажатию клавиш ALT+F3) вызывает программу *Захват экрана*, которая позволяет выделить любую прямоугольную область экрана (или весь экран) и скопировать ее в буфер обмена как растровое изображение.

Выбрав эту команду (или нажав клавиши ALT+F3), надо установить курсор в один из углов копируемой области. Далее, удерживая нажатой левую кнопку мыши, надо провести курсором над копируемой областью (обратите внимание, что проводить курсором можно и над окнами других приложений).



Как только кнопка мыши будет отпущена, выделенная область будет скопирована в буфер обмена (ниже она показана в окне просмотра буфера обмена).

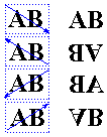
Copyright © StatSoft, 1995



Копируемая область может содержать графики, таблицы результатов, таблицы исходных данных или даже окна других приложений, если они отображаются на экране. Копируемое растровое изображение потом можно вставить в графики системы *STATISTICA* или любое другое приложение, поддерживающее растровый формат.

Обратите внимание, что при выделении захватываемой области ее можно поворачивать на 180° и/или зеркально отображать, выбирая одно из четырех направлений перемещения мыши.

Следующие четыре примера демонстрируют варианты поворота объекта, которые могут быть получены перемещением мыши в разных направлениях над захватываемой областью:

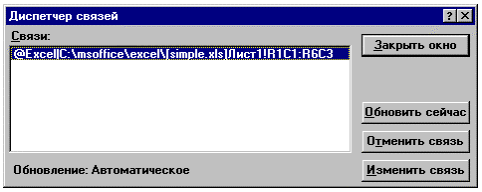


Эта функция может использоваться при анализе графических данных (например, для сравнения частей контурных графиков), а также при создании специальных эффектов для презентаций.

Комбинация клавиш. ALT+F3

Связи

Эта команда открывает диалоговое окно *Диспетчер связей*, в котором можно изменять связи *DDE* (динамического обмена данными) между приложением-источником (или сервером) и системой *STATISTICA* (приложением-клиентом). Эти связи хранятся (и могут быть изменены) в параметре *Длинное имя* диалогового окна таблицы исходных данных *Переменная…*(*Изменение (текущих) спецификаций*, стр. 1170). Установить связь также можно с помощью команды *Установить связь* (стр. 1197). Для подробной информации об установлении связей *DDE* см. раздел *Создание связей (динамический обмен данными)* главы 7 *Управление данными*.



Связи

В этом поле отображаются все связи *DDE* системы *STATISTICA*. Связь можно изменить, выделив ее в этом окне и нажав кнопку *Изменить связь* (см. ниже).

Обновление

Отображает состояние связи *DDE* , выделенной в поле *Связи*. Связи *DDE* могут быть “горячими” (*автоматическое* обновление, перед оператором

StatSoft

Copyright © StatSoft, 1995

связи стоит символ “@”) или “холодными” (*ручное* обновление, перед оператором связи стоит символ “^”). *Автоматическое* обновление означает, что привязанные данные в таблице исходных данных системы *STATISTICA* (файле-клиенте) будут автоматически обновляться при внесении изменений в файл в приложении-сервере. *Ручное* обновление означает, что связь *DDE* временно отключена, и файл-клиент должен будет обновляться вручную с помощью кнопки *Обновить сейчас*, или изменением состояния снова на “горячее” (с помощью кнопки *Изменить связь*).

Обновить сейчас

Эта кнопка предназначена для ручного обновления связи *DDE* (которая выделена в поле *Связи*) в файле-клиенте (таблице исходных данных системы *STATISTICA*).

Отменить связь

Эта кнопка предназначена для временного отключения выделенной связи *DDE*.
Примечание: При проведении анализа с файлом данных, связанным (например, по сети) с файлом, который может быть изменен, нужно использовать этот режим для временного отключения связи, чтобы данные не могли измениться в процессе анализа. Восстановить связь можно, нажав кнопку *Обновить сейчас*.

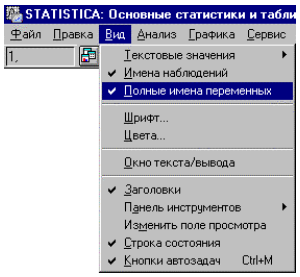
Изменить связь

Нажатие этой кнопки приведет к открытию диалогового окна *Переменная...*(*Изменить (текущие) спецификации*, стр. 1170), в котором можно изменить связь *DDE* с помощью команды *Длинное имя - Связь* (подробную информацию о частях пути *DDE* см. раздел *Создание связей (динамический обмен данными)* главы 7 *Управление данными*).

Copyright © StatSoft, 1995

ВЫПАДАЮЩЕЕ МЕНЮ *ВИД* ТАБЛИЦЫ ИСХОДНЫХ ДАННЫХ

В выпадающем меню *Вид* доступны следующие диалоговые окна и команды,



которые подробно описаны ниже.

Текстовые значения

Этот пункт (выпадающего меню *Вид* таблицы исходных данных) позволяет переключаться между отображением текста, числовых эквивалентов значений и меток значений (если они есть) всех переменных с текстовыми значениями в файле данных таблицы исходных данных. Переключаться между различными отображениями можно, нажимая кнопку панели инструментов (для отображения меток значений нужно при нажатии этой кнопки удерживать нажатой клавишу CTRL).

Имена наблюдений

Этот пункт (выпадающего меню *Вид* таблицы исходных данных) позволяет переключаться между отображением имен наблюдений и номеров наблюдений в файле данных таблицы исходных данных. Если этот пункт отмечен галочкой, по умолчанию в таблице исходных данных будут отображаться имена наблюдений; в противном случае (если нет галочки), будут отображаться только номера наблюдений.

Кнопка панели инструментов. См. кнопку панели инструментов таблицы исходных данных.

Полные имена переменных

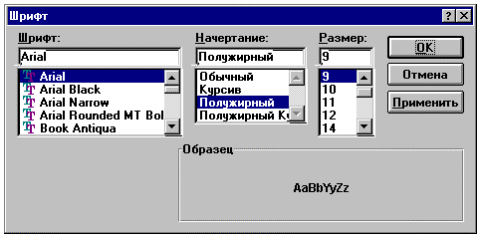
Если формат ширины столбца значений переменных (указанный в диалоговом окне таблицы исходных данных *Переменная…(Изменить (текущие) спецификации)*, стр. 1170) меньше 8 символов (максимальная длина имени переменной), то отображаемое имя переменной будет “обрезано” до формата ширины его значений. Чтобы посмотреть (отобразить) имя переменной полностью, надо выбрать пункт *Полные имена переменных* выпадающего меню *Вид* таблицы исходных данных. Если этот пункт отмечен галочкой, и ширина какого-нибудь столбца меньше, чем соответствующее имя переменной, то ширина этого столбца будет автоматически увеличена, чтобы отображать полное имя переменной.

Диалоговое окно *Шрифт*

Выбор пункта *Шрифт* выпадающего меню *Вид* приведет к открытию диалогового окна *Шрифт*. Здесь можно выбрать тип, стиль и размер шрифта таблицы исходных данных, таблицы результатов или других приложений системы *STATISTICA*. Изменения можно посмотреть в поле *Образец* этого диалогового окна.

StatSoft®

Copyright © StatSoft, 1995



Таблицы исходных данных и таблицы результатов сохраняются с установленными шрифтами и цветами.

Комбинация клавиш. CTRL+F9

Шрифт

Здесь указывается имя используемого шрифта. Для выбора щелкните на типе шрифта.

Начертание

Доступные стили шрифта зависят от выбранного типа шрифта. Сначала нужно выбрать тип шрифта, затем щелкнуть на нужном стиле шрифта.

Размер

Доступные размеры шрифта зависят от типа выбранного шрифта. Сначала нужно выбрать тип шрифта, затем щелкнуть на нужном размере шрифта.

По умолчанию

Эта команда предназначена для сохранения установок шрифта как установок по умолчанию для таблиц исходных данных и таблиц результатов (для таблиц исходных данных и таблиц результатов могут быть заданы разные параметры по умолчанию).

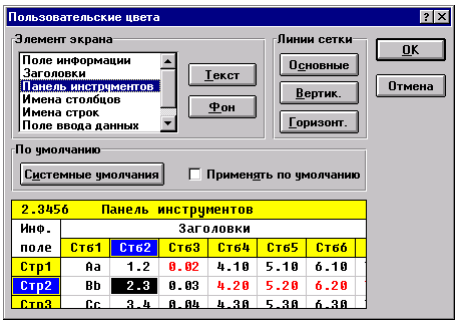
Применить

Эта кнопка может быть использована при вводе размера шрифта, который, возможно, не поддерживается данным типом шрифта. Шрифт этого размера можно посмотреть в окне, перед тем как использовать его в системе *STATISTICA*.

Пользовательские цвета для таблицы результатов и таблицы исходных данных

Эта команда позволяет задавать цвета таблиц результатов и таблиц исходных данных в диалоговом окне *Пользовательские цвета*. Перед тем как сделать окончательный выбор, изменения цветов можно посмотреть в этом диалоговом окне на образце таблицы результатов или таблицы исходных данных, расположенном под списком.

Каждая из кнопок этого диалогового окна открывает стандартное диалоговое окно *Цвет* (стр. `1210), в котором можно выбрать цвета из стандартной цветовой палитры или задать свой цвет (с помощью кнопки *Определить цвет*).



Таблицы исходных данных и таблицы результатов сохраняются вместе с их цветами (например, пользовательскими, если они были заданы).

Комбинация клавиш: SHIFT+F9

Пример. В этом примере будут заданы цвета заголовков таблицы результатов.

Шаг 1. Выбрать пункт *Цвета* из выпадающего меню *Вид*.

Copyright © StatSoft, 1995

StatSoft

Шаг 2. Изменить цвет заголовков таблицы результатов, выбрав элемент экрана *Заголовки*. Нажать на кнопку *Текст* и выбрать желаемый цвет текста заголовков (скажем, синий). Теперь надо выбрать контрастирующий цвет для фона заголовка, нажав на кнопку *Фон* для выбора цвета в диалоговом окне *Цвет*. *Шаг 2* можно повторить для каждого элемента экрана, цвет которого нужно изменить.

Шаг 3. Нажать кнопку *OK*, чтобы закрыть диалоговое окно *Пользовательские цвета* и увидеть измененные цвета текущей таблицы результатов.

Элемент экрана

Здесь можно выбрать элемент таблицы результатов/таблицы исходных данных, который нужно изменить (см. ниже список этих элементов). Каждый элемент таблицы результатов/таблицы исходных данных изменяется отдельно и у каждого элемента можно изменить только цвет текста (кнопка *Текст*), только фон (кнопка *Фон*), или и то, и другое. Для *текста* доступны только сплошные цвета; если выбран полутоновый цвет, то система *STATISTICA* использует ближайший к выбранному сплошной цвет.

- Поле информации (т.е. верхний левый угол таблицы)
- Заголовки
- Панель инструментов
- Имена столбцов
- Имена строк
- Поле ввода данных
- Подсветка

Линии сетки

Здесь задаются цвета линий таблиц результатов или таблиц исходных данных.

Основные. Здесь можно изменить цвет линий, разделяющих имена строк и имена столбцов в таблице результатов или таблице исходных данных.

Вертикальные. Здесь можно изменить цвет вертикальных линий, разделяющих значения в таблице результатов или таблице исходных данных.

Горизонтальные. Здесь можно изменить цвет горизонтальных линий, разделяющих значения в таблице результатов или таблице исходных данных.

Применять по умолчанию

Выбранные цвета можно установить по умолчанию, если выбрать режим *Применять по умолчанию*. Для того, чтобы восстановить цвета по умолчанию системы *STATISTICA*, можно нажать кнопку *Системные умолчания*.

Диалоговое окно *Цвет*

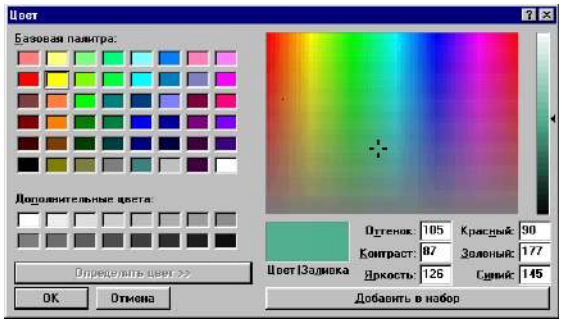
Следующие два диалоговых окна доступны из диалогового окна *Пользовательские цвета* при каждом выборе изменяемого элемента таблицы исходных данных/таблицы результатов.

Базовая палитра. В этой палитре нужно щелкнуть на цвете, чтобы выбрать его для элемента таблицы результатов/таблицы исходных данных.

StatSoft®

Copyright © StatSoft, 1995

Дополнительные цвета . Нажмите кнопку *Определить цвет*, чтобы открыть *Окно уточнения цвета,* в котором можно создать свой цвет и добавить его в палитру *Дополнительные цвета*.



Создать цвет вручную можно, указав для него *оттенок*, *контрастность* и *яркость*, указав значения компонент *красного*, *зеленого и синего* цветов, или перемещая курсор на нужный цвет в *Окне уточнения цвета* и настраивая *полосу яркости*.

Для того, чтобы добавить новый цвет в выбранную ячейку палитры *Дополнительные цвета*, надо нажать кнопку *Добавить в набор*.

Окно текста/вывода

Если выбрать этот пункт выпадающего меню *Вид*, откроется *Окно текста/вывода* (подробную информацию об этом окне см. в главе 5). Весь вывод будет направлен в это окно (вне зависимости от того, производится ли также печать на принтер или в файл).

Если в диалоговом окне *Параметры страницы/вывода* (стр. 1157), установлен какой-либо из режимов *автоотчета* (например, *Авт. печать всех таблиц результатов*), то соответствующий вывод (например, графики или содержание каждой таблицы результатов, отображаемой на экране) будет также направлен в *Окно текста/вывода* (т.е., если эта опция включена, кнопку печати использовать не нужно). Открыть *Окно текста/вывода* можно также, выбрав пункт *Окно* в диалоговом окне *Параметры страницы/вывода* (стр. 1157).

Обратите внимание, что в любом случае открытие этого окна приведет к тому, что вывод результатов анализа будет направлен в это окно, а также на принтер или в файл (в зависимости от установок в диалоговом окне *Параметры страницы/вывода*). В этом окне можно изменять текст и вставлять дополнительную информацию [форматированный текст (в формате *RTF*) и графические объекты из системы *STATISTICA* или других приложений Windows]с помощью стандартных операций *Копировать* и *Вставить* (стр. 1195).

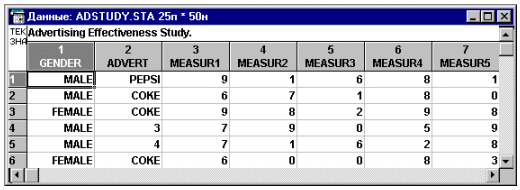
При проведении нескольких анализов “новые” результаты будут добавляться к предыдущим, составляя непрерывный отчет о получаемых результатах. Если полный отчет об анализах необходимо хранить постоянно, нужно выбрать *Файл вывода*. Если необходимо хранить текст и/или графики из *Окна текста/вывода* (включая примечания и вставленный в окно текст), то можно воспользоваться пунктом *Текст. файл*. *Окно текста/вывода* может быть использовано как “блокнот“ для хранения записей или примечаний. У открытого и активного *Окна текста/вывода* есть своя панель инструментов и свои выпадающие меню (см. главу 5).

Copyright © StatSoft, 1995

Техническое замечание. Если таблица исходных данных или таблица результатов выводится в *Окно текста/вывода*, по умолчанию будет использован шрифт, заданный в диалоговом окне *Параметры страницы/вывода* для таблиц исходных данных и таблиц результатов. Поэтому, для того чтобы таблицы исходных данных или таблицы результатов выводились в *Окно текста/вывода* другим шрифтом, его нужно указать в диалоговом окне *Параметры страницы/вывода*. Если в *Окно текста/вывода* вставляется текст из другого приложения, то для вставки его с исходным шрифтом в диалоговом окне *Специальная вставка* надо выбрать формат *RTF*.

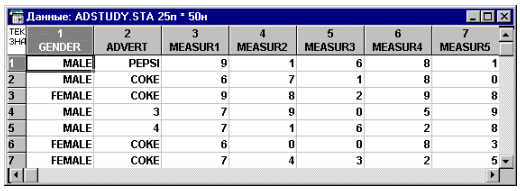
Заголовки

С помощью этого пункта можно выбрать, будет ли таблица исходных данных отображаться с заголовком (если установлена галочка, то таблица исходных данных будет отображаться с заголовком из одной строки)



	1 GENDER	2 ADVERT	3 MEASUR1	4 MEASUR2	5 MEASUR3	6 MEASUR4	7 MEASUR5
1	MALE	PEPSI	9	1	6	8	1
2	MALE	COKE	6	7	1	8	0
3	FEMALE	COKE	9	8	2	9	8
4	MALE	3	7	9	0	5	9
5	MALE	4	7	1	6	2	8
6	FEMALE	COKE	6	0	0	8	3

или без заголовка (в выпадающем меню *Вид* напротив пункта *Заголовок* не будет стоять галочки).



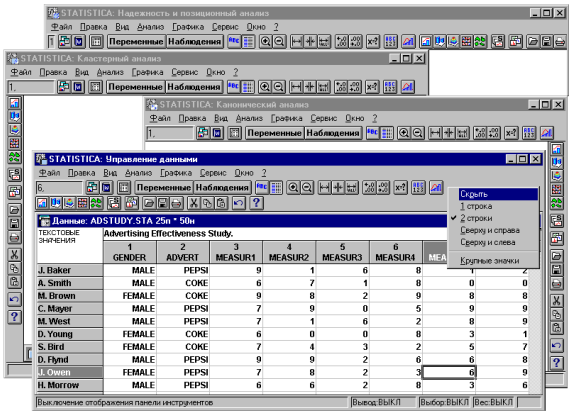
	1 GENDER	2 ADVERT	3 MEASUR1	4 MEASUR2	5 MEASUR3	6 MEASUR4	7 MEASUR5
1	MALE	PEPSI	9	1	6	8	1
2	MALE	COKE	6	7	1	8	0
3	FEMALE	COKE	9	8	2	9	8
4	MALE	3	7	9	0	5	9
5	MALE	4	7	1	6	2	8
6	FEMALE	COKE	6	0	0	8	3
7	FEMALE	COKE	7	4	3	2	5

В качестве заголовка таблицы исходных данных используется заголовок файла данных (изменяется в диалоговом окне *Заголовок файла данных, примечания и Рабочая книга*, стр. 1192).

Отмена выбора заголовков приведет только к тому, что они не будут отображены на экране (т.е. *заголовок файла* не будет удален). Заголовок можно изменить, дважды щелкнув на области заголовка таблицы исходных данных (или выбрав пункт *Заголовки* выпадающего меню *Правка*), чтобы открыть диалоговое окно *Заголовок файла данных, примечания и Рабочая книга* (см. стр. 1192).

Панель инструментов

С помощью этого пункта выпадающего меню *Вид* можно сделать выбор, отображать ли панель инструментов в окне, и, если да, то каким образом (описано ниже).



Скрыть. Этот режим предназначен для того, чтобы временно выключить отображение панели инструментов в окне. (Для того, чтобы включить отображение панели инструментов снова, нужно выбрать один из описанных ниже вариантов.)

1 строка. Если выбран этот вариант, панель инструментов будет отображаться в окне таблицы исходных данных в одну строку.

2 строки. Если выбран этот вариант, панель инструментов будет отображаться в две строки.

StatSoft
Copyright © StatSoft, 1995

Сверху и справа. Если выбран этот вариант, половина панели инструментов будет отображаться в верхней части окна, а вторая половина будет отображаться в правой части окна..

Сверху и слева. Если выбран этот вариант, половина панели инструментов будет отображаться в верхней части окна, а вторая половина будет отображаться в левой части окна.

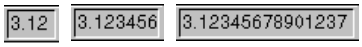
Крупные значки. С помощью этой команды можно переключаться между крупными или нормальными кнопками панели инструментов (если стоит галочка, будут отображаться большие кнопки, что, возможно, будет предпочтительнее при использовании высоких разрешений, например 1024x768, на маленьком мониторе).

Нормальные значки: Крупные значки:

Быстрый вызов. Чтобы изменить конфигурацию, можно дважды щелкнуть на панели инструментов. Меню конфигурации также доступно по нажатию правой кнопки мыши на панели инструментов.

Изменить поле просмотра

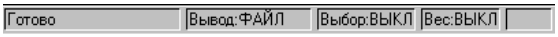
В окошке *Поле просмотра* отображается содержание выделенной ячейки (или позиция курсора, или угол поворота текста в графиках, или номер строки и столбца в *Окне текста/вывода*). Выберите этот пункт, чтобы изменить размер этого поля на один из трех возможных.



Если отображаемая ширина значений переменной, указанная в диалоговом окне таблицы исходных данных *Переменная…*(*Изменение (текущих) спецификаций*, стр. 1170) меньше, чем необходимая для того, чтобы видеть все значащие цифры, можно воспользоваться этой опцией для изменения ширины окошка *Поля просмотра* и отображения в нем всех знаков значения переменной. Кроме того, изменить ширину этого окошка можно, щелкнув на нем курсором мыши.

Строка состояния

С помощью этого параметра можно включать и выключать отображение строки состояния окна таблицы исходных данных (строки в нижней части экрана; см. главу 2). Если установлена галочка, то строка состояния будет отображаться.



Во время выполнения задачи система *STATISTICA* будет использовать строку состояния для отображения состояния выполнения задачи (см. *Индикатор состояния*, ниже).

Комментарий

Эта область строки состояния предназначена для краткого (в одну строку) описания выделенного выпадающего или контекстного меню или команды. Если ничего не выделено (данные не обрабатываются и система *STATISTICA* ожидает следующей команды), будет показано слово *Готово*. Чтобы открыть основное меню *?(Справка)*, нужно дважды щелкнуть на области комментария.

Вывод

В этой области отображается текущее устройство вывода результатов анализа (указанное в диалоговом окне *Параметры страницы/вывода*, стр. 1170). После двойного щелчка на области *Вывод* строки состояния

Copyright © StatSoft, 1995

StatSoft®

откроется диалоговое окно *Параметры страницы/вывода*, в котором можно выбрать, куда будет производиться вывод (*Выкл.*, *Файл*, *Принтер* и/или *Окно*), а также другие параметры.

Выбор

В этой области отображается текущее состояние (*Вкл.* или *Выкл.*) условий выбора наблюдений. После двойного щелчка на области *Выбор* строки состояния откроется диалоговое окно *Условия выбора наблюдений*, в котором можно просмотреть и/или задать условия выбора наблюдений (см. стр. 1223).

Вес

В этой области отображается текущее состояние (*Вкл.* или *Выкл.*) весов переменных. После двойного щелчка на области *Вес* строки состояний откроется диалоговое окно *Задание веса*, в котором можно ввести веса (см. стр. 1225).

Запись

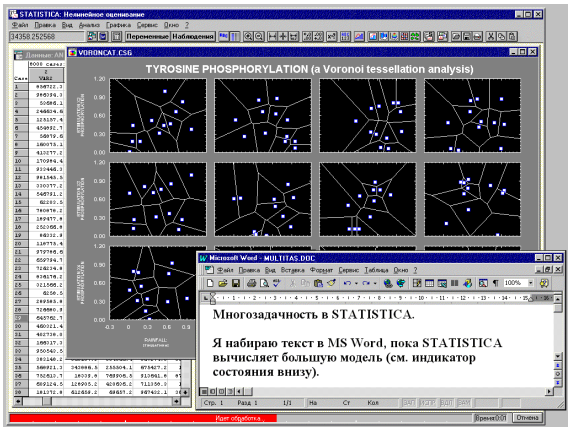
При записи макрокоманды (последовательности нажатия клавиш или движений мыши) в правой части строки состояния будет мигать слово *ЗАП*. Для более подробной информации о записи макрокоманд см. главу 6 *Кнопки автозадач*.

Индикатор состояния

Если система *STATISTICA* выполняет задачу, строка состояния (см. выше) будет отображать состояние анализа.

Состояние

Эта область индикатора состояния отображает, какая часть анализа была произведена.



Затраченное время

Система *STATISTICA* отображает время в секундах, затраченное на выполнение задачи.

Отмена

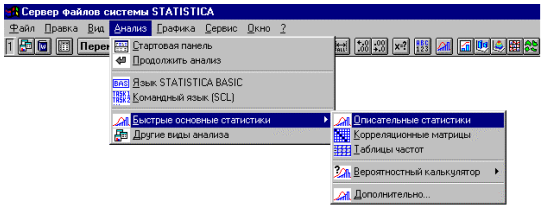
Нажатие на эту кнопку приведет к тому, что выполнение задачи будет прервано. Рисование графика можно прервать щелчком кнопки мыши в любом месте экрана.

Контролировать отображение поля *Индикатор состояния* для *статистических графиков* и *быстрых статистических графиков* можно с помощью диалогового окна *Параметры по умолчанию: статистические графики*.

StatSoft
Copyright © StatSoft, 1995

ВЫПАДАЮЩЕЕ МЕНЮ *АНАЛИЗ* ТАБЛИЦЫ ИСХОДНЫХ ДАННЫХ

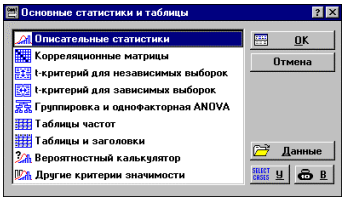
Следующие диалоговые окна и команды доступны из выпадающего меню *Анализ* таблицы исходных данных. Обратите внимание, что список команд, отличных от команд *Стартовая панель, Продолжить анализ*, *Язык STATISTICA BASIC*, *Командный язык SCL*, *Быстрые основные статистики* и *Другие виды анализа*, будет изменяться в зависимости от модуля. Подробную информацию о них можно найти в главах, где описаны конкретные модули. Показанное ниже выпадающее меню *Анализ* относится к модулю *Сервер файлов* и содержит команды, общие для всех модулей системы *STATISTICA*.



Каждая из этих командподробно описана ниже.

Стартовая панель

Эта команда выпадающего меню *Анализ* откроет окно, в котором можно выбрать анализ из списка доступных анализов активного модуля системы *STATISTICA*. Выбор здесь может быть сделан с помощью двойного щелчка на соответствующем пункте или путем выделения его в списке и нажатия кнопки *OK*. Показанная ниже стартовая панель относится к модулю *Основные статистики и таблицы*.



Комбинация клавиш. CTRL+T

Данные

Нажатие на эту кнопку откроет стандартное диалоговое окно *Открытие файла данных* (стр. 1144), в котором можно выбрать файл данных для анализа. Она аналогична подобной команде выпадающего меню *Файл* (см. стр. 1143) и действию кнопки *Открыть* панели инструментов.

Выбор наблюдений (У)

Нажатие на эту кнопку откроет стандартное диалоговое окно *Условия выбора наблюдений* (стр. 1223). В этом диалоговом окне может быть выбрано подмножество наблюдений открытого для анализа файла данных.

Задать вес (В)

Нажатие на эту кнопку откроет диалоговое окно *Задание веса* (стр. 1225), в котором можно задать веса наблюдений для открытого файла данных.

Продолжить анализ

Если проводимый анализ все еще активен, но контрольное диалоговое окно текущего анализа свернуто (значок анализа отображается в левом нижнем углу экрана), то выбор этой команды приведет к продолжению анализа, при этом откроется последнее диалоговое окно задания параметров анализа. Кроме того, для того, чтобы продолжить анализ, можно дважды нажать кнопку Cont или (в большинстве случаев)

Copyright © StatSoft, 1995

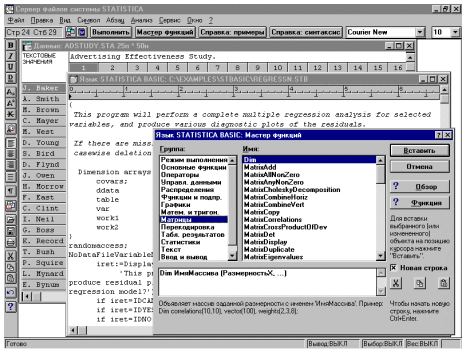
StatSoft®

нажать кнопку *Далее* в верхнем левом углу текущей таблицы результатов или окна графика.

Комбинация клавиш. CTRL+R.

Язык *STATISTICA BASIC*

Выбор пункта *Язык STATISTICA BASIC* выпадающего меню *Анализ* таблицы исходных данных приведет к открытию окна редактора программ на языке *STATISTICA BASIC*.

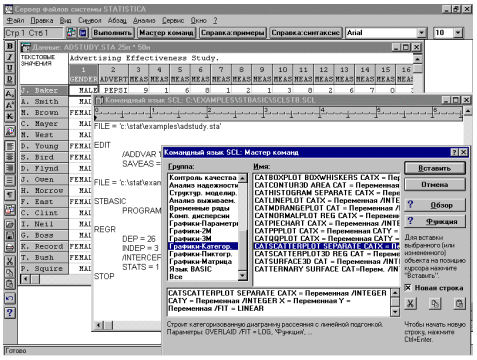


Язык STATISTICA BASIC представляет собой достаточно простой, но мощный язык программирования, который может быть использован для разработки различных типов приложений от простейших преобразований данных, которые не позволяют делать формулы таблицы исходных данных (например, циклы, функции перекодирования), до построения своих сложных постоянных расширений к процедурам вычисления, создания графиков и управления данными в системе *STATISTICA*.

Подробную информацию о языке *STATISTICA BASIC* см. в *Электронном руководстве*.

Командный язык (SCL)

Выбор пункта *Командный язык (SCL)* выпадающего меню *Анализ* приведет к открытию окна редактора программ *Командный язык SCL*.



Командный язык системы STATISTICA (сокращенно *SCL*) позволяет запустить систему *STATISTICA* в командном режиме, то есть позволяет писать “программы”, которые будут сообщать системе, что ей нужно делать. Эти программы на языке *SCL* могут затем быть запущены в любом модуле системы *STATISTICA* без дополнительного ввода. Программы на языке *SCL* могут также быть сохранены для дальнейшего частого использования с помощью команды *Сохранить как* выпадающего меню *Файл*. Среда *Командного языка системы STATISTICA* может быть использована для ускорения вызова проверенных или часто исполняемых программ на языке *SCL* или для написания “безотказных” готовых программ для людей, незнакомых с системой *STATISTICA*. Кроме того, программы на языке *SCL* могут быть поставлены в соответствие кнопкам панели инструментов *Кнопки автозадач* (см. главу 6).

Подробную информацию о *Командном языке системы STATISTICA* можно найти в *Электронном руководстве*.

Быстрые основные статистики

Меню *Быстрые основные статистики* доступно из вызываемого правой кнопкой мыши контекстного меню таблицы исходных данных или таблицы результатов, а также из выпадающего меню *Анализ* любого модуля

StatSoft®

Copyright © StatSoft, 1995

системы *STATISTICA*. Меню *Быстрые основные статистики* может быть использовано в любой точке анализа (например, для получения дополнительной информации при просмотре результатов в любом модуле системы *STATISTICA*).

Эти процедуры включают набор основных статистик, которые можно рассчитать для длинных списков переменных (например, коэффициенты корреляции *всех* переменных в наборе данных). Все анализы можно проводить *по группам* для каждого значения (т.е. кода) выбранной группы переменной.

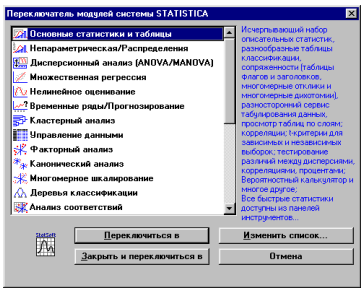
Если меню *Быстрые основные статистики* вызвано из выпадающего меню *Анализ*, то переменные для анализа выбираются в последующих диалоговых окнах выбора переменных.

Если меню *Быстрые основные статистики* вызывается из контекстного меню, то система *STATISTICA* не будет выводить запрос ввода переменных для анализа, так как выбор переменных в этом случае будет зависеть от текущего положения курсора/блока в таблице исходных данных или таблице результатов. Программа выведет запрос переменных, только если выбран пункт *Дополнительно…*, или если выделенный блок не содержит достаточного числа переменных (например, если он содержит только одну переменную для *корреляционных матриц*, или один список переменных для анализа, производимого *по… группам*).

Более подробную информацию о меню *Быстрые основные статистики* см. в главе 10.

Другие виды анализа (Переключатель модулей системы *STATISTICA*)

Выбор пункта *Другие виды анализа* выпадающего меню *Анализ* приведет к открытию диалогового окна *Переключатель модулей* системы *STATISTICA*. Это диалоговое окно содержит список всех доступных модулей системы. Как только модуль будет выделен, в диалоговом окне появится его краткое описание. Это диалоговое окно также содержит еще несколько описанных ниже команд.



Кнопка панели инструментов. См. кнопку панели инструментов таблицы исходных данных.

Быстрый вызов. Для того, чтобы вызвать глобальное контекстное меню, нужно щелкнуть правой кнопкой мыши на фоне приложений. Также для быстрого вызова можно дважды щелкнуть на фоне приложений (аналогично тому, как двойной щелчок на Рабочем столе открывает окно *Переключатель задач Windows*).

Переключиться в

После того, как был выбран (выделен) модуль, в который нужно переключиться, нажатие этой кнопки приведет к переключению в этот модуль. В зависимости режима одного или нескольких приложений, выбранного в диалоговом окне *Параметры по умолчанию: общие настройки* системы *STATISTICA* (см. стр. 1225), при открытии нового модуля текущий открытый модуль (если он есть) либо останется открытым (*режим одного приложения* не выбран), либо будет закрыт (выбран *режим одного приложения*). Кроме того, в этом диалоговом окне можно дважды щелкнуть на имени модуля, чтобы переключиться в него (сохраняя текущий

Copyright © StatSoft, 1995

модуль открытым или закрыв его, см. описание режимов одного и нескольких приложений в главе 1 *Основные соглашения*).

Закрыть и переключиться в

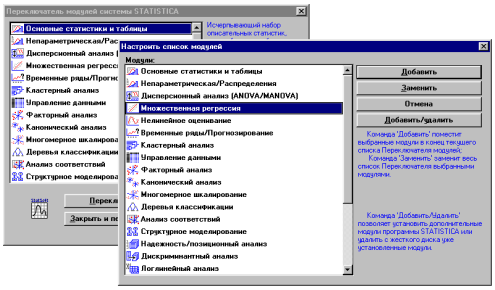
Эта кнопка предназначена для того, чтобы закрыть текущий используемый модуль и переключиться в модуль, выделенный в этом диалоговом окне. Использование этой кнопки в *режиме нескольких приложений* (см. главу 1 *Основные соглашения*, описание режимов одного и нескольких приложений) аналогично запуску системы *STATISTICA* в *режиме одного приложения*. При этом окно приложения текущего модуля будет закрыто, но оно не будет заменено другим окном; вместо этого новый модуль откроется в "следующем" окне приложения.

Изменить список…

Эта кнопка открывает диалоговое окно *Настроить список модулей* (см. ниже), в котором можно изменить (перестроить или сократить) список модулей. В диалоговом окне *Настроить список модулей* можно *заменить* весь список модулей выбранными модулями (которые часто используются) или *добавить* выбранные модули в конец списка в окне *Переключатель модулей* (перемещая редко используемые модули в конец списка).

Настроить список модулей

Диалоговое окно *Настроить список модулей* используется для выбора модулей, доступных в окне *Переключатель модулей* системы *STATISTICA* (см. ниже).



Диалоговое окно *Настроить список модулей* содержит список всех модулей, установленных с системой *STATISTICA* и три другие описанные ниже возможности.

Добавить

В этом режиме выделенные в списке модули будут добавлены в конец списка в окне *Переключатель модулей*.

Заменить

В этом режиме все выделенные в списке модули заменят список в окне *Переключатель модулей*.

Для создания другого порядка модулей в списке окна *Переключатель модулей* надо *заменить* первый модуль (это также удалит предыдущий выбор в окне *Переключатель модулей*); затем *добавить* последующие модули в нужном порядке (по одному за один раз).

Добавить/Удалить

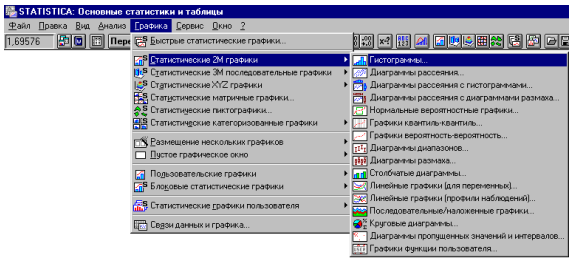
Нажатие этой кнопки приведет к изменению конфигурации системы *STATISTICA*. После нажатия этой кнопки будет запущена программа *Reinst.exe* (расположенная в каталоге системы *STATISTICA*). Она позволяет установить любые модули системы *STATISTICA,* которые не были выбраны во время предыдущей установки, убрать (удалить) любые ранее установленные модули или создать свою рабочую директорию для системы *STATISTICA* при использовании ее с сервера сети.

StatSoft®

Copyright © StatSoft, 1995

ВЫПАДАЮЩЕЕ МЕНЮ *ГРАФИКА* ТАБЛИЦЫ ИСХОДНЫХ ДАННЫХ

Графические возможности выпадающего меню *Графика* подробно рассмотрены в томе II настоящего руководства.



Следующие разделы описывают только основные графики системы *STATISTICA*. Обратите внимание, что все эти типы графиков также доступны через кнопку *Галерея графиков* панели инструментов (см. главу 2).

В дополнение к специализированным графикам, которые доступны из диалоговых окон всех статистических процедур, существуют также две основные категории графиков, доступные из меню и панелей инструментов всех таблиц результатов и таблиц исходных данных:

- *пользовательские графики* и
- *статистические* и *быстрые статистические графики*.

Далее описаны различия между этими типами графиков, которые касаются способов представления исходных данных.

Пользовательские графики. *Пользовательские графики* позволяют визуализировать любые заданные пользователем комбинации значений таблиц результатов и таблиц исходных данных (из строк, столбцов, или строк и столбцов и/или их наборов). При выборе одного из пяти глобальных типов графиков (*2М пользовательские графики*, *3М пользовательские последовательные графики*, *3М пользовательские пространственные графики*, *Пользовательские матричные графики* или *Пользовательские пиктографики*) откроется соответствующее диалоговое окно, в котором можно указать, график какой части текущей таблицы исходных данных или таблицы результатов нужно построить. Вид этого диалогового окна зависит от выбранного глобального типа *пользовательских* графиков (например, *2М пользовательские графики*). Начальный выбор данных для построения графиков, предложенный в этом диалоговом окне, зависит от текущего положения блока (или курсора) в текущей таблице исходных данных или таблице результатов. В каждом из диалоговых окон, в которых задаются *пользовательские* графики, можно выбрать определенный тип графика (внутри глобального типа). Однако, как описано ниже, после того как график создан, его тип можно изменить (с помощью диалоговых окон *Общая разметка* или *Размещение графика*, которые вызываются щелчком на фоне графика или из выпадающего меню графиков *Разметки*).

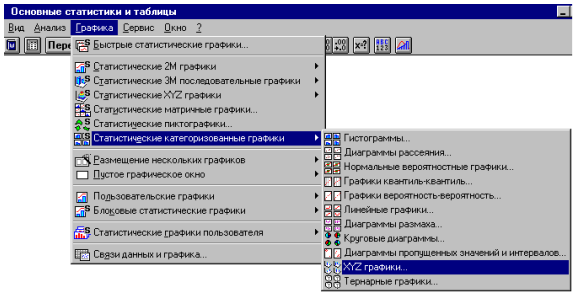
Каждый из этих *пользовательских* типов графиков также доступен через панель инструментов (см. главу 2).

Статистические графики. В отличие от *пользовательских графиков* (которые предоставляют основные инструменты для пользовательской визуализации числового вывода таблиц результатов или исходных данных, см. выше), *статистические графики* являются заданными статистическими графиками. Они доступны из окна *Галерея графиков* системы *STATISTICA,* вызываемой кнопкой *Галерея графиков* панели

Copyright © StatSoft, 1995

StatSoft®

инструментов (см. главу 2) или из выпадающего меню *Графика*.



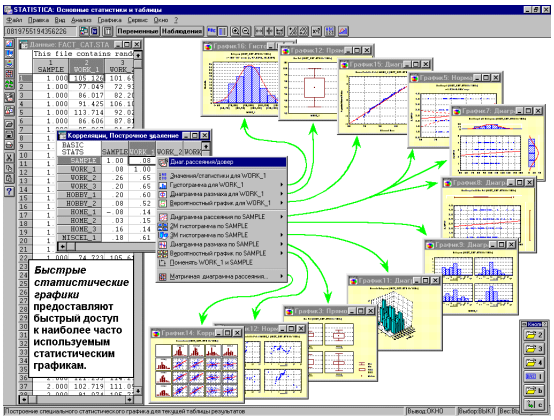
Эти графики предлагают сотни типов графического представления и аналитического обобщения данных. Они не ограничены значениями текущей таблицы результатов и не зависят от выделенного блока и положения курсора; вместо этого, они обрабатывают данные непосредственно из текущего файла данных. Здесь используются как обычные способы графического представления данных (например, построение диаграмм рассеяния, гистограмм, или графиков центральных тенденций, таких как медианы), так и стандартные графические аналитические методы (например, построение категоризованных графиков нормальной вероятности, графиков вероятности с исключенным трендом или графиков доверительных интервалов линий регрессии). При создании статистических графиков система *STATISTICA* принимает во внимание текущие условия выбора и веса наблюдений (см. стр. 1223 1225 соответственно) для выбранных для построения графиков переменных.

Быстрые статистические графики.

Быстрые статистические графики включают наиболее часто используемые типы *статистических графиков* (которые доступны из выпадающего меню *Графика*, см. предыдущий параграф). *Быстрые статистические графики* не предлагают такого количества возможностей, как *статистические графики*, однако, они быстрее выбираются, поскольку, в отличие от *статистических графиков*:

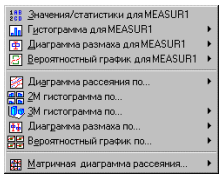
- *быстрые статистические графики* могут быть вызваны из контекстных меню и панелей инструментов (они не требуют обращения к выпадающим меню),
- *быстрые статистические графики* не требуют выбора переменных (выбор переменной определяется текущим положением курсора в таблице результатов или таблице исходных данных), а также
- *быстрые статистические графики* не требуют выбора параметров в промежуточных диалоговых окнах (при их построении используются параметры по умолчанию).

Нажатие кнопки панели инструментов, или выбор пункта *Быстрые статистические графики* (из контекстного меню или выпадающего меню *Графика*) откроет меню, в котором можно выбрать один из статистических графиков, применимых к данной переменной, то есть к переменной, соответствующей текущему положению курсора в таблице результатов или таблице исходных данных.

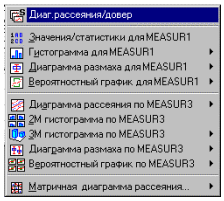


StatSoft®
Copyright © StatSoft, 1995

Во всех таблицах исходных данных и тех таблицах результатов, которые используют только один список переменных (а не два списка или одну группирующую переменную и список), контекстное меню будет иметь следующий вид (здесь “вторичная” переменная не указана, см. пустые поля *по…* в нижней части меню).

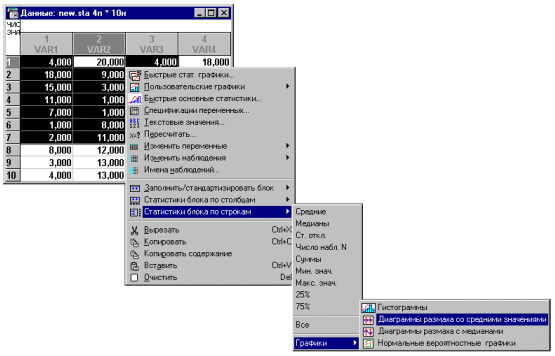


Если таблица результатов имеет матричный формат или формат, при котором положение курсора может указывать не на одну, а на две переменные, то заданные графики двух переменных для указанной пары переменных будут доступны из меню *Быстрые статистические графики* (и система *STATISTICA* не будет выдавать запрос на выбор “второй” переменной для графиков двух переменных).



Если текущее положение курсора не указывает ни на одну из переменных, то выбор любого из *быстрых статистических графиков* приведет к выводу запроса на выбор переменной из списка. При построении *быстрых статистических графиков*, система *STATISTICA* принимает во внимание текущие условия выбора и заданные веса наблюдений (см. стр. 1223 и 1225 соответственно) для выбранных для построения графиков переменных.

Блоковые статистические графики. Эти (пользовательские) графики доступны из пунктов *Статистики блока по столбцам* и *Статистики блока по строкам* (см. стр. 1202) контекстного меню или из окна *Галерея графиков* системы *STATISTICA*.



Выбор одного из этих вариантов приведет к построению соответствующего статистического обобщенного графика, позволяющего сравнивать значения в каждой строке (если выделен пункт *Статистики блока по строкам*) или в каждом столбце (если выделен пункт *Статистики блока по столбцам*) выделенного блока в таблице исходных данных или таблице результатов. Обратите внимание, что эти команды позволяют построить обобщенные статистические графики для данных (строк или столбцов) текущего блока; в этом отношении они похожи на *пользовательские графики* (см. стр. 1219), которые визуализируют данные из текущего блока.

Другие специализированные графики. В дополнение к стандартному набору *быстрых статистических графиков* (см. стр. 1220) некоторые таблицы результатов также предоставляют другие, более специализированные статистические графики (например, пиктографики регрессионных остатков, графики временных последовательностей в модуле *Временные ряды* или контурные графики в модуле *Кластерный анализ*). Как упоминалось выше, другие специализированные статистические графики, которые относятся не к определенной таблице

Copyright © StatSoft, 1995

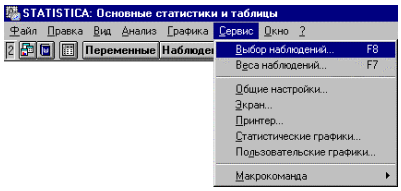
результатов, а к типу анализа (например, графики средних [например итераций] в модуле *Дисперсионный анализ*, графики аппроксимирующих кривых в модуле *Нелинейное оценивание*) доступны непосредственно из диалоговых окон с результатами (т.е., диалоговых окон, которые содержат результаты текущего анализа).

StatSoft®

Copyright © StatSoft, 1995

ВЫПАДАЮЩЕЕ МЕНЮ *СЕРВИС* ТАБЛИЦЫ ИСХОДНЫХ ДАННЫХ

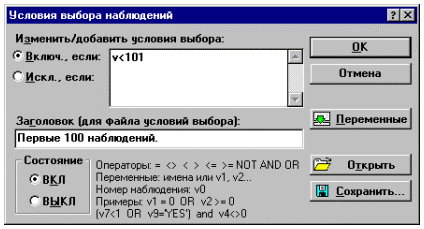
Из выпадающего меню *Сервис* таблицы исходных данных доступны следующие команды.



Ниже рассмотрены все показанные выше пункты меню, за исключением пунктов *Статистические графики* и *Пользовательские графики*, которые описаны в томе II настоящего руководства, а также пункта *Макрокоманды*, который описан в главе 6 *Кнопки автозадач*.

Выбор наблюдений

Команда *Выбор наблюдений* используется для того, чтобы включить в анализ только подмножество наблюдений (т.е. использовать “фильтры наблюдений”).



Окно *Условия выбора наблюдений* доступно по щелчку на поле *Выбор* строки состояния (см. главу 2), при нажатии кнопки *Выбор наблюдений* на *Стартовой панели* или в диалоговом окна задания параметров анализа, а также через пункт *Выбор наблюдений* выпадающего меню *Сервис*.

Комбинация клавиш. F8

Изменить/добавить условия выбора

В поле ввода можно ввести новые или изменить существующие логические условия. Заданные здесь условия выбора определяют, будет ли конкретное наблюдение *включено* в анализ (т.е. наблюдения, удовлетворяющие логическим условиям, будут анализироваться, если используется режим по умолчанию *включить*) или *исключено* из анализа (т.е. наблюдения, удовлетворяющие логическим условиям, буду исключены из анализа, если выбрана режим *исключить*).

Примеры

Следующее условие выбора включает в анализ только наблюдения от 1 до 50.

Включ., если: `v0 <= 50`

Следующее условие выбора включает в анализ только те наблюдения, значение переменной 1 в которых равно *1*.

Включ., если: `v1 = 1`

В условиях выбора наблюдений также могут быть использованы текстовые значения переменных. Обратите внимание, что текстовые значения должны быть заключены в апострофы (как показано ниже).

Включ., если: `v1 = 'MALE'`

Приведенное ниже условие исключает из анализа те наблюдения, у которых значение *v1* меньше либо равно *10, или* те наблюдения, у которых значение *v2* равно значению *v3*.

Искл., если: `(v1 <= 10) OR (v2 = v3)`

Copyright © StatSoft, 1995

StatSoft®

Описание синтаксиса

Далее приведены соглашения о синтаксисе для условий выбора наблюдений. Такие же соглашения действуют во всех случаях, когда в систему *STATISTICA* вводятся задаваемые пользователем условия выбора наблюдений.

- Ссылки на переменные могут производиться по их номерам (например, *v1 = 1*) или по именам (например, *Gender = 1*); обратите внимание, что имена переменных могут быть набраны как в верхнем, так и в нижнем регистре (т.е., *'GENDER'* и *'gender'* воспринимаются как одно и то же).
- Текстовые значения переменной должны заключаться в апострофы (например, *v1 = 'MALE'*); обратите внимание, что интерпретатор игнорирует регистр; текстовые значения могут быть набраны как в верхнем, так и в нижнем регистре (т.е., *'Yes'* и *'YES'* воспринимаются как одно и то же).
- Для задания сложных логических условий (см. пример выше) или изменения порядка выполнения операторов можно использовать скобки.
- Логические операторы:

=	- равно
# , <>, ><	- не равно
<	- меньше
>	- больше
<=	- меньше или равно
>=	- больше или равно
NOT, ~	- логическое НЕ
AND, &	- логическое И
OR, !	- логическое ИЛИ

Заголовок

Здесь может быть введена строка комментария (до 77 символов) о файле условий выбора наблюдений (который будет сохранен с условиями). Заголовок будет включен в вывод анализа, если в диалоговом окне *Параметры страницы/вывода* (page 1157) будет выбран *краткий, средний* или *полный* стиль отчета.

Состояние

Состояние этого диалогового окна отображается в строке состояния окна и показывает, влияют ли на обработку данных условия выбора наблюдений (*Вкл.*) или нет (*Выкл.*).

Переменные

Нажатие этой кнопки приведет к открытию диалогового окна *Переменные*, в котором можно посмотреть имена переменных.

Открыть

Нажатие этой кнопки приведет к открытию диалогового окна *Открытие файла с условиями выбора* (стр. 1144), в котором можно выбрать ранее сохраненный файл, содержащий условия выбора наблюдений.

Сохранить…

Нажатие этой кнопки приведет к открытию диалогового окна *Сохранение условий выбора* (стр. 1144), чтобы сохранить условия выбора в файле. Предварительно сохраненные условия выбора наблюдений могут быть открыты позже с помощью кнопки *Открыть* (см. выше).

Введенные здесь условия выбора наблюдений используются для выделения подмножества наблюдений при анализе данных. При записи в файл они хранятся вместе с заголовком в сжатом

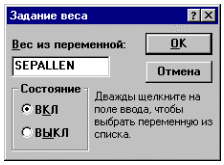
StatSoft®

Copyright © StatSoft, 1995

формате. Сохраненные условия также могут быть использованы (т.е., открыты) во всех диалоговых окнах, где для выделения нескольких подмножеств данных вводятся несколько условий выбора наблюдений (например, *Перекодировать*, *Проверить данные* и *Графики нескольких подгрупп*).

Выбор переменной для задания веса

Эта команда позволяет "изменять" вклад отдельных наблюдений в результат анализа путем "взвешивания" этих наблюдений пропорционально значениям выбранной переменной. Выбрать эту команду можно несколькими способами: щелчком на поле *Вес* строки состояния таблицы исходных данных (см. главу 2), нажатием кнопки *Вес* в окне *Стартовая панель* или в меню задания анализа, а также выбором этого пункта из выпадающего меню *Сервис*. При этом откроется диалоговое окно *Задание веса*, в котором можно указать переменную для взвешивания.



Некоторые статистические процедуры могут использовать дробный вес; в таких случаях при описании соответствующих модулей системы *STATISTICA* приводится дополнительная информация о введении весовой переменной.

Комбинация клавиш. F7

Вес из переменной

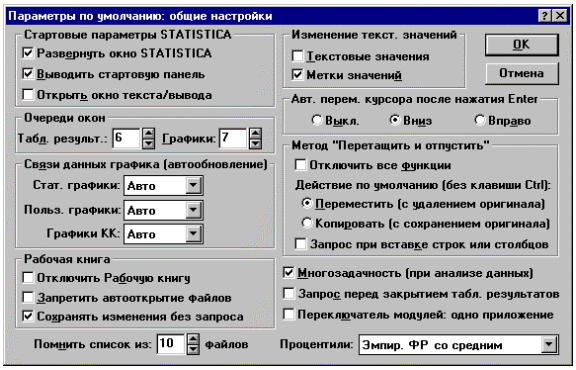
Здесь можно ввести имя переменной, содержащей значения весов, набрав его на клавиатуре или дважды щелкнув на поле ввода переменной, чтобы открыть диалоговое окно *Выбор одной переменной*. В диалоговом окне *Выбор одной переменной* отображается список переменных таблицы исходных данных, из которого можно выбрать переменную для задания веса.

Состояние

Состояние этого диалогового окна отображается в строке состояния окна приложения и показывает, используется ли одна из переменных для задания веса (*Вкл.*) или нет (*Выкл.*).

Параметры по умолчанию: общие настройки

Основные параметры отображения и некоторые операции системы *STATISTICA* могут быть изменены путем выбора (или отмены выбора) режимов, предлагаемых в диалоговом окне *Параметры по умолчанию: общие настройки*.



Заданные здесь значения будут устанавливаться по умолчанию при каждом запуске системы *STATISTICA*.

Стартовые параметры системы *STATISTICA*

Следующие операции будут выполнены при запуске модуля системы *STATISTICA*.

Copyright © StatSoft, 1995

StatSoft®

Развернуть окно *STATISTICA*. Этот режим предназначен для того, чтобы при запуске окно системы *STATISTICA* разворачивалось до размера полного экрана. Если этот режим не выбран, то система *STATISTICA* запустится в окне.

Выводить стартовую панель. Этот режим предназначен для того, чтобы стартовая панель модуля системы *STATISTICA* отображалась каждый раз, когда будет открыт модуль. Если этот режим не выбран, стартовая панель не будет отображаться при каждом запуске модуля (но, как обычно, будет доступна через пункт *Стартовая панель* выпадающего меню *Анализ* или по нажатию клавиш CTRL+T). Обратите внимание, что отключение отображения стартовой панели имеет только визуальные, а не функциональные, последствия и щелчок за ее пределами приведет к тому, что панель будет автоматически свернута.

Открыть *Окно текста/вывода*. Этот режим предназначена для того, чтобы *Окно текста/вывода* автоматически открывалось при запуске модуля системы *STATISTICA*.

Очереди окон

В системе *STATISTICA* таблицы результатов и графики автоматически распределяются по "очередям" на основе принципа "первый пришел – первый ушел". Длина каждой очереди может быть изменена либо здесь, либо в диалоговом окне *Диспетчер окон* (см. стр. 1235). Она задается числом от 0 до 20; однако, если указан *0*, то длина очереди будет бесконечной (в этом случае нужно будет закрывать окна вручную, чтобы избежать загромождения экрана и нецелесообразного расходования ресурсов компьютера).

Табл. результатов. Здесь указывается максимальное число незаблокированных таблиц результатов, которые будут открыты на экране. См. диалоговое окно *Диспетчер окон*, стр. 1235, для подробной информации о блокировании и разблокировании таблиц результатов.

Графики. Здесь указывается максимальное число незаблокированных графиков, которые будут открыты на экране. См. диалоговое окно *Диспетчер окон*, стр. 1235, для подробной информации о блокировании и разблокировании графиков.

Связи данных графика (автообновление)

При создании *статистического графика*, *быстрого статистического графика*, *блокового статистического графика* или *пользовательского графика* из таблицы исходных данных можно сохранить связь данных таблицы с данными графика в *Редакторе данных графика*. Если такая связь существует, то при редактировании данных в таблице соответствующие графики будут автоматически обновляться, отражая внесенные изменения.

Эти параметры контролируют установки по умолчанию для связей между графиками соответствующих категорий и их наборами данных. Четыре режима: *Авто*, *Ручная*, *Блокир.*, и *Нет* соответствуют режимам, контролируемым (для отдельного графика) в диалоговом окне *Связи данных и графика – Обновлять*.

Статистические графики. Выбор этого пункта установит состояние связи по умолчанию для *статистических графиков* и *быстрых статистических графиков*.

Пользовательские графики. Выбор этого пункта установит состояние связи по умолчанию для *пользовательских графиков* и *блоковых статистических графиков*.

Графики КК. Выбор этого пункта установит состояние связи по умолчанию для *Карт контроля качества*.

StatSoft®

Copyright © StatSoft, 1995

Обратите внимание, что внедренные графики (например, *диаграммы рассеяния с гистограммами*, *диаграммы рассеяния с диаграммами размаха*, внедренные графики, созданные с помощью окна *Мастер автоматического размещения графиков*, или составные графики, созданные в модуле *Контроль качества*) также могут обновляться автоматически.

Можно управлять связями между таблицей исходных данных и текущими *статистическими*, *быстрыми статистическими*, *блоковыми статистическими*, *пользовательскими* или графиками *Контроля качества*, а также внедренными графиками из диалогового окна *Связи данных и графика* (доступно из выпадающего меню *Графика* графического окна). Например, связи могут быть временно отключены, чтобы иметь возможность редактировать данные без одновременного обновления графиков.

Рабочая книга

Следующие параметры позволяют контролировать возможности Рабочей книги системы *STATISTICA*. Для подробной информации о Рабочих книгах и их параметрах см. *Заголовок файла данных, примечания и Рабочая книга*.

Отключить Рабочую книгу. Рабочая книга хранит большое количество дополнительных файлов для текущего файла данных и предлагает некоторые варианты управления этими файлами. Если выбран этот режим, Рабочая книга будет отключена. См. параграф *Если Рабочую книгу нужно отключить* в разделе *Панели инструментов* этой главы.

Запретить автооткрытие файлов. Рабочая книга позволяет выбирать до 9 дополнительных файлов, которые будут открываться автоматически при открытии файла данных (см. *Заголовок файла данных, примечания и Рабочая книга* в главе *Панели инструментов* этой главы). При выборе этого режима функция Авто - будет отключена и дополнительных файлов вместе с файлом данных открываться не будет.

Сохранять изменения без запроса. Когда сохраняется текущий файл данных, все изменения в Рабочей книге автоматически сохраняются вместе с ним. Однако если закрыть файл данных, не сохраняя его, программа либо выведет запрос о сохранении изменений в Рабочей книге, если этот режим не выбран, или автоматически сохранит изменения в Рабочей книге, не выводя запроса, если этот режим выбран. Подтверждение сохранения изменений в Рабочей книге может оказаться полезным при очень больших размерах набора данных (их сохранение будет занимать много времени); см. параграф *Если Рабочую книгу нужно отключить* в разделе *Панели инструментов* этой главы.

Помнить список

Здесь указывается число (до 16 включительно) недавно использованных файлов, которые будут отображаться в выпадающем меню *Файл*.

Изменение текстовых значений

Текстовые значения. Если этот пункт не выбран, то при вводе нового текстового значения поверх существующего, новому текстовому значению будет присвоен новый числовой эквивалент; то есть ввод нового текстового значения не повлияет на остальные текстовые значения этой переменной (связи между текстовыми и числовыми значениями текущей переменной не изменятся). Этот режим ввода данных установлен по умолчанию.

Если этот пункт отмечен, то при вводе нового текстового значения поверх существующего, введенный текст заменит существующее текстовое значение в таблице соотношений (связей) между текстовыми и числовыми

Copyright © StatSoft, 1995

значениями, и существующий числовой эквивалент (т.е., числовым значением в текущей ячейке) будет связан с новым текстовым значением. При этом значения всех ячеек текущей переменной, содержащих исходное текстовое значение, будут заменены новым текстовым значением.

Обратите внимание, что эту "двойную запись" значений можно контролировать в диалоговом окне *Диспетчер текстовых значений*.

Метки значений. Если этот пункт выбран, то при вводе новой метки значения поверх существующей, когда таблица исходных данных находится в режиме отображения *меток значений* (включить этот режим можно, нажав клавишу CTRL одновременно с кнопкой панели инструментов таблицы исходных данных), новая метка заменит существующую, а текстовое значение и числовой эквивалент останутся такими же. Этот режим установлен по умолчанию, при этом метки значений рассматриваются как "комментарии", присвоенные значениям (см. также следующий параграф).

Если этот пункт не выбран, то при вводе новой метки значения поверх существующей система *STATISTICA* сначала попытается найти совпадение с существующей меткой значения (для текущей переменной). Если совпадение обнаружено, то будут использованы текстовое значение и числовой эквивалент. Если совпадения нет, то новое текстовое значение будет создано из первых 8 символов, и новое числовое значение будет присвоено этому текстовому значению (в любом случае, изменятся и числовое значение ячейки, и связанное с ним текстовое значение). Обратите внимание, что если в итоге текстовые значения повторяются, то к каждому будет добавлено число, чтобы оно стало уникальным (например, *STATE2*). Этот режим по умолчанию не включен, метки значений рассматриваются как "расширения значений", а не только как "комментарии", присвоенные значениям.

Автоматическое перемещение курсора после нажатия ENTER

Этот параметр позволяет контролировать перемещение курсора после нажатия клавиши ENTER во время ввода или редактирования данных в таблице исходных данных. Перемещать курсор можно *вниз* на следующую ячейку в столбце (или на верхнюю ячейку следующего столбца, если он находится на последней ячейке столбца), *вправо* на следующую ячейку в строке (или в начало следующей строки, если он находится на последней ячейке строки) или *выкл.* (курсор не перемещается после нажатия клавиши ENTER).

Метод "*Перетащить и отпустить*"

Следующие параметры позволяют контролировать возможности метода *Перетащить и отпустить* в системе *STATISTICA*.

Отключить все функции. Если этот режим выбран, все функции метода *Перетащить и отпустить* в таблице исходных данных и таблице результатов будут отключены. Для изменения таблиц исходных данных и таблиц результатов нужно будет использовать стандартные функции редактирования для наблюдений и переменных из выпадающего меню *Правка*, в то время как ни одна из операций с блоками, использующая метод *Перетащить и отпустить*, не сможет быть использована.

Действие по умолчанию (без клавиши Ctrl). Выбранный здесь режим определяет действие по умолчанию метода *Перетащить и отпустить*.

Переместить (с удалением оригинала): В этом режиме, если выделенный блок перетаскивается

StatSoft®

Copyright © StatSoft, 1995

в другое место таблицы исходных данных или таблицы результатов, это блок будет *перемещен* на новое место (т.е. блок будет вырезан со своего места и вставлен на новое). Если в этом режиме нажать клавишу CTRL во время перетаскивания блока, то блок будет *скопирован* на новое место. (Это стандартный стиль Excel метода *Перетащить и отпустить*.)

Копировать (с сохранением оригинала): В этом режиме, если выделенный блок перетаскивается в другое место таблицы исходных данных или таблицы результатов, это блок будет *скопирован* на новое место (т.е. исходный блок останется на своем месте, а на новое место будет вставлена его копия). Если в этом режиме удерживать нажатой клавишу CTRL при перетаскивании блока, то блок будет *перемещен* на новое место. (Этот стиль не стандартен, однако он часто бывает полезным при управлении наборами данных в формате базы данных, когда копирование блоков используется чаще, чем перемещение.)

Запрос при вставке строк или столбцов. Этот режим служит для выведения запроса перед вставкой строк или столбцов во время операций, использующих метод *Перетащить и отпустить*, или вычислении *блоковых статистик*. В запросе указывается, какое число строк или столбцов будет вставлено; для завершения операции нужно нажать кнопку *Да*, для отмены операции – кнопку *Нет*. Если этот режим не выбран, вставка блоков будет производиться без вывода запроса.

Многозадачность (при анализе данных)

Выбор этого режима позволяет переключаться между приложениями Windows или различными модулями системы *STATISTICA* во время обработки данных системой *STATISTICA* (для вывода диалогового окна *Список задач* можно нажать клавиши CTRL+ESC, для переключения между приложениями можно нажать клавиши ALT+TAB). Если идет обработка большого файла данных, которая занимает много времени, то с помощью этого режима во время ожидания может быть использована другая программа. Однако если этот режим не выбран, данные будут обрабатываться несколько быстрее. Обратите внимание, что когда система *STATISTICA* не обрабатывает данные, всегда можно переключаться между приложениями Windows или другими модулями системы *STATISTICA*.

Также обратите внимание, что, если этот режим отключен, то прервать построение графика щелчком мыши или нажатием клавиши будет нельзя.

Запрос перед закрытием таблицы результатов

Если выбран этот режим, то при каждой попытке закрыть измененную таблицу результатов будет появляться предупреждающее сообщение. Это сообщение также предоставляет возможность сохранить измененную таблицу результатов (с расширением **.scr*) до ее закрытия. Если этот режим не выбран, то при закрытии измененной таблицы результатов предупреждения выводиться не будет.

Переключатель модулей: одно приложение

Если выбран этот режим (*Одно приложение*), то переключение между модулями (по двойному щелчку на имени модуля в окне *Переключателе модулей*, см. стр. 1217) во время сеанса системы *STATISTICA* не будет приводить к открытию новых окон приложений. Каждый новый модуль будет открыт в том же окне, заменяя использовавшийся ранее. Если этот режим не выбран (режим *Несколько приложений*), то переключаться между модулями системы *STATISTICA* (с помощью окна *Переключатель*

Copyright © StatSoft, 1995

модулей, стр. 1217) можно будет, не закрывая предыдущие. Обратите внимание, что переключение аналогичное режиму *Одно приложение* возможно и в режиме *Несколько приложений*, если в окне *Переключатель модулей* нажать кнопку *Закрыть и переключиться в* (вместо двойного щелчка на имени модуля).

Процентили

Этот параметр позволяет выбирать из шести методов вычисления значений процентилей, включая медианы (50 процентные точки) и квартили (25 и 75 процентные точки). Выбранный здесь метод будет использован во всех вычислениях медиан, квартилей и процентилей в системе *STATISTICA*, например, при вычислении:

- процентилей точек в модуле *Непараметрическая статистика*;
- медиан и квартилей в модуле *Основные статистики*;
- медиан и квартилей в *2M* и *3M диаграммах размаха*;
- медиан и квартилей в минимаксной диаграмме в модуле *Воспроизводимость и повторяемость*;
- медиан и квартилей в минимаксной диаграмме в модулях *Дискриминантный анализ*, *Надежность и позиционный анализ*, *Основные статистики*, *Множественная регрессия* и других модулях системы *STATISTICA,* у которых в диалоговом окне предлагается построение минимаксных диаграмм.

Пусть для следующих методов n будет числом наблюдений, а p - процентным значением, деленным на 100 (например, *50/100 = 0.5* для медианы). Тогда возможны следующие варианты вычисления процентилей:

Метод взвешенного среднего X_{np}. (Взвешенное среднее с центром в X_{np}). Представим np (n умноженное на p) как $np=j+g$ где j является целой частью np, а g – дробной частью np; затем вычислим:

$$\text{PercentileValue} = (1-g)x_j + gx_{j+1}$$

Обратите внимание, что X_0 в этих вычислениях заменяется на X_1 (например, если $j = 0$, то эта формула будет выглядеть так: $(1-g)x_1 + gx_{(0+1)}$).

Пример: Рассмотрим следующие отсортированные данные: 1, 2, 4, 7, 8, 9, 10, 12, 13

Здесь $n = 9$ и пусть $p = 0.25$ (25-ти процентная точка). Выразим np (n умноженное на p) как

$$np = 9*.25 = 2.25 = j + g$$

Здесь $j = 2$ и $g = 0.25$. Теперь вычислим 25-ти процентную точку:

$$\begin{aligned}25\text{th Percentile} &= (1-.25)x_2 + .25x_3\\ &= (.75)2 + (.25)4\\ &= 2.5\end{aligned}$$

Метод взвешенного среднего $X_{(n+1)p}$. (Взвешенное среднее с центром в $X_{(n+1)p}$). Представим $(n+1)p$ в виде $(n+1)p=j+g$, где j – целая часть $(n+1)p$, а g – дробная часть $(n+1)p$; затем вычислим процентиль по формуле:

$$\text{PercentileValue} = (1-g)x_j + gx_{j+1}$$

Обратите внимание, что X_{n+1} в этих вычислениях заменяется на X_n (например, если $j = n$, то эта формула будет выглядеть так: $(1-g)x_n + gx_n$).

Пример: Рассмотрим следующие отсортированные данные: 1, 2, 4, 7, 8, 9, 10, 12, 13

Здесь $n = 9$ и пусть $p = 0.25$ (25-ти процентная точка). Выразим np (n умноженное на p) как

StatSoft

Copyright © StatSoft, 1995

$(n+1)p = 10*.25 = 2.5 = j + g$

Здесь $j = 2$ и $g = 0.5$. Теперь вычислим 25-ти процентную точку по формуле:

$$25\text{th Percentile} = (1-.5)x_2 + .5x_3$$
$$= (.5)2 + (.5)4$$
$$= 3.0$$

Метод эмпирического распределения функции. Представим *np* (*n* умноженное на *p*) в виде *np=j+g,* где *j* – целая часть *np*, а *g* – дробная часть *np*; затем определим значение процентной точки:

$$\text{PercentileValue} = x_j \quad \text{if } g=0$$
$$\text{PercentileValue} = x_{j+1} \quad \text{if } g>0$$

Пример: Рассмотрим следующие отсортированные данные: 1, 2, 4, 7, 8, 9, 10, 12, 13

Здесь $n = 9$ и пусть $p = 0.25$ (25-ти процентная точка). Выразим *np* (*n* умноженное на *p*) как:

$np = 9*.25 = 2.25 = j + g$

Получим $j = 2$ и $g = 0.25$. Теперь, поскольку $g>0$, вычислим 25-ти процентную точку:

$$25\text{th Percentile} = x_3$$
$$= 4.0$$

Метод эмпирического распределения функции с осреднением. Представим *np* (*n* умноженное на *p*) в виде *np=j+g,* где *j* – целая часть *np*, а *g* – дробная часть - *np*; затем вычислим значение процентной точки по формуле:

$$\text{PercentileValue} = (x_j + x_{j+1})/2 \quad \text{if } g=0$$
$$\text{PercentileValue} = x_{j+1} \quad \text{if } g>0$$

Пример: Рассмотрим следующие отсортированные данные: 1, 2, 4, 7, 8, 9, 10, 12, 13

Здесь $n = 9$ и пусть $p = 0.25$ (25-ти процентная точка). Выразим *np* (*n* умноженное на *p*) как:

$np = 9*.25 = 2.25 = j + g$

Получим $j = 2$ и $g = 0.25$. Теперь, поскольку $g>0$, вычислим 25-ти процентную точку:

$$25\text{th Percentile} = x_3$$
$$= 4.0$$

Метод эмпирического распределения функции с интерполяцией (MS Excel). Представим *(n-1)p* (*(n-1)* умноженное на *p*) в виде *(n-1)p=j+g,* где *j* – целая часть *(n-1)p*, а *g* – дробная часть *(n-1)p*; затем вычислим значение процентной точки:

$$\text{PercentileValue} = x_{j+1} \quad \text{if } g=0$$
$$\text{PercentileValue} = x_{j+1} + g(x_{j+2} - x_{j+1}) \quad \text{if } g>0$$

Пример: Рассмотрим следующие отсортированные данные: 1, 2, 4, 7, 8, 9, 10, 12, 13

Здесь $n = 9$ и пусть $p = 0.25$ (25-ти процентная точка). Выразим *np* (*n* умноженное на *p*) как:

$(n-1)p = 8*.25 = 2.0 = j + g$

Получим $j = 2$ и $g = 0$. Теперь, поскольку $g=0$, вычислим 25-ти процентную точку как:

$$25\text{th Percentile} = x_3$$
$$= 4.0$$

Метод ближайшего наблюдения. (Наблюдение, ближайшее к *np*). Определим *j* как целую часть *np+1/2*, затем вычислим:

$$\text{PercentileValue} = x_j$$

Пример: Рассмотрим следующие отсортированные данные: 1, 2, 4, 7, 8, 9, 10, 12, 13

Здесь $n = 9$ и пуст ь $p = 0.25$ (25-ти процентная точка). Выразим *np* (*n* умноженное на *p*) как:

$np + 1/2 = (9*.25) + .5 = 2.75 = j + g$

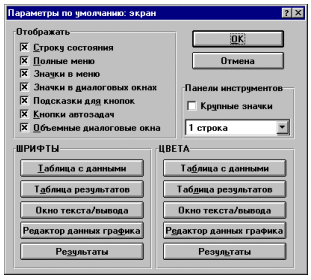
Получим $j = 2$. Теперь вычислим 25-ти процентную точку как:

Copyright © StatSoft, 1995

25th Percentile = x_2
= 2.0

Параметры по умолчанию: экран

Отображение окон системы *STATISTICA* может быть изменено путем выбора (или отмены выбора) некоторых режимов из диалогового окна *Параметры по умолчанию: экран*. Выбранные здесь параметры будут по умолчанию использованы для отображения окон при каждом запуске системы *STATISTICA*.



Обратите внимание, что выбранные здесь цвета и шрифт не будут отображены в текущей таблице исходных данных или таблице результатов (для этого нужно использовать соответствующие команды меню *Вид*, см. стр. 1208). Этот выбор повлияет только на конфигурацию, используемую системой *STATISTICA* при создании новых таблиц исходных данных (с помощью команды *Создать данные*, стр. 1142) или таблиц результатов.

Отображать строку состояния

Если выбран этот пункт, то будет отображаться строка состояния (в нижней части экрана, см. главу 2). Если этот пункт не выбран, строка состояния отображаться не будет.

Полные меню

Если этот пункт выбран, то на экране будут отображаться полные названия меню (названия команд и соответствующие им комбинации клавиш, если они есть). Если он не выбран, то для всех выпадающих меню будут отображаться только названия команд.

Значки в меню

Этот параметр позволяет установить отображение значков в пунктах меню системы *STATISTICA*. Если этот пункт не выбран, то не будут отображаться значки в меню, вызываемых правой кнопкой мыши, кнопками панели инструментов, а также в выпадающих меню (после того как выбор этого режима отменен, надо закрыть модуль и открыть его снова, чтобы значки выпадающего меню исчезли).

Значки в диалоговых окнах

Если выбран этот пункт, то значки будут использованы в диалоговых для обозначения доступных вариантов. Отмена выбора этого режима приведет к тому, что значки отображаться не будут.

Подсказки для кнопок

Если этот пункт выбран, то всплывающие *подсказки для кнопок* будут отображаться при перемещении курсора к каждой из кнопок панели инструментов. Если этот пункт не выбран, то *подсказки для кнопок* отображаться не будут.

Кнопки автозадач

Если в этом диалоговом окне (или из выпадающего меню *Вид*) выбрать пункт *Кнопки автозадач*, в окне системы *STATISTICA* будет (по умолчанию) отображена панель инструментов *Кнопки автозадач* (см. главу 6). Если этот пункт не выбран, панель инструментов *Кнопки автозадач* отображаться не будет. Обратите внимание, что все “горячие” клавиши, назначенные для *кнопок автозадач*, могут быть использованы вне зависимости от установок отображения панели инструментов.

StatSoft®
Copyright © StatSoft, 1995

Комбинация клавиш. Для включения или выключения панели инструментов *Кнопки автозадач* можно использовать клавиши CTRL+M.

Объемные диалоговые окна

По умолчанию диалоговые окна системы *STATISTICA* отображаются с эффектом объема для улучшения читаемости надписей на элементах диалогового окна. Эффект объема можно убрать, отменив выбор этого пункта.

Шрифты

Нажав на соответствующие кнопки, можно выбрать шрифт по умолчанию для содержания *таблиц результатов*, *таблиц с данными*, *редактора данных графика* или *результатов* (общего раздела диалоговых окон с результатами).

Цвета

Цвет *таблиц с данными*, *таблиц результатов* и *редактора данных графика* можно изменить, нажав соответствующую кнопку для вывода диалогового окна *Настройка цветов по умолчанию*. Можно также нажать кнопку *Результаты*, чтобы открыть диалоговое окно *Текст результатов – Цвета* (см. ниже), где можно задать цвета для отображения различных частей результата.

Панели инструментов

Этот параметр позволяет выбрать способ отображения панели инструментов в окне по умолчанию.

Крупные значки. Выберите этот режим, чтобы на панели инструментов отображались кнопки большого размера. Это может оказаться полезным, когда высокое разрешение (например, 800x600 или 1024x768) используется на мониторе маленького размера.

Вид панели инструментов. Здесь можно выбрать вид отображения панели инструментов.

Скрыть. Выбор этого варианта приведет к тому, что панель инструментов будет временно удалена из окна. Чтобы снова включить панель инструментов, нужно выбрать один из описанных ниже вариантов.

1 строка. Если выбран этот вариант, панель инструментов будет расположена в одну строку.

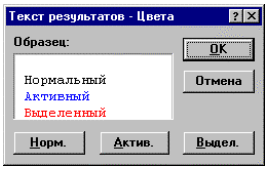
2 строки. Если выбран этот вариант, панель инструментов будет расположена в две строки.

Сверху и справа. Если выбран этот вариант, часть панели инструментов будет отображаться в верхней части окна, а другая часть - вдоль правой границы окна.

Сверху и слева. Если выбран этот вариант, часть панели инструментов будет отображаться в верхней части окна, а другая часть - вдоль левой границы окна.

Текст результатов - Цвета

Это диалоговое окно позволяет изменять цвета отдельных компонент общего раздела диалоговых окон результатов.



После нажатия на соответствующую кнопку откроется диалоговое окно *Цвет* (стр. 1209), в котором можно выбрать цвет *нормального* текста (текст, которым написаны названия отдельных полей), *активного* текста (текст, которым

Copyright © StatSoft, 1995

StatSoft®

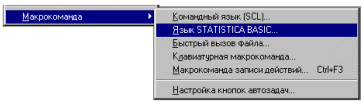
написаны значения полей) или *выделенного* текста (цвет текста в специальных диалоговых окнах, таких как *Множественная регрессия – Результаты*, который используется для обозначения значимости). В области *Образец* до закрытия диалогового окна можно посмотреть сделанные изменения.

Параметры по умолчанию: принтер

При выборе пункта *Принтер* выпадающего меню *Сервис* таблицы исходных данных открывается диалоговое окно *Параметры страницы/вывода*. Параметры, задаваемые в этом окне, подробно описаны на стр. 1157.

Сервис - Макрокоманды

Показанные ниже команды доступны через пункт *Макрокоманда* выпадающего меню *Сервис* таблицы исходных данных.



Выбор одной из этих команд приведет к открытию соответствующего диалогового окна *Кнопки автозадач* для создания выбранного типа макрокоманды. Подробную информацию о всплывающей или фиксированной панели инструментов *Кнопки автозадач* можно найти в главе 6.

Командный язык (*SCL*)

Выбор этого пункта из выпадающего меню *Сервис* приведет к открытию диалогового окна *Изменение кнопки автозадач*. Здесь можно выбрать файл на командном языке (*SCL)* системы *STATISTICA* и присвоить его *кнопке автозадач*.

Язык *STATISTICA BASIC*

Выбор этого пункта из выпадающего меню *Сервис* приведет к открытию диалогового окна *Изменение кнопки автозадач*, в котором можно выбрать файл на языке *STATISTICA BASIC* и присвоить его *кнопке автозадач*.

Быстрый вызов файла

Выбор этого пункта из выпадающего меню *Сервис* приведет к открытию диалогового окна *Изменение кнопки автозадач*, в котором можно выбрать таблицу исходных данных, таблицу результатов, график или файл отчета и присвоить его *кнопке автозадач*.

Клавиатурная макрокоманда

Выбор этого пункта из выпадающего меню *Сервис* приведет к открытию диалогового окна *Запись или ввод клавиатурной макрокоманды*. Здесь можно ввести или записать новую клавиатурную макрокоманду или открыть сохраненную ранее клавиатурную макрокоманду и присвоить ее *кнопке автозадач*.

Макрокоманда записи действий

Выбор этого пункта из выпадающего меню *Сервис* приведет к открытию диалогового окна *Макрокоманда записи действий*, в котором можно записать новую макрокоманду или открыть ранее сохраненную макрокоманду и присвоить ее *кнопке автозадач*.

Настройка кнопок автозадач

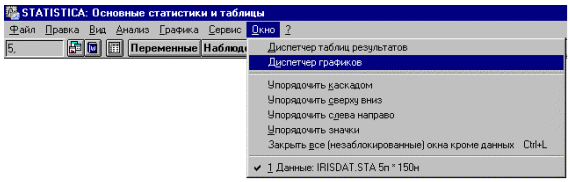
Выбор этого пункта из выпадающего меню *Сервис* приведет к открытию диалогового окна *Настройка кнопок автозадач*, где можно задать и изменить отдельные кнопки на панели инструментов *Кнопки автозадач*.

StatSoft®

Copyright © StatSoft, 1995

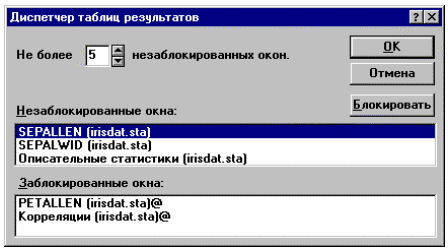
ВЫПАДАЮЩЕЕ МЕНЮ *ОКНО* ТАБЛИЦЫ ИСХОДНЫХ ДАННЫХ

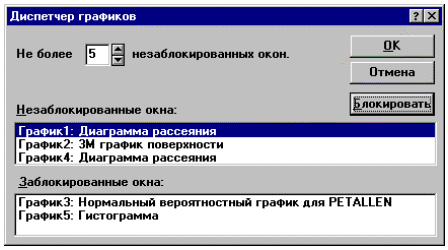
Выпадающее меню *Окно* содержит показанные ниже команды.



Диалоговые окна *Диспетчер таблиц результатов* и *Диспетчер графиков*

Выбор одного из этих пунктов, дает возможность изменять число (незаблокированных, см. ниже) таблиц результатов или графиков, которые будут одновременно отображаться на экране.





В системе *STATISTICA* таблицы результатов и графики распределяются по отдельным “очередям” на основе метода “первый пришел, первый ушел”.

Например, если длина очереди таблиц результатов равна 3 и создается четвертая таблица, то первая таблица результатов будет автоматически закрыта. Рекомендуемой длины очереди не существует, число мест в очереди нужно выбирать в соответствии с типом анализа, размером экрана и личными предпочтениями. Изменить длину очереди можно в диалоговом окне *Диспетчер окон*, или в диалоговом окне *Параметры по умолчанию: общие настройки* (стр. 1225). Число окон (таблиц результатов или графиков) на экране можно также контролировать, блокируя их или закрывая самостоятельно.

Очередь

Здесь указывается максимальное число незаблокированных окон, которые будут отображаться на экране. Число окон может быть в диапазоне от 0 до 20; при этом значение 0 соответствует бесконечной очереди (в этом случае окна нужно будет закрывать самостоятельно, чтобы избежать загромождения экрана и траты ресурсов компьютера).

Незаблокированные окна

Это список всех незаблокированных (незащищенных) окон, отображенных на экране. Если выделить в этом списке окно и нажать кнопку *Блокировать* (см. ниже), окно будет заблокировано и удалено из очереди.

Заблокированные окна

Это список всех заблокированных окон, отображенных на экране. Добавить окно в это список можно, выделив его имя в списке *Незаблокированные окна* (см. выше) и нажав кнопку *Блокировать* (см. ниже). Чтобы удалить с экрана заблокированное окно, его нужно закрыть вручную.

Copyright © StatSoft, 1995

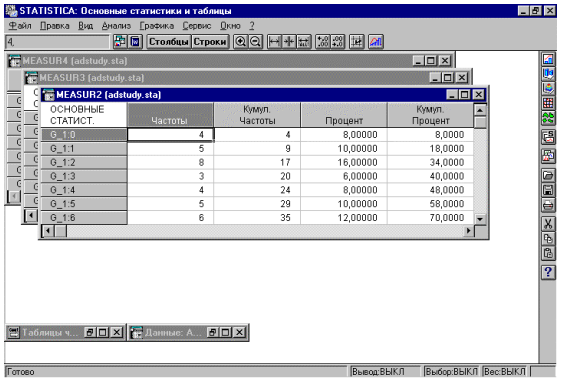
StatSoft®

Блокировать

Отдельные окна можно убрать из очереди, заблокировав их с помощью кнопки *Блокировать*. Чтобы заблокировать окно, сначала нужно выделить его имя в списке *Неблокированные окна* (см. выше), а затем нажать кнопку *Блокировать*. Заблокированные окна (они отмечаются символом "@" в конце заголовка) останутся на экране до тех пор, пока не будут закрыты вручную, или пока не будет закрыт модуль или система *STATISTICA*.

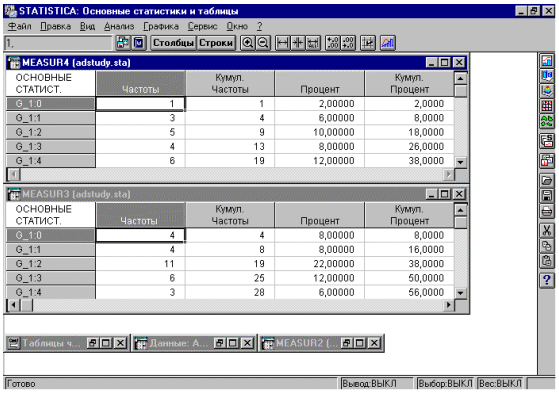
Упорядочить каскадом

Выбор этого пункта приведет к тому, что открытые окна системы *STATISTICA* будут расположены на экране, частично перекрывая друг друга, при этом будет видна строка заголовка каждого окна.



Упорядочить сверху вниз

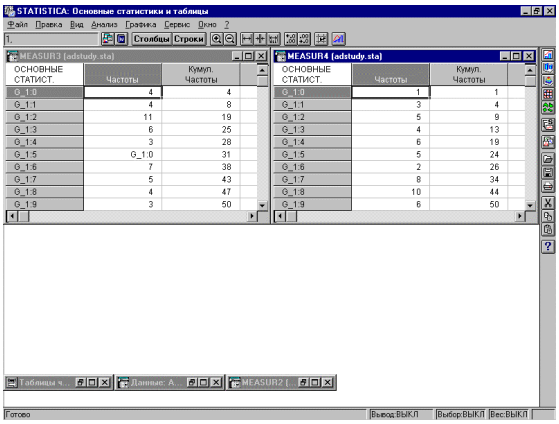
Выбор этого пункта приведет к тому, что открытые окна системы *STATISTICA* будут расположены по горизонтали сверху вниз (бок о бок).



Если выбран этот режим, система *STATISTICA* автоматически оптимизирует отображение открытых окон (предпочтение будет отдано упорядочиванию по горизонтали). Графические окна сохранят свои пропорции, если выбран режим *Фиксировать пропорции* выпадающего меню *Вид* (см. том II настоящего руководства).

Упорядочить слева направо

Выбор этого пункта приведет к тому, что открытые окна системы *STATISTICA* будут расположены вертикально.



Если выбран этот режим, система *STATISTICA* автоматически оптимизирует отображение

StatSoft
Copyright © StatSoft, 1995

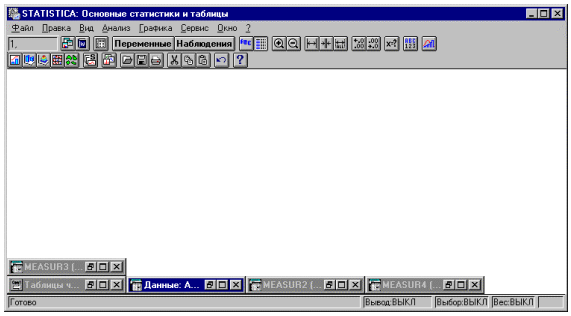
открытых окон (предпочтение будет отдано упорядочиванию по вертикали).

Графические окна сохранят свои пропорции, если в графическом окне выпадающего меню *Вид* выбран режим *Фиксировать пропорции* (см. том II настоящего руководства).

Комбинация клавиш. ALT+SHIFT+F6

Упорядочить значки

Выбор этого пункта приведет к тому, что все свернутые окна системы *STATISTICA* будут расположены по строкам.



Закрыть все (незаблокированные) окна, кроме данных

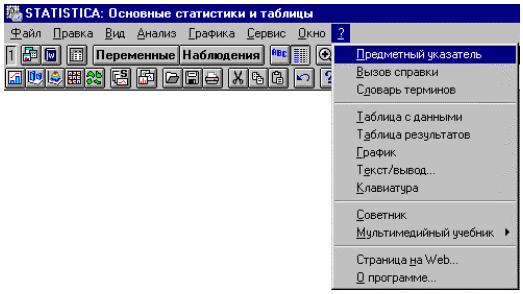
Выбор этой команды приведет к тому, что все открытые (и незаблокированные) таблицы результатов, графики и относящиеся к ним окна (например, графические данные) системы *STATISTICA* будут закрыты; таблица исходных данных останется при этом открытой. Этот режим может быть использован, когда для начала нового анализа нужно “очистить” экран. Обратите внимание, что окна открытых файлов блокируются автоматически и не закрываются при выборе этой опции.

Комбинация клавиш. CTRL+L

Copyright © StatSoft, 1995

ВЫПАДАЮЩЕЕ МЕНЮ *?(СПРАВКА)* ТАБЛИЦЫ ИСХОДНЫХ ДАННЫХ

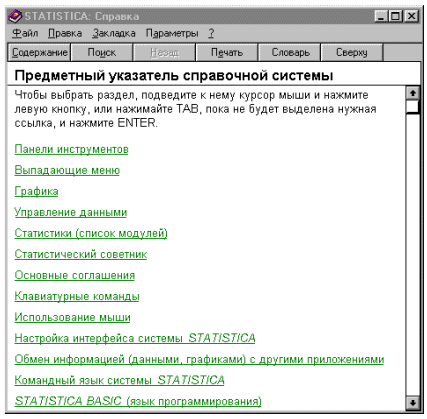
Из выпадающего меню *?(Справка)* доступны диалоговые окна и команды, предоставляющие доступ к различным частям *Электронного руководства* системы *STATISTICA*.



Каждая из этих команд открывает окно *Справка*, содержащее информацию имеющую отношение к основной теме (см. ниже).

Меню *?(Справка)* – Предметный указатель

Выбор этой команды приведет к открытию окна *Справка - Предметный указатель* системы *STATISTICA*.



Предметный указатель содержит список наиболее часто используемых заголовков *справочного руководства*. Для получения информации по теме достаточно просто щелкнуть на нужном заголовке.

Меню *?(Справка)* - Таблица с данными

Эта команда открывает раздел *Электронного руководства*, в котором приведен список меню и других параметров окна таблицы исходных данных. Чтобы получить больше информации по одной из перечисленных здесь тем, достаточно щелкнуть на нужном названии.

Меню *?(Справка)* - Таблица результатов

Аналогично пункту *Таблица с данными*, выбор этого пункта откроет раздел *Электронного руководства*, в котором приведен список меню и других параметров окна таблицы результатов. Чтобы получить информацию по теме, достаточно щелкнуть на нужном названии.

Меню *?(Справка)* - График

Выбор этого пункта приведет к тому, что откроется раздел *Электронного руководства*, в котором будет приведен список меню и других параметров графического окна. Чтобы получить информацию по теме, достаточно щелкнуть на нужном названии.

Меню *?(Справка)* - Текст/вывод

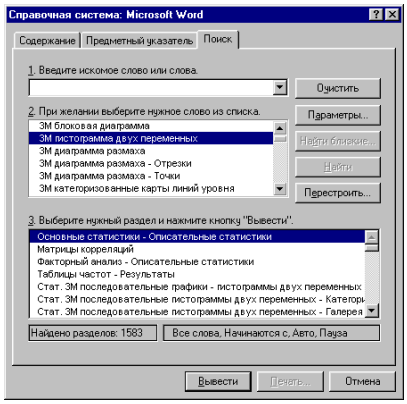
Выбор этого пункта приведет к тому, что откроется раздел *Электронного руководства*, в

StatSoft®
Copyright © StatSoft, 1995

котором будет приведен список меню и других параметров *Окна текста/вывода*. Чтобы получить информацию по теме, достаточно щелкнуть на нужном названии.

Меню *?(Справка)* - Вызов справки…

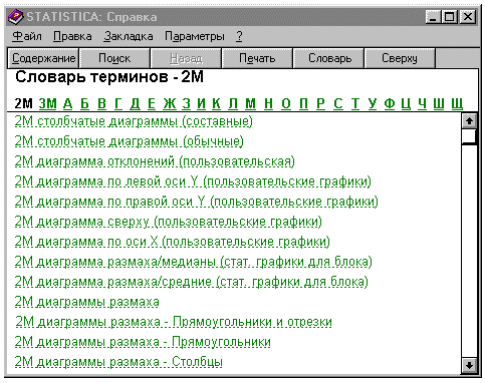
При выборе этого пункта выпадающего меню *?(Справка)* откроется стандартное диалоговое окно Windows.



Это диалоговое окно может быть использовано в качестве интерактивной справочной системы для поиска определенных заголовков в *Электронном руководстве* системы *STATISTICA* путем ввода ключевых слов и выборе тем, связанных с этими словами.

Меню *?(Справка)* - Словарь терминов

Эта команда открывает раздел *Электронного руководства*, содержащий список всех записей словаря системы *STATISTICA*, отсортированный по алфавиту.



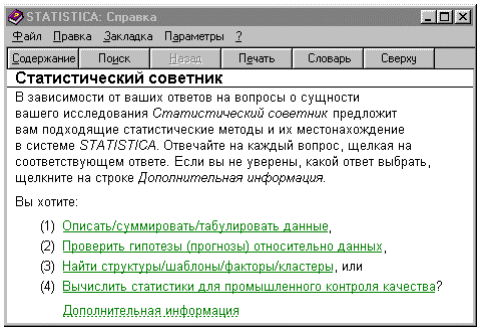
Чтобы получить информацию о нужной теме, достаточно щелкнуть на записи словаря в этом окне.

Меню *?(Справка)* - Клавиатура

Эта команда открывает раздел *Электронного руководства*, содержащий комбинации клавиш, которые могут быть использованы для ускорения доступа к различным процедурам системы *STATISTICA*.

Меню *?(Справка)* - Советник

С помощью этой команды запускается справочная программа *Статистический советник* системы *STATISTICA*.



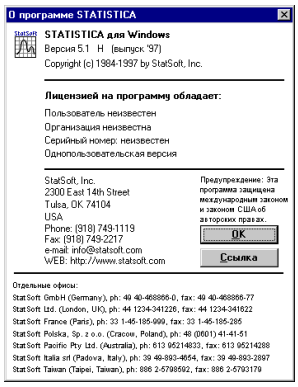
Copyright © StatSoft, 1995

StatSoft®

Если из выпадающего меню *?(Справка)* выбран пункт *Советник*, программа задаст несколько простых вопросов о природе исследуемой проблемы и типе данных, а затем предложит наиболее подходящие статистические процедуры и укажет, как их найти в системе *STATISTICA*.

Меню *?(Справка)* - О программе

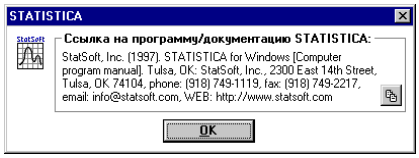
Если выбран этот пункт, откроется окно с указанием версии системы *STATISTICA*, информацией об авторских правах и лицензии (см. ниже).



Имя владельца/пользователя и название организации, отображаемые в этом диалоговом окне, записываются при первой установке системы *STATISTICA*.

Ссылка на систему *STATISTICA*

Если в диалоговом окне *О системе STATISTICA* нажать эту кнопку, то откроется окно, в котором отображается ссылка на систему *STATISTICA*



Если нажать кнопку, эта информация будет скопирована в буфер обмена для упрощения ссылки на систему *STATISTICA*.

StatSoft®
Copyright © StatSoft, 1995

ПРЕДМЕТНЫЙ УКАЗАТЕЛЬ

А

Автозаполнение, 1135
Автоматическая выдача каждой заполненной страницы, 1160, 1161
Автоматическая выдача страниц при каждом выводе, 1159, 1160
Автоматическая печать всех графиков, 1163
Автоматическая печать всех таблиц результатов, 1160
Автоматический выход из таблиц результатов и графиков, 1163
Автоматический пересчет формул в таблице исходных данных, 1184
Автоматическое перемещение курсора после нажатия ENTER, 1228
Автоотчет (автоматическая печать всех графиков), 1163
Автоотчет (автоматическая печать всех таблиц результатов), 1160

Б

Блокнот, 1211
Блоковые статистические графики, 1204
Буфер обмена и графические объекты, 1196
Быстрые основные статистики, 1216
Быстрые статистические графики (обзор), 1220
Быстрый экспорт, 1151
  разделители в свободном ASCII-формате, 1152
  файлы Excel, 1151, 1152
  файлы Lotus, 1151, 1152
  файлы Quattro, 1151, 1152
  файлы в свободном ASCII-формате, 1152

В

Вес, 1225
Вид
  заголовки, 1212
  имена наблюдений, 1208
  метки значений, 1208
  панель инструментов, 1212
  полные имена переменных, 1208
  строка состояния, 1213
  текстовые значения, 1208
Вставить
  в графики, 1196
  в таблицу исходных данных, 1196
  установить связь, 1197
Вставка с помощью метода OLE, 1197
Выбор наблюдений, 1223
Выбор переменной для задания веса, 1225
Вывод, 1161
Вызов справки…, 1240
Вырезать
  из таблицы исходных данных, 1195
Выход, 1166

Г

Галерея графиков, 1219

Д

Даты, 1180, 1183
Диалоговое окно
  открытия/сохранения файла, 1144
Диалоговое окно Параметры страницы/вывода, 1234
Динамический обмен данными
  связь, 1138, 1174, 1197, 1198, 1206
Диспетчер графиков, 1236
Диспетчер окон, 1236
Длинное имя, 1139
  метка, 1173
  связь, 1174
  формула, 1173
Длинное имя переменной, 1139
Добавить наблюдения, 1188
  как вставить наблюдения перед первым наблюдением, 1189
Добавить переменные, 1167
Дробные ранги, 1186

З

Заблокированные окна, 1236
Заголовки, 1212
Заголовок, 1193
Заголовок файла, примечания и Рабочая книга, 1192
Закрыть все, 1238
Закрыть измененную таблицу результатов, 1229
Заменить, 1199
Заполнить случайными значениями, 1201
Заполнить/копировать вниз, 1201
Заполнить/копировать вправо, 1201
Запрос подтверждения перед закрытием таблицы результатов, 1229
Захват экрана, 1205
Значения переменных
  даты, 1180

И

Изменение текстовых значений
  параметры по умолчанию, 1227
Изменить поле просмотра, 1213
Имена наблюдений, 1191, 1208
  диспетчер, 1191, 1192
  параметры, 1192
Имена переменных, 1208
Импорт
  текстовые (ASCII) файлы, 1148
  файлы ASCII (текстовые), 1148
Импорт текстовых (ASCII) файлов, 1148
Импорт файлов в формате ASCII, 1148
Индикатор состояния, 1214
Информация о файле и примечания, 1193

К

Как вставить наблюдения перед первым наблюдением, 1189, 1190
Как добавить блок перед первой переменной, 1167
Как поместить блок до первой переменной, 1168, 1169
Кнопки автозадач, 1234
Командный язык (SCL), 1140, 1216
Комбинации клавиш, 1240

Copyright © StatSoft, 1995

Копировать
в таблице исходных данных, 1195
наблюдения, 1190
наблюдения, как вставить наблюдения до наблюдения номер один, 1190
содержание, 1196
Копировать вниз, 1201
Копировать вправо, 1201
Копировать переменные, 1169
как поместить блок до первой переменной, 1169

Л

Лаг переменных, 1184

М

Мастер функций, 1175
формулы в таблице с данными, 1175
Метки значения, 1179
Метод *Перетащить и отпустить*
автозаполнение, 1135
вставка блока, 1133
выделение блока, 1132
копирование блока, 1133
копирование значений в таблице исходных данных, 1133
обзор, 1132
очистка блока, 1134
параметры по умолчанию, 1228
переменная скорость выделения блока, 1136
перемещение блока, 1133
перемещение значений в таблице исходных данных, 1133
экстраполяция блока, 1135
экстраполяция блока влево или вверх, 1136
Многозадачность, 1229
Модуль запуска, 1140

Н

Наблюдения
добавить, 1188
копировать, 1190
переместить, 1189
удалить, 1191
Найти, 1198
Настроить список модулей, 1218
Незаблокированные окна, 1236

О

О программе *STATISTICA*, 1241
Обзор синтаксиса формул таблиц исходных данных, 1176
Объемные диалоговые окна, 1233
Окна, разделение отображения, 1136
Окно текста/вывода, 1211
Операции над датами, 1180
Операции с блоками (метод *Перетащить и отпустить*), 1132
Операции с буфером обмена
вставить, 1196
вырезать, 1195
копировать, 1195
Описание синтаксиса для условий выбора наблюдений, 1224
Открыть
графический файл, 1143
другие…, 1142
Командный язык (SCL), 1144
новый файл данных, 1142
программа на языке *STATISTICA* BASIC, 1144
таблица исходных данных, 1143
таблица результатов, 1143
файл, 1144
файл данных, 1143
файл текста/вывода, 1143
Открыть несколько файлов данных, 1143
Отображение разделенных окон, 1136
Очереди, 1226, 1236
Очистить, 1197
Ошибки DDE, 1198

П

Параметры вывода текста на принтер, 1155
Параметры по умолчанию
изменение текстовых значений, 1227
метод *Перетащить и отпустить*, 1228
общие настройки, 1225
печать, 1161
принтер, 1234
экран, 1232
Параметры принтера, 1155
Параметры страницы/вывода – диалоговое окно, 1161
Переключатель модулей, 1217, 1218
одно приложение, 1229
Перекодирование значений, 1186
Переменная скорость
выделение блока, 1136
прокрутка, 1136
Переменная, текущие спецификации, 1170
Переместить наблюдения, 1189
как вставить наблюдения до наблюдения номер один, 1189
Переместить переменные, 1168
как поместить блок до первой переменной, 1168
Пересчитать переменные, 1184
Печать
автоматическая выдача каждой заполненной страницы, 1160
автоматическая печать/выдача страниц при каждом выводе, 1159
автоотчет (автоматическая печать всех таблиц результатов), 1160
графики и таблицы результатов, 1156
данные, 1154
диалоговое окно параметров печати/вывода, 1161
длина вывода, 1158
параметры печати, 1165
параметры по умолчанию системы *STATISTICA*, 1161
параметры принтера, 1155
печать графиков и таблиц результатов, 1156
печать данных, 1154
печать/выдача страниц при выводе, 1155
печать/выдача текущих страниц, 1153
совместная, 1156
тип отчета, 1158
файлы, 1156
Подсказки для кнопок, 1232
Поиск, 1198, 1199
Полутоновые цвета, 1162

StatSoft®
Copyright © StatSoft, 1995

Пользовательские графики (обзор), 1219
Пользовательские графики и статистические графики, 1219
Поля, 1162
  поля вывода текста, 1159, 1164
  поля графика, 1164
Поля вывода текста, 1164
Поля графика, 1162, 1164
Правка
  имена наблюдений, 1191
  текст/числовое присвоение, 1179
  текстовые значения, 1178
  ячейка, 1167
Предметный указатель меню ?(Справка), 1239
Преобразование
  из даты в текстовое значение, 1183
  из текстового значения в дату, 1183
Преобразование в таблицу результатов, 1153
Принтер…, 1165
Присвоение текста/числа, 1178
Продолжить анализ, 1215
Просмотр
  спецификации переменных, 1178

P

Рабочая книга, 1227
Разделение переменной, содержащей дату, 1181, 1182
Разделенная прокрутка в таблицах исходных данных и таблицах результатов, 1136
Разделители списка, 1176
Ранг
  дробный, 1186
  дробный в процентах, 1186
  обычный, 1186
  присвоить ранг 1, 1186
  совпадающие значения, 1186
Ранжирование
  значений, 1185
Распределения
  обзор синтаксиса, 1177
Результаты
  цвет текста, 1233
Ручка заполнения, 1134, 1135, 1136

C

Связи DDE, 1140
Связи данных графика, 1226
Связь DDE, 1138, 1174, 1197, 1198, 1206
Сдвинуть переменные, 1184
Словарь, 1240
Совместная печать, 1156
Совпадающие значения при ранжировании, 1186
Создание даты, 1182
Создание даты из 2 или 3 переменных, 1181
Создание нового текст/числового присвоения, 1179
Создать данные, 1142
Сохранить
  данные под другим именем, 1150
  таблицу результатов под другим именем, 1150
  файл, 1144
Специализированные графики (обзор), 1221
Справка (Электронное руководство), 1239, 1240
Ссылка на систему *STATISTICA*, 1241
Стандартизировать столбцы блока, 1202
Стандартизировать строки блока, 1202
Стартовая панель, 1215
Статистики блока
  по столбцам, 1202
  по строкам, 1203
Статистические графики (обзор), 1219
Статистический советник, 1240
Строка заголовка, 1193
Строка состояния, 1213, 1232

T

Таблица исходных данных
  автоматический пересчет переменных, 1184
  выпадающее меню ?(Справка), 1239
  выпадающее меню *Анализ*, 1215
  выпадающее меню *Вид*, 1208
  выпадающее меню *Графика*, 1219
  выпадающее меню Макрокоманды, 1234
  выпадающее меню *Окно*, 1236
  выпадающее меню *Правка*, 1167
  выпадающее меню *Сервис*, 1223
  выпадающее меню *Файл*, 1142
  добавить наблюдения, 1188
  добавить переменные, 1167
  заголовки, 1212
  заголовок файла, примечания и Рабочая книга, 1192
  заполнить случайными значениями, 1201
  заполнить/копировать вниз, 1201
  заполнить/копировать вправо, 1201
  изменение (текущих) спецификаций, 1170
  изменение имен наблюдений, 1191
  изменение текстовых значений, 1178
  изменение ячейки, 1167
  индикатор состояния, 1214
  копировать вниз, 1201
  копировать вправо, 1201
  копировать наблюдения, 1190
  копировать переменные, 1169
  метод *Перетащить и отпустить*, 1132
  окно, 1142
  панель инструментов, 1212
  параметры имен наблюдений, 1192
  перекодировать значения, 1186
  переместить наблюдения, 1189
  переместить переменные, 1168
  пересчитать, 1184
  просмотр имен наблюдений, 1208
  просмотр меток значений, 1208
  просмотр полных имен переменных, 1208
  просмотр текстовых значений, 1208
  разделение прокрутки, 1136
  ранжирование значений, 1185
  связи, 1206
  сдвинуть переменные, 1184
  спецификации всех переменных, 1178

Copyright © StatSoft, 1995

стандартизировать столбцы блока, 1202
стандартизировать строки блока, 1202
строка состояния, 1213
транспонировать блок, 1200
транспонировать файл данных, 1200
удалить наблюдения, 1191
удалить переменные, 1170
формулы, 1176
формулы распределений, 1177
цвета, 1209
шрифт, 1208
Таблица результатов
диспетчер таблиц результатов, 1236
метод *Перетащить и отпустить*, 1132
разделение прокрутки, 1136
создание из таблицы исходных данных, 1153
цвета, 1209
шрифт, 1208
Текст
даты/значения, 1183
значения, 1208
метки значений, 1208
Текстовые значения, 1178
Текущие спецификации, 1170
Тип отчета в распечатках, 1158
Транспонировать
блок, 1200
файл данных, 1200

У

Удалить наблюдения, 1191
Удалить переменные, 1170
Упорядочить
каскадом, 1237
сверху вниз, 1237
слева направо, 1237
Упорядочить значки, 1238
Установить связь, 1138

Ф

Файл данных
заголовок, примечания и Рабочая книга, 1192
Файлы Рабочей книги, 1192, 1193
если Рабочую книгу нужно отключить, 1195
Формулы
имена переменных, 1176
комментарии, 1176
константы, 1177
логические операторы, 1176
операторы, 1177
операторы условий, 1176
разделители списка, 1176
распределения, 1177
функции, 1177
Формулы в таблице исходных данных
автоматический пересчет переменных, 1184

Ц

Цвета, 1209, 1233
Цветовая палитра
базовая, 1210
дополнительная, 1211

Ш

Шрифт, 1208

Э

Экспорт, 1151
разделители в свободном ASCII-формате, 1152
файлы Excel, 1151, 1152
файлы Lotus, 1151, 1152
файлы Quattro, 1151, 1152
файлы в свободном ASCII-формате, 1152
Экстраполяция блока, 1135
Электронное руководство, 1239, 1240

Я

Язык *STATISTICA* BASIC, 1216

StatSoft®

Copyright © StatSoft, 1995

Глава 4:

ОКНО ТАБЛИЦЫ РЕЗУЛЬТАТОВ: МЕНЮ, ДИАЛОГОВЫЕ ОКНА, СВОЙСТВА

Оглавление

Операции с блоками и *Автозаполнение* в таблицах результатов 1248
Выпадающее меню *Файл* таблицы результатов 1252
Выпадающее меню *Правка* таблицы результатов 1258
Выпадающее меню *Вид* таблицы результатов 1262
Выпадающее меню *Анализ* таблицы результатов 1267
Выпадающее меню *Графики* таблицы результатов 1269
Выпадающее меню *Сервис* таблицы результатов 1269
Выпадающее меню *Окно* таблицы результатов 1269
Выпадающее меню *?(Справка)* таблицы результатов 1269
Предметный указатель 1270

Подробное оглавление

ОПЕРАЦИИ С БЛОКАМИ И *АВТОЗАПОЛНЕНИЕ* В ТАБЛИЦАХ РЕЗУЛЬТАТОВ 1248
Возможности метода *Перетащить и отпустить* 1248
Выделение блока 1248
Перемещение блока 1248
Копирование блока 1249
Очистка блока 1249
Экстраполяция блока (*Автозаполнение*) 1249
Изменение скорости выделения блока 1250

РАЗДЕЛЕНИЕ ПРОКРУТКИ В ТАБЛИЦАХ ИСХОДНЫХ ДАННЫХ И ТАБЛИЦАХ РЕЗУЛЬТАТОВ 1250

ВЫПАДАЮЩЕЕ МЕНЮ *ФАЙЛ* ТАБЛИЦЫ РЕЗУЛЬТАТОВ 1252
Открыть таблицу результатов 1252
Открыть другие 1252
Сохранить таблицу результатов 1252
Сохранить как… 1253
Сохранить как файл данных 1253
Параметры печати в выпадающем меню *Файл* таблицы результатов 1253
Печать/выдача текущих страниц 1253
Печать таблицы результатов 1253

Copyright © StatSoft, 1995

Диапазон ... 1254
Маркированные ячейки ... 1254
Печать после выхода из окна ... 1254
Параметры ... 1255
Параметры вывода текста на принтер ... 1255
Печать графических файлов и таблиц результатов ... 1255
Параметры страницы/вывода ... 1256
Принтер ... 1256
Выход ... 1257

ВЫПАДАЮЩЕЕ МЕНЮ *ПРАВКА* ТАБЛИЦЫ РЕЗУЛЬТАТОВ ... 1258
Ячейка таблицы результатов (изменение) ... 1258
Спецификации столбца таблицы результатов (изменение) ... 1258
Первая строка ... 1258
Вторая строка ... 1258
Формат ... 1258
Ширина столбцов ... 1259
Ширина полей ... 1259
Имя строки таблицы результатов (изменение) ... 1259
Имя ... 1259
Ширина ... 1259
Заголовки таблицы результатов (изменение) ... 1259
Заголовок окна ... 1260
Первая строка ... 1260
Вторая строка ... 1260
Выравнивание ... 1260
Информационное поле таблицы результатов (изменение) ... 1260
Строка 1 ... 1260
Строка 2 ... 1260
Строка 3 ... 1260
Строка 4 ... 1260
Добавить столбцы/строки в таблицу результатов ... 1260
Операции с буфером обмена в таблице результатов ... 1261
Заполнить/стандартизировать блок ... 1261
Статистики блока/по строкам и столбцам ... 1261
Захват экрана ... 1261

ВЫПАДАЮЩЕЕ МЕНЮ *ВИД* ТАБЛИЦЫ РЕЗУЛЬТАТОВ ... 1262
Ширина таблицы результатов ... 1262
Ширина столбца ... 1262
Длина имени строки ... 1262
Ширина полей ... 1262
Шрифт таблицы результатов ... 1262
Увеличить таблицу результатов ... 1263
Уменьшить таблицу результатов ... 1263

StatSoft®
Copyright © StatSoft, 1995

Диалоговое окно Параметры увеличения ... 1263
OK (по умолчанию) ... 1263
Выбрать шрифт ... 1264
Цвета таблицы результатов ... 1264
Маркировать выделенное ... 1264
Кнопка Параметры панели инструментов таблицы результатов ... 1264
Окно текста/вывода ... 1264
Заголовки таблицы результатов ... 1265
Панель инструментов таблицы результатов ... 1265
Изменить поле просмотра ... 1265
Строка состояния таблицы результатов ... 1265
Комментарий ... 1266
Вывод ... 1266
Выбор ... 1266
Вес ... 1266
ЗАП ... 1266
Кнопки автозадач таблицы результатов ... 1266

ВЫПАДАЮЩЕЕ МЕНЮ *АНАЛИЗ* ТАБЛИЦЫ РЕЗУЛЬТАТОВ ... 1267
Стартовая панель ... 1267
Продолжить анализ ... 1267
Две кнопки *Далее* в таблице результатов ... 1267
Кнопка *Далее* и “всплывающая” кнопка *Далее* в таблице результатов ... 1267
Язык STATISTICA BASIC ... 1268
Командный язык ... 1268
Быстрые основные статистики ... 1268
Другие виды анализа (Переключатель модулей системы *STATISTICA*) ... 1268

ВЫПАДАЮЩЕЕ МЕНЮ *ГРАФИКА* ТАБЛИЦЫ РЕЗУЛЬТАТОВ ... 1269

ВЫПАДАЮЩЕЕ МЕНЮ *СЕРВИС* ТАБЛИЦЫ РЕЗУЛЬТАТОВ ... 1269

ВЫПАДАЮЩЕЕ МЕНЮ *ОКНО* ТАБЛИЦЫ РЕЗУЛЬТАТОВ ... 1269

ВЫПАДАЮЩЕЕ МЕНЮ *?(СПРАВКА)* ТАБЛИЦЫ РЕЗУЛЬТАТОВ ... 1269

ПРЕДМЕТНЫЙ УКАЗАТЕЛЬ ... 1270

Copyright © StatSoft, 1995

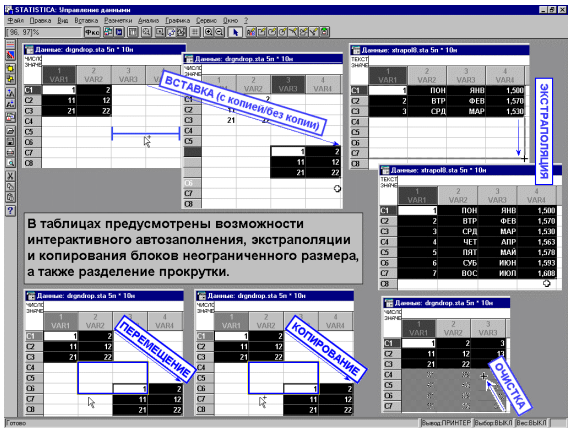
Глава 4:

ОКНО ТАБЛИЦЫ РЕЗУЛЬТАТОВ: МЕНЮ, ДИАЛОГОВЫЕ ОКНА, СВОЙСТВА

ОПЕРАЦИИ С БЛОКАМИ И *АВТОЗАПОЛНЕНИЕ* В ТАБЛИЦАХ РЕЗУЛЬТАТОВ

Возможности метода *Перетащить и отпустить*

Система *STATISTICA* поддерживает полный набор описанных ниже стандартных возможностей метода *Перетащить и отпустить* (стиль Excel).



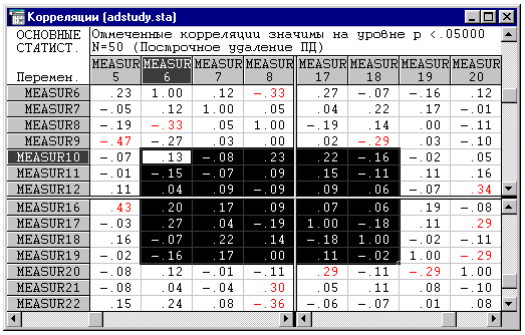
Таблицы результатов поддерживают те же возможности метода *Перетащить и отпустить*, что и таблицы исходных данных за исключением вставки. Обратите внимание, что операции метода *Перетащить и отпустить* могут быть отключены, если выбрать опцию *Отключить все функции метода Перетащить и отпустить* в диалоговом окне *Параметры по умолчанию: общие настройки* (см. главу 3).

Выделение блока

Блоки можно выделить следующими способами: (1) провести курсором, удерживая при этом нажатой левую кнопку мыши, (2) щелкнуть на одном углу выделяемого блока, а затем, прокрутив таблицу до противоположного угла (первая ячейка останется выделенной), щелкнуть на последней ячейке, одновременно нажав клавишу SHIFT (щелкнуть с клавишей SHIFT), или (3) использовать клавиши перемещения курсора, держа нажатой клавишу SHIFT.

Для расширения ранее выделенного блока можно воспользоваться клавишей SHIFT и клавишами перемещения курсора или прокрутить таблицу с помощью мыши и нажать клавишу SHIFT, щелкнув при этом в нужном углу блока.

Чтобы выделить большой блок в режиме “разделения окон” (как показано ниже), надо щелкнуть на ячейке в одном окне, затем прокрутить таблицу до противоположной ячейки в другом окне и щелкнуть с клавишей SHIFT, выделяя блок.


Корреляции (adstudy.sta)

ОСНОВНЫЕ СТАТИСТ. — Отмеченные корреляции значимы на уровне p <.05000 N=50 (Построчное удаление ПД)

Перемен.	MEASUR 5	MEASUR 6	MEASUR 7	MEASUR 8	MEASUR 17	MEASUR 18	MEASUR 19	MEASUR 20
MEASUR6	-.23	1.00	.12	-.33	.27	-.07	-.16	.12
MEASUR7	-.05	.12	1.00	.05	.04	.22	.17	-.01
MEASUR8	-.19	-.33	.05	1.00	-.19	.14	.00	-.11
MEASUR9	-.47	-.27	.03	.00	.02	-.29	.03	-.10
MEASUR10	-.07	.13	-.08	.23	.22	-.16	-.02	.05
MEASUR11	-.01	-.15	-.07	.09	.15	-.11	.11	.16
MEASUR12	.11	.04	.09	-.09	.09	.06	-.07	.34
MEASUR16	.43	.20	.17	.09	.07	.06	.19	-.08
MEASUR17	-.03	.27	.04	-.19	1.00	-.18	.11	.29
MEASUR18	.16	-.07	.22	.14	-.18	1.00	-.02	-.11
MEASUR19	-.02	-.16	.17	.00	.11	-.02	1.00	-.29
MEASUR20	-.08	.12	-.01	-.11	.29	-.11	-.29	1.00
MEASUR21	-.08	.04	-.04	.30	.05	.11	.08	-.10
MEASUR22	.15	.24	.08	-.36	-.06	-.07	.01	.08

Когда в таблице исходных данных или таблице результатов выделен блок, над ним можно производить описанные ниже операции с помощью метода *Перетащить и отпустить*.

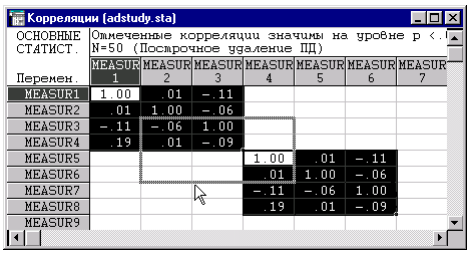
Перемещение блока

Выделенный блок данных можно переместить, подведя курсор мыши к границе выделения

StatSoft®

Copyright © StatSoft, 1995

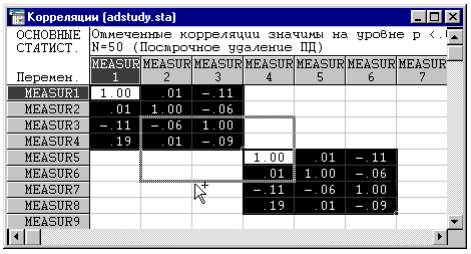
(курсор измениться на стрелку) и перетаскивая блок на новое место.



После перемещения блока данных в пустые ячейки будут помещены пропущенные значения. См. также *Перемещение и копирование значений в таблице результатов* ниже.

Копирование блока

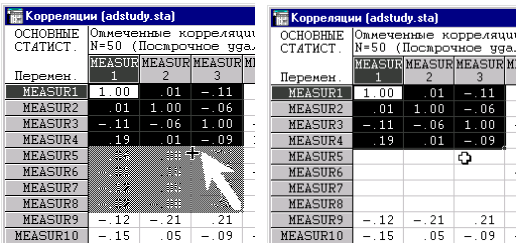
Выделенный блок данных можно скопировать, подведя курсор мыши к границе выделения (курсор изменится на стрелку) и перетаскивая блок на новое место, нажав клавишу CTRL, при этом рядом с курсором появится знак “плюс”.



Можно изменить действие, происходящее по умолчанию при перемещении блока без нажатия клавиши CTRL (с перемещения на копирование). Этот параметр меняется в диалоговом окне *Параметры по умолчанию: общие настройки* (см. главу 3).

Очистка блока

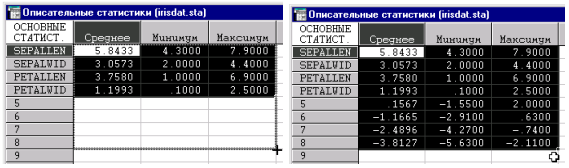
Часть выделенного блока (или весь блок) может быть очищена перемещением в его собственных пределах.



После выделения блока надо переместить курсор в нижний правый угол блока, к *Ручке заполнения*. Когда курсор изменится на знак “плюс, « надо нажать левую кнопку мыши и переместить его в пределах выделенного блока. Во время перетаскивания значения в блоке будут затенены. Когда кнопка мыши будет отпущена, затененные значения будут удалены.

Экстраполяция блока (*Автозаполнение*)

Горизонтальные или вертикальные ряды блока могут быть экстраполированы перетаскиванием *Ручки заполнения* (маленького квадрата в правом нижнем углу блока). Обратите внимание, что если строка или столбец выделены полностью, то *Ручка заполнения* отображается как в верхней, так и в нижней части столбца, или в начале и в конце строки. В системе *STATISTICA* существует возможность создания серий значений, таких как последовательные числа или линейные экстраполяции (например, серию *1*, *2*, *3* можно расширить так, чтобы она включала *4*, *5*, *6*, ...).



Экстраполяция блока (*Автозаполнение*) осуществляется одним из перечисленных ниже способов.

Copyright © StatSoft, 1995

StatSoft®

- Если исходное выделение содержит повторяющиеся значения, то они дублируются в расширенном блоке.
- Для неидентичных значений вычисляется экстраполяция с помощью линейной регрессии.

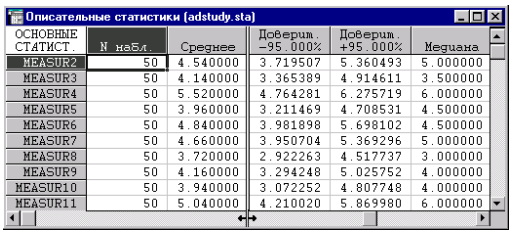
Экстраполяция (*Автозаполнение*) блока вверх или влево. Помимо экстраполяции вниз или вправо, блок можно экстраполировать вверх или влево, перетаскивая *Ручку заполнения* вверх или влево от начала блока. Обратите внимание, что если переместить *Ручку заполнения* вверх или влево в пределах начального выделения, не выходя за верхнюю или левую границы блока, то данные в пределах выделения будут удалены (во время перетаскивания эти значения отображаются серым цветом, см. стр. 1249). Для копирования первой строки или первого столбца ячеек блока в смежные ячейки выделенного блока можно воспользоваться командами *Заполнить блок* контекстного меню или выпадающего меню *Правка*.

Изменение скорости выделения блока

Существует возможность изменения скорости прокрутки при расширении блока за пределы текущего окна. Переместив курсор на небольшое расстояние от таблицы результатов, можно прокручивать выделенный блок по одной строке. Для постраничной прокрутки нужно переместить курсор на большее расстояние от таблицы результатов.

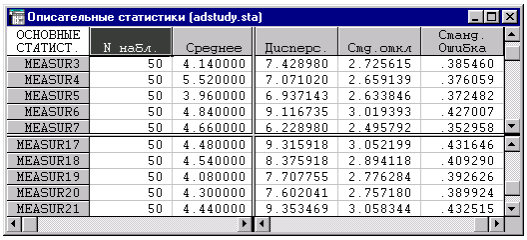
РАЗДЕЛЕНИЕ ПРОКРУТКИ В ТАБЛИЦАХ ИСХОДНЫХ ДАННЫХ И ТАБЛИЦАХ РЕЗУЛЬТАТОВ

Таблицы исходных данных могут быть разделены на две или четыре секции (окна) перемещением полосы разделения (черного прямоугольника в верхней части вертикальной полосы прокрутки или в левой части горизонтальной). Это может оказаться полезным для одновременного просмотра результатов из разных частей таблицы исходных данных или таблицы результатов в том случае, когда объем данных очень большой.



ОСНОВНЫЕ СТАТИСТ.	N набл.	Среднее	Доверит. -95.000%	Доверит. +95.000%	Медиана
MEASUR2	50	4.540000	3.719507	5.360493	5.000000
MEASUR3	50	4.140000	3.365389	4.914611	3.500000
MEASUR4	50	5.520000	4.764281	6.275719	6.000000
MEASUR5	50	3.960000	3.211469	4.708531	4.500000
MEASUR6	50	4.840000	3.981898	5.698102	4.500000
MEASUR7	50	4.660000	3.950704	5.369296	5.000000
MEASUR8	50	3.720000	2.922263	4.517737	3.000000
MEASUR9	50	4.160000	3.294248	5.025752	4.000000
MEASUR10	50	3.940000	3.072252	4.807748	4.000000
MEASUR11	50	5.040000	4.210020	5.869980	6.000000

Для того чтобы разделить окно таблицы исходных данных или таблицы результатов по вертикали или по горизонтали, нужно переместить курсор мыши к вертикальной или горизонтальной “полосе разделения” (черному прямоугольнику над верхней стрелкой вертикальной прокрутки или слева от левой стрелки горизонтальной прокрутки) и перетащить “курсор разделения” (↔ или ↕) к нужной строке или нужному столбцу. Таблицы исходных данных или таблицы результатов могут быть разделены на два (см. выше) или четыре (см. ниже) “окна”.



ОСНОВНЫЕ СТАТИСТ.	N набл.	Среднее	Дисперс.	Смд.откл	Сманд. Ошибка
MEASUR3	50	4.140000	7.428980	2.725615	.385460
MEASUR4	50	5.520000	7.071020	2.659139	.376059
MEASUR5	50	3.960000	6.937143	2.633846	.372482
MEASUR6	50	4.840000	9.116735	3.019393	.427007
MEASUR7	50	4.660000	6.228980	2.495792	.352958
MEASUR17	50	4.480000	9.315918	3.052199	.431646
MEASUR18	50	4.540000	8.375918	2.894118	.409290
MEASUR19	50	4.080000	7.707755	2.776284	.392626
MEASUR20	50	4.300000	7.602041	2.757180	.389924
MEASUR21	50	4.440000	9.353469	3.058344	.432515

Обратите внимание, что окна, разделенные по вертикали, прокручиваются вместе при вертикальном прокручивании, а окна, разделенные по горизонтали прокручиваются вместе при горизонтальном прокручивании.

Чтобы убрать разделение окон из таблицы исходных данных или таблицы результатов,

StatSoft®

Copyright © StatSoft, 1995

нужно перетащить полосу разделения к прокручивающим стрелкам на любом конце полосы прокрутки.

Чтобы выделить большой блок в режиме “разделения окон”, нужно щелкнуть на ячейке в одном окне, затем прокрутить таблицу до противоположной ячейки в другом окне и щелкнуть с клавишей SHIFT (см. ниже *Выделение блока*). Далее к выделенному блоку могут быть применены операции с блоками системы *STATISTICA* (т.е., метод *Перетащить и отпустить*, см. выше, и *Блоковые статистики-/Столбцы /Строки /Графики*, см. стр. 1261).

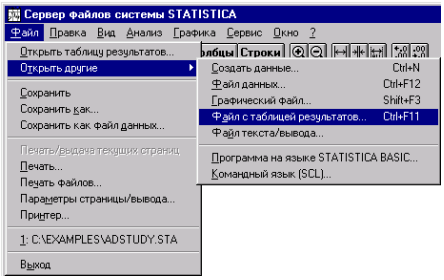
Отображение разделенных окон также возможно в *Редакторе данных графика*, *Менеджере мегафайлов* и других окнах таблиц результатов.

Copyright © StatSoft, 1995

StatSoft®

ВЫПАДАЮЩЕЕ МЕНЮ *ФАЙЛ* ТАБЛИЦЫ РЕЗУЛЬТАТОВ

Из выпадающего меню *Файл* таблицы результатов доступны следующие диалоговые окна и опции.



Ниже приведено их подробное описание.

Открыть таблицу результатов

Эта опция открывает диалоговое окно *Открытие файла с таблицей результатов,* в котором можно выбрать файл открываемой таблицы (с расширением **.scr*). Файлы таблиц результатов содержат полученные данные, а также информацию о форматировании и настройках (например, выделение, пользовательские цвета и т.п.).

Кнопка панели инструментов. См. кнопку панели инструментов *Окна таблицы результатов*.

Комбинация клавиш. CTRL+F11

Открыть другие

Список этого пункта меню является общим для всех меню (таблиц исходных данных, таблиц результатов, графиков, *Окна текста/вывода*) и предоставляет доступ ко всем типам файлов системы *STATISTICA*. Эти опции позволяют *Создать данные*, открыть существующий *Файл данных*, *Графический файл*, *Файл таблицы результатов*, *Файл текста/вывода*, *Программу на языке STATISTICA BASIC* или программу на *Командном языке (SCL*) (подробную информацию о последних двух опциях см. в *Электронном руководстве*). Если выбрать опцию *Создать данные*, откроется диалоговое окно *Создание данных: имя файла*, в котором можно задать имя файла для сохранения новой таблицы. После выбора имени файла будет создана таблица размером 10 переменных на 10 наблюдений.

Выбор любой из остальных опций откроет соответствующее диалоговое окно *Открытие/Сохранение файла,* в котором можно выбрать имя открываемого файла.

Сохранить таблицу результатов

С помощью этой опции можно сохранить таблицу результатов (файл с расширением **.scr*). Если текущая таблица результатов ранее не сохранялась (т.е. имя файла не задано), то откроется диалоговое окно *Сохранение таблицы результатов,* в котором можно указать имя сохраняемого файла и место на диске, где он будет расположен. Если имя и место на диске указаны, то при каждом выборе опции *Сохранить* таблица результатов будет сохраняться под одним и тем же именем файла, удаляя его предыдущие копии. Если необходимо изменить имя или место расположения файла на диске, нужно выбрать опцию *Сохранить как*.

Кнопка панели инструментов. См. кнопку панели инструментов *Окна таблицы результатов*.

Комбинация клавиш. SHIFT+F12

StatSoft®

Copyright © StatSoft, 1995

Сохранить как…

Эта команда также используется для сохранения таблицы результатов. При каждом выборе этой опции открывается диалоговое окно *Сохранение таблицы результатов*, в котором можно указать имя файла и место на диске, куда будет записана текущая таблица.

Комбинация клавиш. F12

Сохранить как файл данных

В системе *STATISTICA* различные типы результатов анализа (таблицы результатов и графики) могут быть преобразованы в исходные данные. Данная команда открывает диалоговое окно *Сохранение таблицы результатов*, в котором можно сохранить содержание таблицы результатов в формате таблицы исходных данных (файла данных с расширением **.sta*). Это позволяет в дальнейшем использовать содержание таблицы результатов в качестве исходных данных для анализа, или экспортировать его в формат других файлов. Когда система *STATISTICA* создает таблицу исходных данных из таблицы результатов, она использует столбец *Имена строк* таблицы результатов в качестве *Имен наблюдений* и присваивает *Заголовки столбцов* таблицы результатов *Именам переменных* (первая строка имени столбца) и *Длинным меткам переменных* (вторая строка имени столбца) таблицы исходных данных. Система *STATISTICA* также использует *Заголовок* таблицы результатов в качестве *Заголовка файла* таблицы исходных данных.

Параметры печати в выпадающем меню *Файл* таблицы результатов

Описанные ниже опции доступны из выпадающего меню *Файл* как для таблицы исходных данных, так и для таблицы результатов. Более подробную информацию о них см. в главе 3.

Печать/выдача текущих страниц

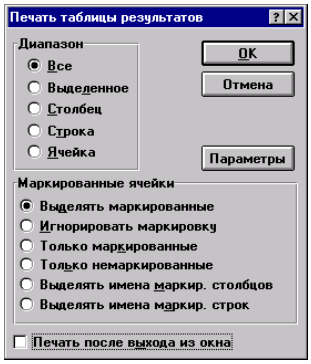
Если в диалоговом окне *Параметры страницы/вывода* (см. стр. 1256) установлен *Вывод* на *Принтер* или в *Файл*, то процедура вывода будет зависеть от выбора опции *Авт. выдача каждой заполненной страницы* в этом же диалоговом окне. При выборе этой опции система *STATISTICA* будет сохранять вывод во внутреннем буфере до тех пор, пока не заполнится страница. Если эта опция не выбрана, вывод будет храниться во внутреннем буфере до окончания анализа. Команда *Печать/выдача текущих страниц* предотвращает внутреннее сохранение и сразу выводит на печать (или посылает в файл) накопившийся вывод (технически, происходит “выдача страницы ”).

Обратите внимание, что эта опция не действует в *Окне текста/вывода* (см. главу 5).

Печать таблицы результатов

С помощью команды *Печать* выпадающего меню *Файл* таблицы результатов можно вывести активную таблицу результатов в файл (для дальнейшей печати, исправления или включения в другие документы), на принтер (для непосредственной печати) и/или в *Окно текста/вывода* согласно установкам, сделанным в диалоговом окне *Параметры страницы/вывода* (стр. 1256). Если выбрать эту опцию, откроется диалоговое окно *Печать таблицы результатов*.

Copyright © StatSoft, 1995



Быстрый вызов. Чтобы распечатать таблицу результатов, не открывая диалоговое окно *Печать таблицы результатов* (т.е. используя ранее сохраненные параметры или параметры по умолчанию), нужно нажать на кнопку *Печать* панели инструментов (см. ниже). Обратите внимание, что если в таблице результатов выделен блок, то будет напечатан только этот блок; в противном случае будет напечатана вся таблица.

Кнопка панели инструментов. См. кнопку панели инструментов *Окна таблицы результатов Window*.

Комбинация клавиш. CTRL+P

Диапазон

Здесь можно выбрать диапазон печати таблицы результатов.

Все. Печатать всю активную таблицу результатов.

Выделенное. Печатать только выделенный блок ячеек активной таблицы результатов.

Столбец. Печатать только столбец, на котором расположен курсор.

Строка. Печатать только строку, на которой расположен курсор.

Ячейка. Печатать только ячейку, на которой расположен курсор.

Маркированные ячейки

Следующие пункты имеют отношение только к печати маркированных ячеек таблицы результатов. Маркированными называются ячейки, содержание которых помечено с помощью опции *Маркировать выделенное* (стр. 1264), или ячейки, которые автоматически маркированы системой *STATISTICA* во время анализа (например, в случае корреляционного анализа выделены значимые корреляции).

Выделять маркированные. Эта опция позволит распечатать весь заданный диапазон таблицы результатов (см. выше), выделяя маркированные ячейки с помощью расположенной справа звездочки (*).

Игнорировать маркировку. Эта опция позволит печатать весь выделенный диапазон таблицы результатов (см. выше), но маркированные ячейки выделяться не будут.

Только маркированные. Эта опция позволит напечатать только маркированные ячейки таблицы результатов. При этом будут напечатаны заголовки строк и столбцов, а немаркированные ячейки останутся пустыми.

Выделять имена маркир. столбцов. Если в выделенном диапазоне таблицы результатов есть маркированные ячейки (см. выше), то при выводе будут выделены только *Имена столбцов,* в которых находятся такие ячейки.

Выделять имена маркир. строк. Если в выделенном диапазоне таблицы результатов есть маркированные ячейки (см. выше), то при выводе будут выделены только *Имена строк,* в которых находятся такие ячейки.

Печать после выхода из окна

Если выбрана эта опция, распечатка автоматически будет производиться в файл или на принтер при нажатии кнопки *OK* в этом диалоговом окне или при нажатии кнопки

StatSoft®

Copyright © StatSoft, 1995

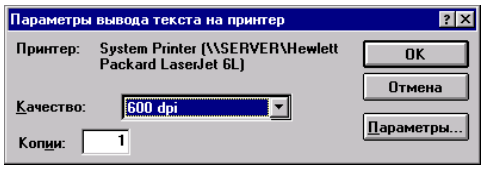
Печать на панели инструментов таблицы результатов (текст, оставшийся во внутреннем буфере, также будет распечатан). При этом каждая распечатка (таблица) будет начинаться с новой страницы. Если эта опция не выбрана, вывод будет сохраняться во внутреннем буфере и посылаться на принтер в зависимости от состояния поля *Авт. выдача каждой заполненной страницы* (либо каждая заполненная страница, либо все сразу, например, при выходе из модуля или очистке буфера с помощью опции *Печать/выдача текущих страниц*, см. стр. 1253). Обратите внимание, что эта опция применяется только локально (временно) в текущем модуле системы *STATISTICA* на текущем сеансе, тогда как выбор опции *Авт. печать/выдача страниц при каждом выводе* в диалоговом окне *Параметры страницы/вывода* (см. стр. 1256) действует глобально (постоянно).

Параметры

Эта кнопка открывает диалоговое окно *Параметры вывода текста на принтер* (см. ниже), в котором можно выбрать качество (т.е. разрешение) распечатки и число копий.

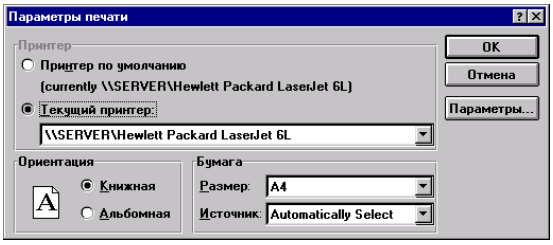
Параметры вывода текста на принтер

Это диалоговое окно вызывается из диалогового окна *Печать таблицы результатов* (см. выше) кнопкой *Параметры*. Здесь можно выбирать качество печати (т.е. разрешение) и число копий для вывода текста.



В этом диалоговом окне также отображается тип выбранного принтера. Он может быть изменен в диалоговом окне *Параметры печати* (см. стр. 1256), которое вызывается с помощью кнопки *Параметры*.

В диалоговом окне *Параметры печати* можно выбрать принтер (из установленных на данном компьютере), ориентацию страницы, размер и способ подачи бумаги.

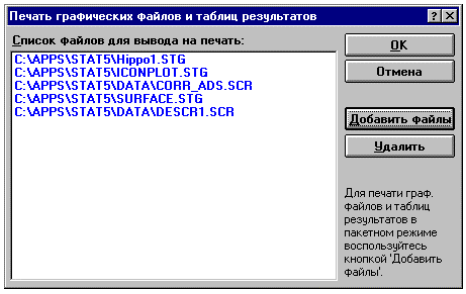


Обратите внимание, что границы вывода текста могут быть изменены в диалоговом окне *Поля вывода текста* (оно вызывается кнопкой *Поля/Принтер* в диалоговом окне *Параметры страницы/вывода*, см. ниже).

Печать графических файлов и таблиц результатов

Команда *Печать файлов* из выпадающего меню *Файл* позволяет распечатывать сохраненные ранее графики и таблицы результатов в пакетном режиме. Эта опция может оказаться полезной, если необходимо распечатывать графики и таблицы результатов не в процессе анализа, а позднее. Для этого графики или таблицы результатов сохраняются по мере их получения (с помощью опций *Сохранить таблицу результатов* или *Сохранить график*, см. стр. 1252), а затем вызывается показанное ниже диалоговое окно.

Copyright © StatSoft, 1995



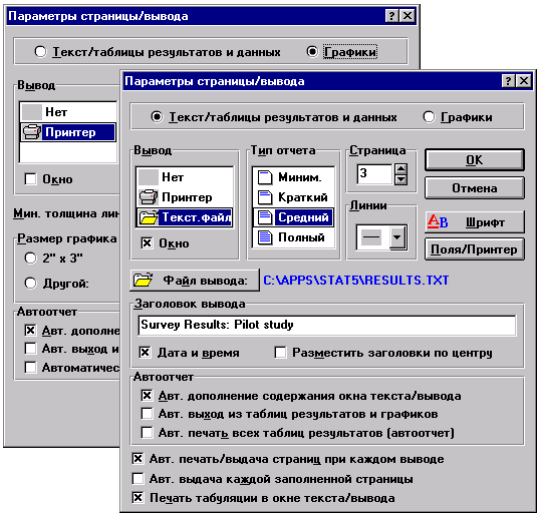
В этом окне можно выбрать нужные графики и таблицы результатов (см. ниже) и нажать кнопку *OK* для начала совместной печати.

После нажатия *OK* сначала будут напечатаны все выбранные графики (на отдельных страницах, в том порядке, в котором они были выделены), а затем все таблицы результатов вместе (также в выбранном порядке).

Подробную информацию об этой опции см. в главе 3.

Параметры страницы/вывода

Это диалоговое окно с закладками позволяет указывать, куда будет производиться вывод, задавать объем дополнительной информации, размеры границ и другие параметры. Оно вызывается с помощью двойного щелчка на поле *Вывод* строки состояния, через опцию *Параметры страницы/вывода* выпадающего меню *Файл* или через опцию *Принтер* выпадающего меню *Сервис*. Все выбранные здесь параметры будут использоваться по умолчанию (глобально) при каждом запуске любого из модулей системы *STATISTICA*.



Перед тем как печатать текст, таблицу исходных данных или таблицу результатов, нужно выбрать один из вариантов вывода: на принтер, в файл и/или в *Окно текста/вывода* (на закладке *Текст-/Таблицы результатов/Таблицы исходных данных*) или на принтер и/или в *Окно текста/вывода* (на закладке *Графики*), а также установить другие опции соответствующей закладки этого диалогового окна.

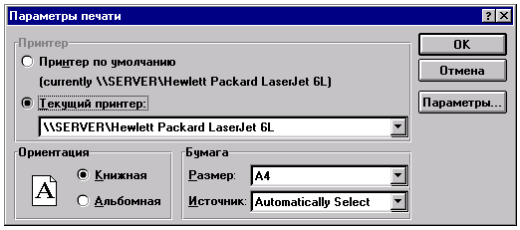
Подробную информацию об этом окне см. в главе 3.

Принтер

Эта опция выпадающего меню *Файл* открывает диалоговое окно *Параметры печати,* в котором можно выбрать принтер, ориентацию страницы, а также размер и способ подачи бумаги для печати графического и текстового вывода.

StatSoft®

Copyright © StatSoft, 1995



Обратите внимание, что можно сохранить различные параметры печати для графического и текстового вывода (т.е. вывода таблиц результатов и таблиц исходных данных).

Подробную информацию об этом окне см. в главе 3.

Выход

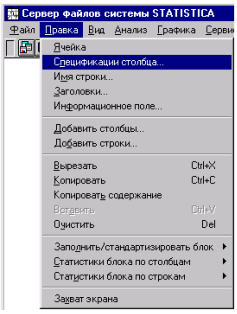
Эта команда завершает текущий сеанс системы *STATISTICA*. Если выбрать опцию *Выход* до того, как будут сохранены изменения в таблицах исходных данных, таблицах результатов, тексте или графиках, система *STATISTICA* выведет запрос о сохранении измененных объектов.

Комбинация клавиш. ALT+F4

Copyright © StatSoft, 1995

ВЫПАДАЮЩЕЕ МЕНЮ *ПРАВКА* ТАБЛИЦЫ РЕЗУЛЬТАТОВ

Из выпадающего меню *Правка* таблицы результатов доступны следующие диалоговые окна и опции.



Далее подробно рассмотрены показанные выше команды.

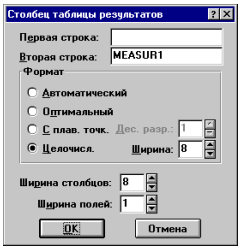
Ячейка таблицы результатов (изменение)

Опция *Ячейка* позволяет изменять выделенную ячейку таблицы результатов, не удаляя ее содержание. После выбора эта команда будет затенена до окончания изменения ячейки.

Комбинация клавиш: F2 или двойной щелчок на ячейке.

Спецификации столбца таблицы результатов (изменение)

Выбрав опцию *Спецификации столбца* выпадающего меню *Правка* (или дважды щелкнув на *Имени столбца*), можно изменить параметры выделенного столбца в диалоговом окне *Столбец таблицы результатов*.



Спецификации каждого столбца изменяются отдельно от остальных [за исключением параметров *Ширина столбцов* и *Ширина полей* (см. ниже), которые применяются ко всем столбцам таблицы результатов].

Комбинация клавиш. CTRL+F2

Первая строка

Здесь можно ввести первую строку заголовка столбца, которая может содержать любые печатаемые символы. Длина этой строки ограничена текущей шириной столбца (см. ниже опцию *Ширина столбцов*). Если вторая строка (см. ниже) пуста, а длина текста первой строки больше текущей ширины столбца, то при отображении в таблице текст будет перенесен на вторую строку, хотя изменять его по-прежнему нужно будет в этом поле.

Вторая строка

Здесь можно ввести вторую строку заголовка столбца. Длина строки также ограничена шириной столбца, а заголовок может содержать любые печатаемые символы.

Формат

Формат отображения значений в столбцах таблицы результатов можно изменить с помощью описанных ниже опций.

Автоматический. Если выбрана эта опция, все числа в столбце будут автоматически выровнены в соответствии с наибольшим числом (положительным или отрицательным), так что в

StatSoft®
Copyright © StatSoft, 1995

каждой ячейке столбца десятичная точка будет находиться на одной и той же позиции.

Оптимальный. Если выбрана эта опция, значения в столбце не будут выравниваться по десятичным разрядам, однако каждая ячейка по отдельности будет изменена, чтобы выглядеть "оптимально". Значения в столбце будут заполнять ширину столбца в зависимости от значений параметров *Ширина столбцов* и *Ширина полей*.

С плав. точк. С помощью этой опции можно указать число десятичных разрядов для значений данного столбца. Минимальное число десятичных разрядов - 1, максимальное - 15.

Целочисл. Если выбрана эта опция, все десятичные числа будут округлены до целых. Для отображения целых чисел можно указать их длину (минимальная = 1, максимальная = текущей ширине столбца).

Ширина столбцов

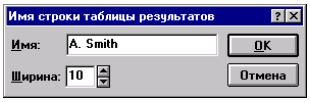
Указанное здесь значение (минимальное = 1, максимальное = 80) определяет ширину каждого столбца таблицы результатов. (Это глобальная операция, которая применяется ко всем столбцам таблицы результатов.) Данная опция также доступна из диалогового окна *Настройка ширины* (стр. 1262).

Ширина полей

Значения в каждом столбце могут иметь заданные поля с обеих сторон для того, чтобы отделить значения разных столбцов друг от друга. Эта опция позволяет указать ширину полей (минимальная = 0, максимальная = 10) для каждого столбца таблицы результатов. (Это глобальная операция, которая применяется ко всем столбцам таблицы результатов.) Данная опция также доступна из диалогового окна *Настройка ширины* (стр. 1262).

Имя строки таблицы результатов (изменение)

При выборе этой опции (или двойном щелчке на *Имени строки*), откроется диалоговое окно *Имя строки таблицы результатов*, где можно изменить *Имя строки,* выделенной в таблице результатов.



Комбинация клавиш. ALT+F2

Имя

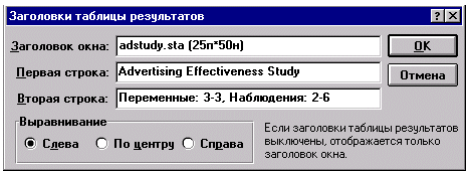
Здесь можно ввести или изменить *Имя строки*, которое должно содержать не более 40 любых печатаемых символов.

Ширина

Здесь указывается число (минимальное = 1, максимальное = 40), которое определяет (глобально) ширину *Имен строк* в таблице результатов. Этот параметр также можно указать в диалоговом окне *Настройка ширины* (стр. 1262).

Заголовки таблицы результатов (изменение)

Заголовки таблицы результатов можно редактировать в диалоговом окне *Заголовки*, которое вызывается из выпадающего меню *Правка* или с помощью двойного щелчка на заголовке таблицы результатов.



Copyright © StatSoft, 1995

Заголовок окна

По умолчанию *Заголовок окна* содержит название проводимого анализа, а также имя файла. Здесь можно изменить эти значения.

Первая строка

В дополнение к *Заголовку окна* в заголовок таблицы результатов можно ввести еще две строки. В этом поле вводится первый заголовок для таблицы результатов (до 63 печатаемых символов).

Вторая строка

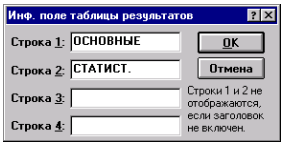
Здесь вводится вторая строка заголовка таблицы результатов (до 63 печатаемых символов). Эта строка не обязательна, но в то же время пространство, зарезервированное в таблице под заголовки, не уменьшится, даже если используется только одна строка. Обратите внимание, что если в поле *Заголовки* выпадающего меню *Вид* (стр. 1265) не стоит галочка (опция не выбрана), то эти две строки отображаться в таблице результатов не будут.

Выравнивание

Здесь можно выбрать выравнивание *По центру*, *Слева* или *Справа* для *Первого* и *Второго заголовка*.

Информационное поле таблицы результатов (изменение)

Информационное поле - это область слева от *Заголовков* таблицы результатов (иногда скрытая под кнопкой *Далее*), которая в дальнейшем может быть использована для записи пользовательской информации о таблице результатов. Содержание *Информационного поля* можно изменить в диалоговом окне *Инф. поле таблицы результатов*.



Здесь можно ввести до 4 строк информации о таблице результатов. Каждая из них может содержать число символов, не превышающее текущую длину строки, которая указана в диалоговом окне *Имя строки* (стр. 1259).

Строка 1

По умолчанию в этой строке система *STATISTICA* записывает название текущего модуля, и продолжает его (если необходимо) на *Строке 2*.

Строка 2

Эта строка по умолчанию является продолжением имени модуля со *Строки 1*. Обратите внимание, что если заголовок таблицы результатов отключен (т.е., в выпадающем меню *Вид* не выбрана опция *Заголовки*, см. стр. 1265), то *Строки 1* и *2* не будут отображаться в *Информационном поле*.

Строка 3

По умолчанию эта строка оставляется пустой в большинстве таблиц результатов.

Строка 4

Эта строка зарезервирована (по умолчанию) для названий *Имен строк*: например, *Переменные*.

В каждой из 4-х строк текст может быть изменен.

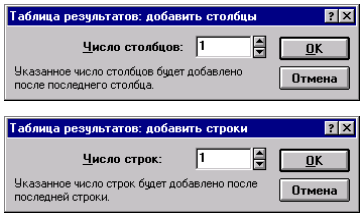
Добавить столбцы/строки в таблицу результатов

Команды *Добавить столбцы* и *Добавить строки* позволяют добавлять пустые столбцы или строки

StatSoft®

Copyright © StatSoft, 1995

после последнего столбца или последней строки текущей таблицы результатов.



После выбора нужной опции из выпадающего меню *Правка* откроется диалоговое окно, в котором нужно указать число добавляемых строк или столбцов. Чтобы добавить указанное число столбцов или строк в таблицу результатов, нужно нажать кнопку *OK* в этом диалоговом окне. Новое пространство может быть использовано для примечаний или информации из таблиц результатов или других окон.

Операции с буфером обмена в таблице результатов

Операции *Вырезать*, *Копировать*, *Копировать содержание*, *Вставить* и *Очистить* аналогичны подобным операциям в окне таблицы исходных данных. Подробную информацию о них можно найти в главе 3.

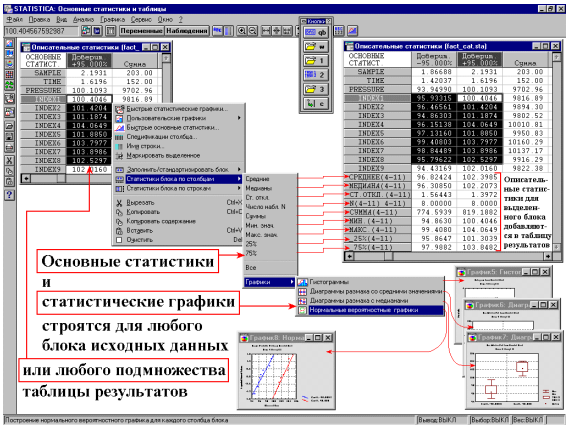
Заполнить/стандартизировать блок

Операции *Заполнить случайными значениями*, *Заполнить/копировать вниз*, *Заполнить/копировать вправо*, *Стандартизировать строки* и *Стандартизировать столбцы* выполняются аналогично таким же операциям в окне таблицы исходных данных.

Подробную информацию о них можно найти в главе 3.

Статистики блока/по строкам и столбцам

Операции, доступные через этот пункт меню, действуют так же, как и в окне таблицы исходных данных. Подробную информацию о них см. в главе 3.



Захват экрана

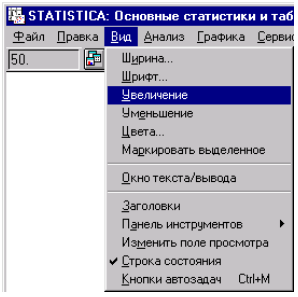
Эта команда доступна из выпадающего меню *Правка* таблицы результатов, а также в любой момент при нажатии клавиш ALT+F3. Она запускает программу *Захват экрана*, которая позволяет выбрать любой прямоугольный участок экрана (или весь экран) и скопировать его в буфер обмена как растровое изображение. Копируемая область может содержать графики, таблицы результатов, таблицы исходных данных и даже окна других приложений, если они в это время отображаются на экране. Скопированное растровое изображение затем можно поместить на графики системы *STATISTICA* или в другие приложения, которые поддерживают растровый формат. Подробную информацию и примеры см. в главе 3.

Copyright © StatSoft, 1995

StatSoft®

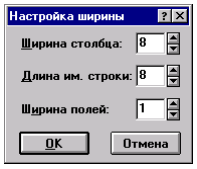
ВЫПАДАЮЩЕЕ МЕНЮ *ВИД* ТАБЛИЦЫ РЕЗУЛЬТАТОВ

Из выпадающего меню *Вид* таблицы результатов доступны показанные и подробно описанные ниже диалоговые окна и опции.



Ширина таблицы результатов

Эта опция выпадающего меню *Вид* открывает диалоговое окно *Настройка ширины,* в котором можно изменять *Ширину столбца*, *Длину имени строки и Ширину полей*. Эти установки являются глобальными (т.е. выбранные здесь параметры будут применены ко всем столбцам, строкам и границам таблицы результатов).



Кроме того, параметры *Ширина столбца* и *Ширина полей* могут быть изменены в диалоговом окне *Столбцы таблицы результатов* (стр. 1258).

Кнопка панели инструментов. См. кнопку панели инструментов *Окна таблицы результатов*.

Ширина столбца

Введенное здесь значение (минимальное = 1, максимальное = 80) определяет ширину каждого столбца таблицы результатов.

Длина имени строки

Здесь указывается число (минимальное = 1, максимальное = 40), которое определяет (глобально) длину имен строк таблицы результатов. Длину имени строки также можно изменить в диалоговом окне *Имя строки таблицы результатов* (стр. 1259).

Ширина полей

В каждом столбце можно ввести поля с обеих сторон всех элементов, чтобы отделить друг от друга значения разных столбцов. Данная опция позволяет указать ширину полей (минимальная = 0, максимальная = 10) для каждого столбца таблицы результатов.

Обратите внимание, что ширина столбца также может быть изменена в процессе работы перетаскиванием границы между именами столбцов или нажатием кнопок и панели инструментов.

Шрифт таблицы результатов

Команда *Шрифт* выпадающего меню *Вид* открывает одноименное диалоговое окно. Здесь можно изменить тип, стиль и размер шрифта текущей таблицы результатов. При выборе параметров можно увидеть образец текста в поле *Образец* этого диалогового окна. Это же окно отображается при выборе шрифта в таблице исходных данных; подробнее см. главу 3.

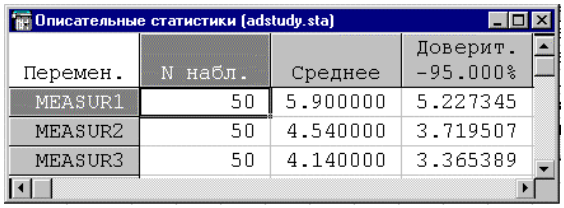
Чтобы выбрать шрифт по умолчанию для всех новых таблиц результатов, нужно воспользоваться опцией *Экран* из выпадающего меню *Сервис* (см. стр. 1269)

StatSoft®

Copyright © StatSoft, 1995

Увеличить таблицу результатов

Эта команда, аналогично соответствующей кнопке панели инструментов таблицы результатов (см. ниже), позволяет постепенно увеличивать шрифт таблицы результатов на значение, определяемое типом шрифта.



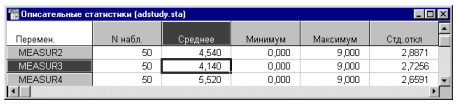
При первом нажатии на эту кнопку может появиться диалоговое окно *Параметры увеличения* (см. ниже) (в зависимости от типа шрифта в таблице результатов) с предупреждением, что текущий шрифт не масштабируется, и с предложением заменить его на другой.

Шрифт в таблице результатов может быть также уменьшен (см. *Уменьшить таблицу результатов* ниже) на то же значение, на которое был увеличен.

Кнопка панели инструментов. См. кнопку панели инструментов *Окна таблицы результатов*.

Уменьшить таблицу результатов

Эта команда, как и соответствующая кнопка панели инструментов таблицы результатов (см. ниже), позволяет постепенно уменьшать шрифт таблицы результатов (на значение, определяемое типом шрифта).



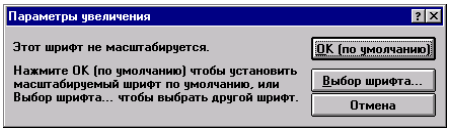
При первом нажатии на эту кнопку может появиться диалоговое окно *Параметры увеличения* (см. ниже) (в зависимости от типа шрифта в таблице результатов) с предупреждением, что текущий шрифт не масштабируется, и с предложением заменить его на другой.

Шрифт в таблице результатов может быть также увеличен (см. *Увеличить таблицу результатов* выше) на то же значение, на которое он был уменьшен.

Кнопка панели инструментов. См. кнопку панели инструментов *Окна таблицы результатов*.

Диалоговое окно Параметры увеличения

При первом нажатии кнопок *Увеличить* или *Уменьшить* на панели инструментов (см. главу 2) или при выборе опций *Увеличить таблицу результатов* или *Уменьшить таблицу результатов* (см. выше), в том случае, если используемый шрифт не масштабируется, откроется диалоговое окно *Параметры увеличения*.



Размер не масштабируемых шрифтов нельзя изменять непрерывно, а только на относительно большие значения.

OK (по умолчанию)

Эту кнопку можно нажать, чтобы установить шрифт по умолчанию и закрыть диалоговое окно.

Copyright © StatSoft, 1995

StatSoft®

Выбрать шрифт

Эта кнопка открывает диалоговое окно *Шрифт,* в котором можно выбрать другой шрифт для таблицы результатов или таблицы исходных данных.

Цвета таблицы результатов

В диалоговом окне *Пользовательские цвета* можно задать цвета для текущих таблиц результатов и таблиц исходных данных. Изменения цвета можно посмотреть на образце таблицы в нижней части диалогового окна. Каждая из его опций открывает стандартное диалоговое окно *Цвет*, в котором можно выбрать цвета из стандартной палитры или задать свой цвет (с помощью кнопки *Определить цвет*). Это же диалоговое окно открывается при выборе цветов для таблицы исходных данных (подробнее см. главу 3).

Чтобы задать цвета для использования по умолчанию во всех новых таблицах результатов, нужно выбрать опцию *Экран* из выпадающего меню *Сервис* (см. стр. 1269)

Маркировать выделенное

Эта команда доступна также через кнопку панели инструментов и из контекстного меню, вызываемого правой кнопкой мыши из любой ячейки. Она позволяет маркировать (или убрать маркировку) отдельные значения или блок значений в таблице результатов, для того, чтобы они отличались от остальных. Сначала нужные ячейки необходимо выделить с помощью курсора, а затем выбрать опцию *Маркировать выделенное* из выпадающего меню *Вид*. Маркированные значения в таблице результатов в окне системы *STATISTICA* имеют цвет, отличный от цвета остальных ячеек (по умолчанию они отображаются красным цветом, см. *Сервис – Экран*, стр. 1269), а при печати по умолчанию отмечаются символом * (звездочка) рядом со значением. Другие опции для отображения выделенных ячеек в распечатываемых и сохраняемых отчетах могут быть выбраны в диалоговом окне *Печать таблицы результатов* (см. стр. 1253).

Часто при создании таблиц результатов некоторые ячейки или блоки маркируются автоматически, чтобы “выделить” статистически значимые результаты (например, коэффициенты корреляции). Для таких таблиц результатов обычно предусмотрена возможность изменения критерия выделения ячеек (например, уровня значимости) с помощью кнопки *Параметры* на панели инструментов таблицы результатов (см. ниже).

Кнопка панели инструментов. См. кнопку панели инструментов *Окна таблицы результатов*.

Кнопка Параметры панели инструментов таблицы результатов

Эта кнопка панели инструментов таблицы результатов позволяет изменять (или выбирать) отдельные параметры (в зависимости от используемого модуля системы *STATISTICA*), которые будут влиять на результаты в таблице.

Например, для некоторых корреляционных матриц эта кнопка откроет диалоговое окно, в котором можно установить предельное значение уровня значимости для тех коэффициентов корреляции, которые будут маркированы в таблице результатов (по умолчанию p = ,05).

Окно текста/вывода

Так же, как и в выпадающем меню *Вид* таблицы исходных данных, эта команда открывает *Окно текста/вывода,* которое предназначено для

StatSoft®

Copyright © StatSoft, 1995

накопления печатаемого вывода (текста и графиков), а также для ввода замечаний и комментариев. Подробную информацию об этом окне можно найти в главе 5.

Заголовки таблицы результатов

С помощью этого пункта меню можно выбрать отображение таблицы результатов с заголовком или без него. Если в выпадающем меню *Вид* напротив пункта *Заголовки* поставлена галочка, то таблица результатов будет отображаться с заголовком из одной или двух строк. В противном случае заголовок отображаться не будет.

Эта опция определяет только отображение заголовков. При выключении отображения информация из них не удаляется. Если дважды щелкнуть на области заголовка таблицы результатов (или выбрать пункт *Заголовки* из выпадающего меню *Правка*), откроется диалоговое окно *Заголовки таблицы результатов,* в котором каждый заголовок можно изменить (стр. 1259).

Панель инструментов таблицы результатов

Выбрав эту опцию из выпадающего меню *Вид*, можно задать один из описанных ниже способов отображения панели инструментов или скрыть ее.

Скрыть. Эта опция временно убирает панель инструментов из окна. Чтобы снова ее показать, нужно выбрать один из описанных ниже способов отображения.

1 строка. Если выбрана эта опция, панель инструментов будет отображаться в окне таблицы в одну строку.

2 строки. Если выбрана эта опция, панель инструментов будет отображаться в две строки.

Сверху и справа. Часть панели инструментов будет отображаться в верхней части окна, а вторая часть - в правой части окна.

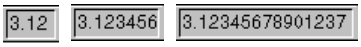
Сверху и слева. Часть панели инструментов будет отображаться в верхней части окна, а вторая часть - в левой части окна.

Крупные значки. Эта опция изменяет размер кнопок панели инструментов. Если стоит галочка, то отображаются крупные значки (это может пригодиться при использовании дисплея с высоким разрешением, например, 1024x768).

Быстрый вызов. Для изменения расположения панели инструментов можно дважды щелкнуть на ней мышью.

Изменить поле просмотра

В *Поле просмотра* отображается содержание выделенной ячейки (или положение курсора на графике). Используя эту команду, можно переключаться между тремя размерами (длинами) *Поля просмотра* (см. главу 2).



Если ширина столбцов (заданная в диалоговом окне *Столбцы таблицы результатов*, стр. 1258) меньше, чем необходимо для отображения всех значащих символов, то с помощью этой опции можно изменить длину *Поля просмотра* и посмотреть значения полностью. Эти изменения можно сделать и с помощью щелчка мышью на самом *Поле просмотра*.

Строка состояния таблицы результатов

С помощью этой опции можно включать (поставив галочку) и выключать отображение

Copyright © StatSoft, 1995

строки состояния в окне таблицы результатов (в нижней части экрана, см. главу 2). Во время выполнения задач система *STATISTICA* использует строку состояния для отображения хода процесса.

Комментарий

Эта область строки состояния используется для краткого (в одну строку) описания активного выпадающего или контекстного меню или команды. Если нет активных меню, данные не обрабатываются, и система *STATISTICA* ожидает следующей команды, то отображается слово *Готово*. Чтобы открыть основное меню *?(Справка),* можно дважды щелкнуть на области комментария строки состояния.

Вывод

В этом поле показано, куда производится вывод анализа, причем это значение соответствует установке, сделанной в диалоговом окне *Параметры страницы/вывода* (см. стр. 1256). Дважды щелкнув на поле *Вывод,* можно открыть это диалоговое окно и задать способ (*Выкл.*, *Файл*, *Принтер* и/или *Окно*) и другие параметры вывода результатов.

Выбор

Здесь отображается текущее состояние (*Вкл.* или *Выкл.*) условий выбора наблюдений. Дважды щелкнув на поле *Выбор* строки состояния, можно открыть диалоговое окно *Условия выбора наблюдений,* где вводятся или редактируются соответствующие условия (см. главу 2).

Вес

Здесь отображается текущее состояние (*Вкл.* или *Выкл.*) условий взвешивания переменных. Дважды щелкнув на поле *Вес* строки состояния, можно открыть диалоговое окно *Вес* где задаются значения веса (см. главу 2).

ЗАП

Это сообщение будет мигать в правой части строки состояния при записи макрокоманды (последовательности нажатия клавиш или движений мыши). Подробнее о записи макрокоманд см. главу 6 *Кнопки автозадач.*

Кнопки автозадач таблицы результатов

Эта опция позволяет включать и выключать отображение панели инструментов с *Кнопками автозадач*. Обратите внимание, что все “горячие” клавиши, присвоенные *Кнопкам автозадач,* могут использоваться вне зависимости от того, отображается эта панель инструментов или нет.

Комбинация клавиш. CTRL+M.

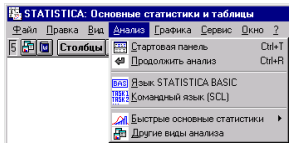
Подробную информацию о *Кнопках автозадач* см. в главе 6.

StatSoft®

Copyright © StatSoft, 1995

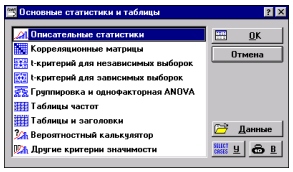
ВЫПАДАЮЩЕЕ МЕНЮ *АНАЛИЗ* ТАБЛИЦЫ РЕЗУЛЬТАТОВ

Из выпадающего меню *Анализ* доступны показанные и описанные ниже диалоговые окно и опции.



Стартовая панель

Если выбрать эту опцию из выпадающего меню *Анализ*, откроется окно, в котором можно выбрать один из типов анализа данных активного модуля системы *STATISTICA*. Содержание стартовой панели зависит непосредственно от модуля (на приведенном ниже примере показана стартовая панель модуля *Основные статистики и таблицы*).



Выбрать нужный тип анализа можно, дважды щелкнув на нем или выделив его и нажав кнопку *OK*.

Комбинация клавиш. CTRL+T

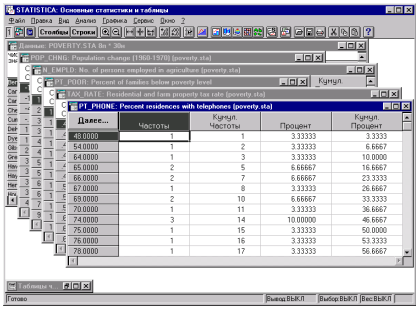
Продолжить анализ

Если производимый анализ еще активен и его свернутое диалоговое окно находится в левом нижнем углу экрана, то можно выбрать эту опцию, чтобы продолжить текущий анализ, т.е. открыть его диалоговое окно. Кроме того, чтобы продолжить анализ, можно дважды щелкнуть на значке Cont или (в большинстве случаев) нажать кнопку *Далее* (см. ниже) в верхнем левом углу таблицы результатов или окна графика.

Комбинация клавиш. CTRL+R

Две кнопки *Далее* в таблице результатов

Часто при анализе данных выводится несколько таблиц результатов. Когда их число превысит установленную длину очереди (см. *Диспетчер таблиц результатов* в главе 3), то на последней таблице появится кнопка *Далее*. При нажатии на эту кнопку обработка продолжится и появится следующая таблица результатов.



Если в результате анализа была открыта только одна таблица результатов (или это последняя таблица в очереди), то нажатие кнопки *Далее* откроет диалоговое окно предыдущего анализа.

Обратите внимание, что можно повторно открыть диалоговое окно анализа и не открывать находящиеся в очереди дополнительные таблицы результатов, для этого нужно дважды щелкнуть на значке анализа в нижнем углу окна системы *STATISTICA*.

Кнопка *Далее* и “всплывающая” кнопка *Далее* в таблице результатов

Функции этих двух кнопок не отличаются, если число таблиц результатов или окон графиков,

Copyright © StatSoft, 1995

StatSoft®

последовательно открываемых при анализе, не превышает текущую, установленную для них "длину очереди" (3 по умолчанию). Но в том случае, если часть документов не открыта и ожидает своей очереди, то:

- нажатие кнопки *Далее* на таблице результатов или графике приведет к тому, что будут открыты последующие таблицы результатов и графики, ожидающие в очереди, тогда как
- нажатие "всплывающей" кнопки *Далее* в нижней части экрана – которая, фактически, представляет свернутое диалоговое окно выбора вывода – приведет к открытию диалогового окна, нарушая очередность вывода таблиц результатов или графиков.

Язык STATISTICA BASIC

Опция *STATISTICA BASIC* выпадающего меню *Анализ* таблицы результатов открывает окно редактора программ языка *STATISTICA BASIC*. Подробную информацию о языке *STATISTICA BASIC* см. в *Электронном руководстве*.

Командный язык

Опция *Командный язык (SCL)* выпадающего меню *Анализ* открывает окно редактора программ *Командного языка (SCL)*. Подробную информацию о *Командном языке (SCL)* см. в *Электронном руководстве*.

Быстрые основные статистики

Опции *Быстрые основные статистики* доступны из контекстного меню или выпадающего меню *Анализ* таблицы результатов или таблицы исходных данных в любом модуле системы *STATISTICA*. Они могут использоваться на любой стадии анализа данных (например, для получения дополнительной информации при просмотре вывода любого модуля системы *STATISTICA*). Эти процедуры позволяют вычислить ряд основных статистик для длинных списков переменных (например, коэффициенты корреляции для *всех* переменных в наборе данных), в том числе и *по группам*, т.е. для каждого значения (кода) выбранной группирующей переменной.

Если вычисление *Быстрых основных статистик* задается из выпадающего меню *Анализ* таблицы результатов, то анализируемые переменные выбираются в последующих диалоговых окнах выбора переменных. Подробную информацию о *Быстрых основных статистиках* см. в главе 10.

Другие виды анализа (Переключатель модулей системы *STATISTICA*)

Опция *Другие виды анализа* выпадающего меню *Анализ* таблицы результатов открывает диалоговое окно *Переключатель модулей*. Это окно содержит список всех доступных модулей системы *STATISTICA*. При выделении конкретного модуля в этом диалоговом окне приводится его краткое описание.

Кнопка панели инструментов. См. кнопку панели инструментов таблицы результатов.

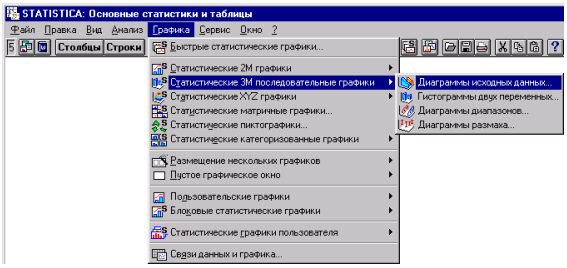
Быстрый вызов. Можно щелкнуть на приложении правой кнопкой мыши для вызова глобального контекстного меню или дважды щелкнуть на приложении (аналогично двойному щелчку на рабочем столе, по которому вызывается *Переключатель задач Windows*).

StatSoft®

Copyright © StatSoft, 1995

ВЫПАДАЮЩЕЕ МЕНЮ *ГРАФИКА* ТАБЛИЦЫ РЕЗУЛЬТАТОВ

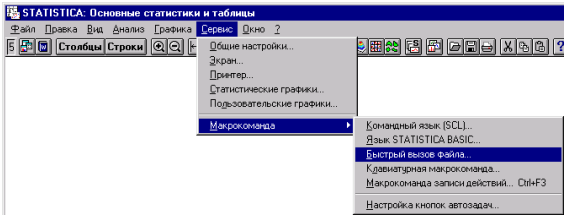
Опции выпадающего меню *Графика* подробно описаны в томе II настоящего руководства.



Краткий обзор этих опций приведен в главе 2.

ВЫПАДАЮЩЕЕ МЕНЮ *СЕРВИС* ТАБЛИЦЫ РЕЗУЛЬТАТОВ

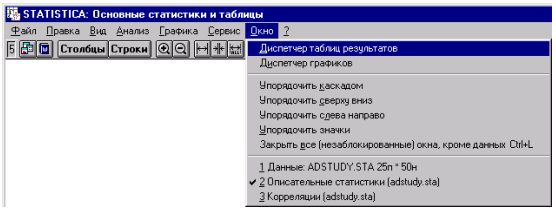
Из выпадающего меню *Сервис* таблицы результатов доступны показанные ниже диалоговые окна и опции.



Все они подробно описаны в главе 3, за исключением опций *Статистические графики* и *Пользовательские графики*, которые рассматриваются в томе II настоящего руководства.

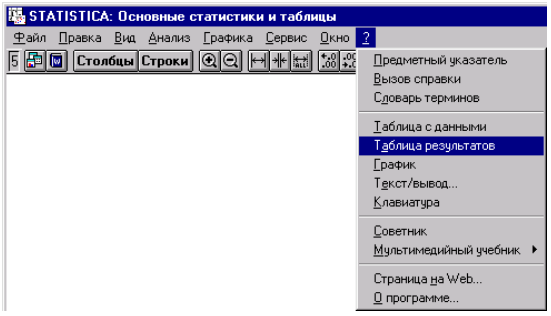
ВЫПАДАЮЩЕЕ МЕНЮ *ОКНО* ТАБЛИЦЫ РЕЗУЛЬТАТОВ

Выпадающее меню *Окно* таблицы результатов содержит показанные ниже опции, которые подробно описаны в главе 3.



ВЫПАДАЮЩЕЕ МЕНЮ *?(СПРАВКА)* ТАБЛИЦЫ РЕЗУЛЬТАТОВ

Ниже показаны опции выпадающего меню *?(Справка)* таблицы результатов.



Каждая из них открывает окно *Справка,* содержащее информацию по указанной теме (более подробно содержание соответствующих разделов описано в главе 3).

Copyright © StatSoft, 1995

StatSoft®

ПРЕДМЕТНЫЙ УКАЗАТЕЛЬ

А

Автозаполнение, 1249

Б

Быстрые основные статистики, 1268

В

Вид
- заголовки, 1265
- панель инструментов, 1265
- строка состояния, 1265
- ширина столбца/полей, 1262

Выход, 1257

Д

Диалоговое окно Параметры страницы/вывода, 1256
Длина имени строки (в таблицах результатов), 1262
Добавить столбцы, 1260
Добавить строки, 1260

З

Захват экрана, 1261

И

Изменение скорости
- выделение блока, 1250
- прокрутка, 1250

Изменение ширины полей, 1262
Изменение ширины столбцов в таблицах результатов, 1259
Изменить поле просмотра, 1265

К

Кнопка Параметры, 1264
Кнопки Далее, 1267
Командный язык (SCL), 1268

М

Маркировать выделенное, 1264
Метод *Перетащить и отпустить*
- Автозаполнение, 1249
- выделение блока, 1248
- изменение скорости выделения блока, 1250
- копирование блока, 1249
- обзор, 1248
- очистка блока, 1249
- перемещение блока, 1248
- экстраполяция блока, 1249
- экстраполяция блока вверх или влево, 1250

О

Окна, разделение отображения, 1250
Окно текста/вывода, 1264
Операции с блоками (метод *Перетащить и отпустить*), 1248
Открыть
- графический файл, 1252
- создать данные, 1252
- текстовый файл, 1252
- файл данных, 1252
- файл таблицы результатов, 1252

Отображение разделенных окон, 1250

П

Пакетная печать, 1255
Параметры вывода текста на принтер, 1255
Параметры по умолчанию
- печать, 1256

Параметры принтера, 1255
Параметры увеличения, 1263
Переключатель модулей, 1268
Печать
- выпадающее меню Файл таблицы результатов, 1253
- графические файлы и таблицы результатов, 1255
- диалоговое окно Параметры страницы/вывода, 1256
- пакетная, 1255
- параметры по умолчанию, 1256
- параметры принтера, 1255
- печать/выдача текущих страниц, 1253
- принтер, 1256
- файлы, 1255

Печать после выхода из окна, 1254
Поля
- изменение ширины, 1259

Правка
- заголовки, 1259
- имена строк, 1259
- информационное поле, 1260
- ячейка, 1258

Принтер, 1256
Продолжить анализ, 1267

Р

Разделение прокрутки в таблицах исходных данных и таблицах результатов, 1250
Ручка заполнения, 1249, 1250

С

Создать данные, 1252
Стартовая панель, 1267
Строка состояния, 1265

Т

Таблица исходных данных
- метод *Перетащить и отпустить*, 1248
- разделение прокрутки, 1250

Таблица результатов
- выпадающее меню ?(Справка), 1269
- выпадающее меню Анализ, 1267
- выпадающее меню Вид, 1262
- выпадающее меню Графика, 1269
- выпадающее меню Окно, 1269
- выпадающее меню Правка, 1258
- выпадающее меню Сервис, 1269
- выпадающее меню Файл, 1252
- добавить столбцы/строки, 1260
- заголовки, 1265
- заполнить/стандартизировать блок, 1261
- изменить заголовки, 1259
- изменить имена строк, 1259
- изменить информационное поле, 1260

StatSoft®

Copyright © StatSoft, 1995

изменить спецификации столбца, 1258
изменить ячейку, 1258
кнопка Параметры, 1264
кнопки Далее, 1267
маркированные ячейки, 1254
маркировать выделенное, 1264
метод *Перетащить и отпустить*, 1248
операции с буфером обмена, 1261
панель инструментов, 1265
параметры увеличения, 1263
печать, 1253
просмотр ширины, 1262
разделение прокрутки, 1250
сохранить, 1252
сохранить как, 1253
сохранить как файл данных, 1253
Строка состояния, 1265
увеличить, 1263
уменьшить, 1263

У

Увеличить
таблицу результатов, 1263
Уменьшить
таблицу результатов, 1263

Ш

Ширина столбца (в таблицах результатов), 1262

Э

Экстраполяция блока, 1249

Я

Язык STATISTICA BASIC, 1268

Copyright © StatSoft, 1995

StatSoft®

Copyright © StatSoft, 1995

Глава 5:

ОКНО ТЕКСТА/ВЫВОДА: МЕНЮ, ДИАЛОГОВЫЕ ОКНА, СВОЙСТВА

Оглавление

Вводная часть 1276
Как открыть *Окно текста/вывода* 1277
Сохранение файлов из *Окна текста/вывода* 1278
Печать из *Окна текста/вывода* 1280
Форматирование текста в *Окне текста/вывода* 1281
Другие пункты выпадающих меню *Окна текста/вывода* 1285
Предметный указатель 1290

Подробное оглавление

ВВЕДЕНИЕ 1276
Вывод некоторых (или всех) таблиц результатов и графиков 1276
Просмотр текстовых файлов 1277
Блокноты 1277

КАК ОТКРЫТЬ *ОКНО ТЕКСТА/ВЫВОДА* 1277
Как открыть *Окно текста/вывода* из диалогового окна *Параметры по умолчанию: общие настройки* 1277
Открыть текстовый файл 1277
Окно текста/вывода (через меню *Вид*) 1278
Как открыть *Окно текста/вывода* через диалоговое окно *Параметры страницы/вывода* 1278
Авт. дополнение содержания *Окна текста/вывода* 1278
Авт. печать всех таблиц результатов и графиков (Автоотчет) 1278

СОХРАНЕНИЕ ФАЙЛОВ ИЗ *ОКНА ТЕКСТА/ВЫВОДА* 1278
Файлы в формате *RTF* 1279
Текстовые файлы 1279
Сохранить 1279
Сохранить как… 1279
Как текст 1279

ПЕЧАТЬ ИЗ *ОКНА ТЕКСТА/ВЫВОДА* 1280
Печать выделенного текста 1280

Copyright © StatSoft, 1995

StatSoft®

Поля вывода текста ... 1280

ФОРМАТИРОВАНИЕ ТЕКСТА В *ОКНЕ ТЕКСТА/ВЫВОДА* ... **1281**

Форматирование символов ... 1281
Размеры и стили шрифта ... 1281
Эффекты шрифта ... 1282
Цвет ... 1282
Форматирование абзацев ... 1282
Нормальный ... 1282
Выровнять по левому краю ... 1282
Выровнять по центру ... 1283
Выровнять по правому краю ... 1283
Одиночный интервал ... 1283
Двойной интервал ... 1283
Отступ слева ... 1283
Отступ справа ... 1283
Висячая строка ... 1284
Линейка ... 1284
Задание позиций табуляции в *Окне текста/вывода* ... 1284
Удаление позиций табуляции ... 1284
Печать табуляции при выводе текстовых таблиц ... 1284
Разрыв страницы ... 1285

ДРУГИЕ ПУНКТЫ ВЫПАДАЮЩИХ МЕНЮ *ОКНА ТЕКСТА/ВЫВОДА* ... **1285**

Новое окно ... 1285
Специальная вставка ... 1285
Найти значение ... 1285
Найти ... 1286
Направление ... 1286
Учитывать регистр ... 1286
Найти далее ... 1286
Заменить значение ... 1286
Найти ... 1286
Заменить на ... 1286
Поиск ... 1286
Подтверждать каждую замену ... 1286
Выделить все ... 1287
Перейти ... 1287
Отмена в *Окне текста/вывода* ... 1287
Добавлять результаты ... 1287
Вставить рисунок из файла ... 1287
Размер объекта ... 1288
Высота и ширина картинки ... 1288
Единицы измерения ... 1288
Изменение размеров объекта ... 1288

StatSoft®

Copyright © StatSoft, 1995

Объект 2x3 1288
Объект 3x4,5 1288
Объект 4x6 1288
Пользовательский размер 1288
Размер графического объекта и минимальная толщина линии 1289

ПРЕДМЕТНЫЙ УКАЗАТЕЛЬ 1290

Copyright © StatSoft, 1995

Глава 5:

ОКНО ТЕКСТА/ВЫВОДА: МЕНЮ, ДИАЛОГОВЫЕ ОКНА, СВОЙСТВА

ВВЕДЕНИЕ

Окно текста/вывода действует так же, как и текстовый редактор, и позволяет создавать текстовые и графические отчеты практически неограниченного размера. Оно совместимо с форматом *RTF* (*R*ich *T*ext *F*ormat – расширенный текстовый формат), поддерживает форматированный текст и графику (см. стр.1279).

Окно текста/вывода можно использовать:

(1) для просмотра и редактирования отчета, который был выведен на экран как таблицы результатов и/или графики, полученные в ходе анализа (предусмотрена возможность автоматического перенаправления вывода текста и/или графических изображений в это окно, см. стр. 1278);

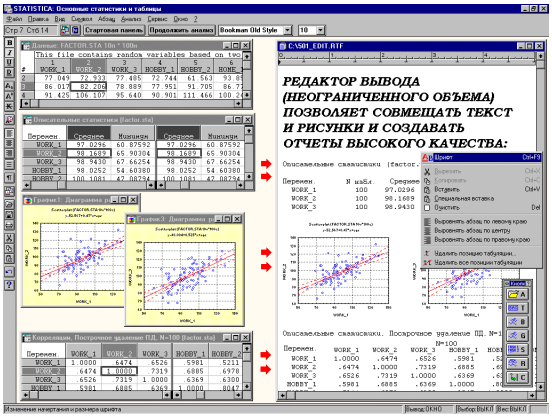
(2) для открытия (а также просмотра или редактирования) файлов вывода или любых других текстовых или текстовых и графических (*RTF*) файлов; а также

(3) как локальный текстовый и графический блокнот.

Каждое из этих трех возможных применений будет кратко описано в следующих пунктах.

Вывод некоторых (или всех) таблиц результатов и графиков

Чтобы открыть *Окно текста/вывода* , необходимо выбрать пункт *Окно текста/вывода* в выпадающем меню *Вид*. Если это окно будет открыто, то при каждом выводе таблицы исходных данных, графика, таблицы результатов или выделенного блока ячеек, здесь будет появляться выводимый элемент.



Этот канал вывода может быть также определен в диалоговом окне *Параметры страницы/вывода,* которое можно вызвать, дважды щелкнув на поле *Вывод* строки состояний окна системы *STATISTICA*, или из пункта *Принтер* выпадающего меню *Сервис.*

Если в этом диалоговом окне также выбран вывод на *Принтер* или в *Файл*, то результаты будут направлены одновременно в *Окно текста/вывода* и в этот канал. Кроме того, вы можете выбрать одну из опций *Автоотчета* (см. главу 6), и тогда при каждом отображении таблиц результатов и/или графиков на экране, они будут посылаться в текущий канал вывода и одновременно в *Окно текста/вывода* (если оно открыто).

StatSoft

Copyright © StatSoft, 1995

Просмотр текстовых файлов

Независимо от текущего открытого *Окна текста/вывода* (см. предыдущий абзац), записанные ранее текстовые файлы могут быть открыты в новом окне путем выбора пункта *Файл текста/вывода...* в списке команды *Открыть другие* (выпадающее меню *Файл*). В *Окне текста/вывода* можно открыть и отредактировать все текстовые (ASCII) файлы и файлы с форматированным текстом (**.rtf*), содержащие графические изображения.

Блокноты

Вне зависимости от того, как было открыто *Окно текста/вывода* (см. предыдущие абзацы), его всегда можно использовать, как блокнот (вы можете набирать/редактировать текст или вставлять текст и графики или другие изображения из буфера обмена).

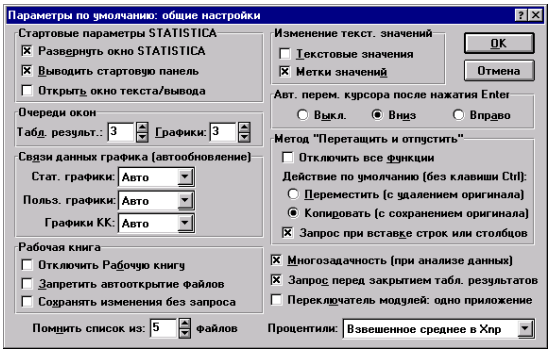
КАК ОТКРЫТЬ *ОКНО ТЕКСТА/ВЫВОДА*

В системе *STATISTICA* предусмотрено несколько возможностей открыть *Окно текста/вывода*. После того, как оно было открыто, вы можете направлять в него вывод текста и графических изображений.

Если вы включите какую-либо из опций *Автоотчета* (например, *Авт. печать всех таблиц результатов*) в диалоговом окне *Параметры страницы/вывода* (см. Главу 3), то соответствующий вывод (например, графики или данные каждой таблицы результатов, отображенной на экране) также будет автоматически направлен в *Окно текста/вывода*.

Как открыть *Окно текста/вывода* из диалогового окна *Параметры по умолчанию: общие настройки*

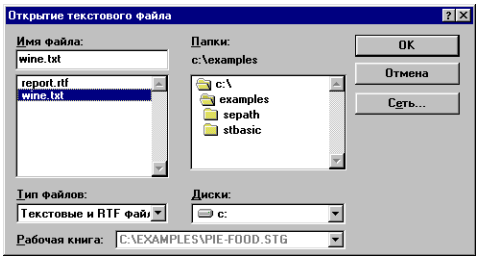
Чтобы *Окно текста/вывода* автоматически открывалось каждый раз, когда открывается модуль системы *STATISTICA*, выберите пункт *Открыть Окно текста/вывода* в диалоговом окне *Параметры по умолчанию: общие настройки* системы *STATISTICA* (см. третью опцию в группе *Стартовые параметры*).



Если вы выберете эту опцию, она будет использоваться по умолчанию до тех пор, пока вы не измените выбор.

Открыть текстовый файл

В *Окне текста/вывода* можно открыть ранее сохраненный текстовый файл (расширение **.txt*) или файл в формате *RTF* (расширение**.rtf*).



Для этого нужно выбрать пункт *Открыть текст* из выпадающего меню *Файл* или нажать кнопку

Copyright © StatSoft, 1995

StatSoft®

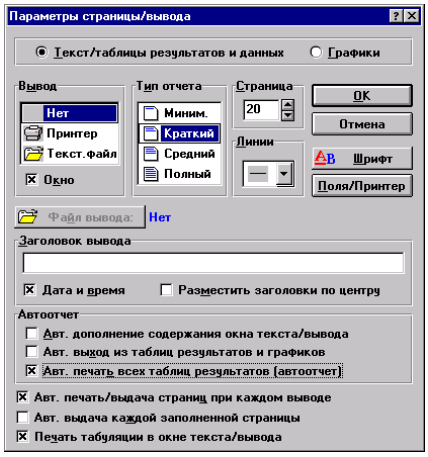
на панели инструментов или комбинацию клавиш CTRL+O для вывода диалогового окна *Открытие текстового файла*.

Окно текста/вывода (через меню *Вид*)

Если вы выберете этот пункт из выпадающего меню *Вид*, будет открыто *Окно текста/вывода,* которое было сохранено последним. Весь вывод будет перенаправлен в это окно (независимо от того, производится ли также вывод на принтер или в файл).

Как открыть *Окно текста/вывода* через диалоговое окно *Параметры страницы/вывода*

Вы также можете открыть *Окно текста/вывода*, выбрав опцию *Окно* в диалоговом окне *Параметры страницы/вывода* (см. также главу 3).



При выборе этой опции независимо от того, куда выводятся все таблицы результатов или исходных данных (в *Файл* или на *Принтер*), они будут одновременно направлены в *Окно текста/вывода*. Если в этом диалоговом окне также выбрана опция *Авт. печать всех таблиц результатов/графиков (Автоотчет)*, то вывод результатов анализов также будет автоматически “напечатан” в *Окно текста/вывода* (см. ниже).

Авт. дополнение содержания *Окна текста/вывода*

Если вы выберете эту опцию в диалоговом окне *Параметры страницы/вывода* , система *STATISTICA* автоматически откроет последний сохраненный текстовый файл (с расширением **.txt*) или файл в формате *RTF* (с расширением **.rtf*) в новом *Окне текста/вывода* (например, чтобы добавить текущий вывод к результатам предыдущего анализа).

Если эта опция не выбрана, то из диалогового окна *Параметры страницы/вывода* будет открыто пустое *Окно текста/вывода* (если выбрана опция *Окно,* см. выше).

Авт. печать всех таблиц результатов и графиков (Автоотчет)

Если в диалоговом окне *Параметры страницы/вывода* выбран этот режим печати, то весь вывод анализа и/или графиков будет автоматически направлен на принтер или в файл и/или в *Окно текста/вывода* (если в этом диалоговом окне выбрана опция *Окно*). Этот режим печати удобен, если вы хотите сохранить полный отчет о результатах, отображаемых на экране во время анализа. Для получения подробной информации см. главу 3.

СОХРАНЕНИЕ ФАЙЛОВ ИЗ *ОКНА ТЕКСТА/ВЫВОДА*

В *Окно текста/вывода* может быть вставлен форматированный текст и графические изображения из системы *STATISTICA* или других приложений Windows. Эта информация потом может быть сохранена в одном из описанных ниже форматов: как текстовый файл или как файл в формате *RTF*.

StatSoft®

Copyright © StatSoft, 1995

Файлы в формате *RTF*

Файлы в формате *RTF* (*R*ich *T*ext *F*ormat – расширенный текстовый формат) являются стандартными файлами, содержащими форматированный текст и графические изображения, используемыми системой *STATISTICA* для сохранения файлов отчета (т.е., вывода из *Окна текста/вывода*). Если файлы сохраняются в формате *RTF* (**.rtf*), все форматирование будет также сохранено, чтобы эти файлы можно было использовать в других приложениях, поддерживающих формат *RTF* (например, в программе MS Word). Формат *RTF* является стандартом компании Microsoft для сохранения форматированного текста и графических изображений с целью обмена данными между приложениями.

Текстовые файлы

При сохранении содержания *Окна текста/вывода* в виде текстового файла (с расширением **.txt*; воспользуйтесь опцией *Как текст*, описанной ниже) форматирование, в отличие от файлов формата *RTF*, сохранено не будет. Поэтому текстовый файл не будет содержать графических изображений и информации о шрифтах, отступах, разметке страницы и пр. Тем не менее, этот файл может быть в дальнейшем использован в приложениях, которые не поддерживают стандарт *RTF*.

Если вы собираетесь постоянно сохранять текст и/или графические изображения из *Окна текста/вывода* (включая ваши примечания и прочие вставки), вы можете воспользоваться следующими пунктами из выпадающего меню *Файл*.

Сохранить

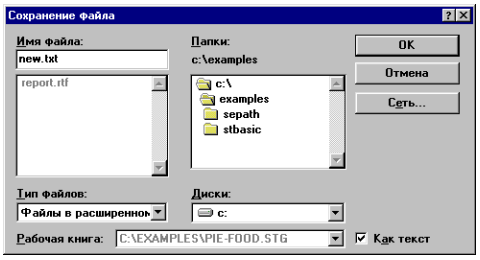
Когда вы выберете этот пункт из меню или нажмете кнопку на панели инструментов, система *STATISTICA* сохранит содержание *Окна текста/вывода* под текущим именем и расширением файла (т.е., **.rtf* или **.txt*), не выводя промежуточного диалогового окна. Если вы хотите сохранить содержание *Окна текста/вывода* под другим именем, воспользуйтесь опцией *Сохранить как…,* описанной ниже.

Комбинация клавиш. Нажмите CTRL+S.

Сохранить как…

При выборе этого пункта (или нажатии клавиши F12) появится диалоговое окно *Сохранение файла*.

Комбинация клавиш. Нажмите F12.



В этом окне можно задать параметры для сохранения содержания *Окна текста/вывода* как текстового файла (с расширением **.txt*; используйте пункт *Как текст*, описанный ниже) или как файла в формате *RTF* (с расширением **.rtf*).

Как текст

Этот пункт доступен только из *Окна текста/вывода* [включая программные файлы языка *STATISTICA BASIC*, программные файлы командного языка *SCL*, командные файлы пакета *SEPATH* и прочие приложения *Окна*

Copyright © StatSoft, 1995

StatSoft®

текста/вывода]. Воспользуйтесь этой опцией для сохранения содержания *Окна текста/вывода* в текстовом файле стандартного ASCII формата (расширение **.txt*).

Если эта опция *не выбрана*, система *STATISTICA* запишет файл в формате *RTF* , чтобы сохранить форматирование (шрифт, расстояния, графические изображения, и т. д.). Обычно формат по умолчанию (*RTF*) предпочтительнее, но такой файл можно изменить только в программах, поддерживающих формат *RTF*. Если вы сохраните файл в текстовом формате, то сможете использовать его практически в любой программе, однако, форматирование и графические изображения в этом случае будут утеряны.

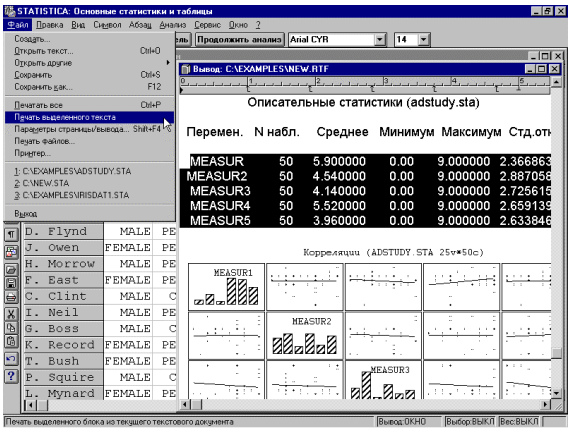
ПЕЧАТЬ ИЗ *ОКНА ТЕКСТА/ВЫВОДА*

Из-за особого статуса этого окна печать из него производится особым образом, отличным от печати из таблиц исходных данных и таблиц результатов. Когда вы нажмете кнопку *Печать* , комбинацию клавиш CTRL+P или выберете пункт *Печатать все* из выпадающего меню *Файл Окна текста/вывода*, все содержание этого окна будет отправлено непосредственно на принтер [вне зависимости от текущих настроек поля *Вывод* строки состояния, которая располагается в нижней части окна системы *STATISTICA*]. Например, оно будет отправлено на принтер, даже если вывод на *Принтер* отключен (*Off*) или перенаправлен в файл.

Чтобы напечатать только выделенный блок текста из *Окна текста/вывода*, вы можете воспользоваться пунктом *Печать выделенного текста* (см. ниже). Самый быстрый способ сохранить полный постоянный отчет о всех отображаемых на экране результатах - это выбрать некоторые опции *Автоотчета* (диалоговое окно *Параметры страницы/вывода*, см. главу 3). При этом все отображаемые на экране таблицы результатов и/или окна с графиками автоматически направляются в *Окно текста/вывода.*

Печать выделенного текста

Воспользуйтесь этим пунктом выпадающего меню *Файл Окна текста/вывода* для печати только выбранного (выделенного) блока текста в *Окне текста/вывода*.



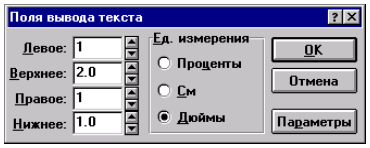
Чтобы напечатать весь текст, воспользуйтесь пунктом *Печатать все* выпадающего меню *Файл*.

Поля вывода текста

В этом диалоговом окне (опция *Поля/Принтер* в диалоговом окне *Параметры страницы/вывода*), вы можете задать расположение текста (например, из *Окна текста/вывода*, таблиц исходных данных или таблиц результатов) на печатаемой странице путем задания полей (определения значений отступа слева, сверху, справа и снизу) в соответствии с единицами измерения (*Проценты*, *См* или *Дюймы*).

StatSoft®

Copyright © StatSoft, 1995



Обратите внимание, что эти параметры полей не зависят от установок полей для вывода графических изображений (которые задаются в диалоговых окнах *Предварительный просмотр* и *Поля графиков*).

Для задания размера и ориентации бумаги (*книжная* или *альбомная*) воспользуйтесь пунктом *Принтер* выпадающего меню *Файл*.

ФОРМАТИРОВАНИЕ ТЕКСТА В *ОКНЕ ТЕКСТА/ВЫВОДА*

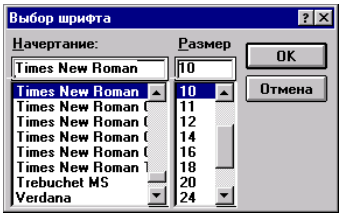
Результаты анализа, также как и добавленный текст могут быть отформатированы с использованием различных шрифтов, выравнивания абзацев, внедренных графических изображений и других объектов. Для сохранения форматированного текста из этого окна его необходимо записать в файл в формате *RTF* (*R*ich *T*ext *F*ormat – расширенный текстовый формат) (см. стр. 1279).

Форматирование символов

Далее рассмотрены возможности управления форматом символов в *Окне текста/вывода*. Эти опции доступны из выпадающего меню *Символ* или, в некоторых случаях, через панель инструментов *Окна текста/вывода* (см. *Кнопки панели инструментов Окна текста/вывода*, глава 2). Также для некоторых из этих опций приведены соответствующие комбинации клавиш и кнопки панели инструментов.

Размеры и стили шрифта

В диалоговом окне *Выбор шрифта* (вызывается выбором пункта *Шрифт* выпадающего меню *Символ* или нажатием клавиш CTRL+F9) можно изменять стиль шрифта (*Начертание*) и его *Размер* для выделенного и нового текста в *Окне текста/вывода*.



Кроме того, для задания необходимого стиля и размера шрифта вы можете воспользоваться полем *Выбор шрифта* (см. ниже) на панели инструментов *Окна текста/вывода*.

Обратите внимание, что если текущий выделенный блок текста в *Окне текста/вывода* содержит различные шрифты, то при вставке нового текста или замене блока будет использован тот тип шрифта, название которого показано в поле *Шрифт* панели инструментов *Окна текста/вывода*.

Установка шрифта по умолчанию. Когда таблица исходных данных или таблица результатов отображается в *Окне текста/вывода*, по умолчанию будет использован шрифт, заданный в диалоговом окне *Параметры страницы/вывода* (см. главу 3) для вывода таблиц исходных данных и таблиц результатов. Поэтому, если вы хотите, чтобы таблицы исходных данных или таблицы результатов отображались в *Окне текста/вывода* другим шрифтом, вам нужно будет задать этот шрифт в диалоговом окне *Параметры страницы/вывода*.

Copyright © StatSoft, 1995
StatSoft®

Если вы вставляете текст из другого приложения, то для вставки текста в *Окно текста/вывода* с использованием оригинального шрифта выберите в диалоговом окне *Специальная вставка* (см. стр. 1285) в качестве формата вставки тип *RTF*.

Эффекты шрифта

В *Окне текста/вывода* могут быть использованы следующие атрибуты шрифтов.

Обычный. При выборе этого пункта выпадающего меню *Символ* [или нажатии клавиш CTRL+0 (ноль)] шрифт станет обычным (например, не курсивным).

Полужирный. При выборе этой опции (кнопка панели инструментов или клавиши CTRL+B) начертание последующих или выделенных символов станет **полужирным**.

Курсив. При выборе этой опции (кнопка панели инструментов или клавиши CTRL+I) символы будут выводиться *курсивом*.

Подчеркивание. При выборе этой опции (кнопка панели инструментов или клавиши CTRL+U) последующие или выделенные символы будут подчеркнуты.

Двойное подчеркивание. При выборе этой опции (кнопка панели инструментов или клавиши CTRL+D) последующие или выделенные символы будут подчеркнуты дважды.

Нижний индекс. При выборе этой опции (кнопка панели инструментов) последующие или выделенные символы будут написаны нижним индексом (например, нижний $_{индекс}$).

Верхний индекс. При выборе этой опции (кнопка панели инструментов) последующие или выделенные символы будут написаны верхним индексом (например, верхний ^{индекс}).

Зачеркивание. При выборе этой опции (кнопка панели инструментов) последующие или выделенные символы будут ~~зачеркнуты~~.

Скрытый. При выборе этого пункта из выпадающего меню *Символ* последующие или выделенные символы будут скрыты. Вы можете включать и выключать отображение скрытых символов, нажимая кнопку панели инструментов или выбирая пункт *Скрытые символы* в меню *Вид*.

Цвет

Выбрав этот пункт (кнопка панели инструментов или клавиши SHIFT+F9), можно поменять цвет шрифта (с помощью стандартного для Windows диалогового окна *Цвет*).

Форматирование абзацев

Эти опции позволяют задавать форматирование абзацев в *Окне текста/вывода*. Каждая из них доступна как пункт выпадающего меню *Абзац*. Кроме того, можно изменить выравнивание и интервалы абзаца, нажав соответствующую кнопку панели инструментов (см. ниже). Обратите внимание, что переключение режимов отображения и скрытия символов абзаца (¶) и скрытого текста в окне осуществляется с помощью кнопки панели инструментов или пункта *Скрытые символы* выпадающего меню *Вид*.

Нормальный

Выбор этой опции вернет форматирование абзаца по умолчанию (т. е. выравнивание по левому краю и одиночный интервал).

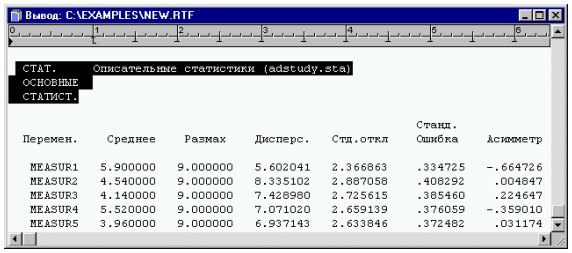
Выровнять по левому краю

Выбор этой опции (или нажатие кнопки панели инструментов) приведет к выравниванию

StatSoft®

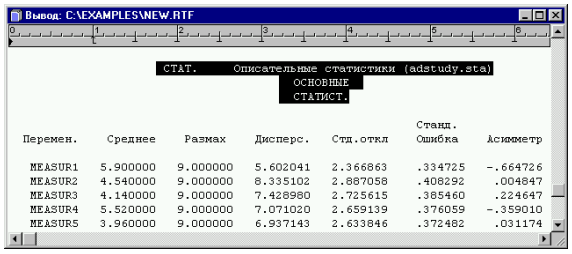
Copyright © StatSoft, 1995

абзаца по левому краю страницы (см. выделенный текст на рисунке).



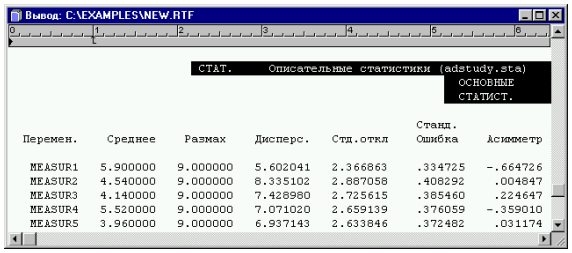
Выровнять по центру

Выбор этой опции (или нажатие кнопки панели инструментов) приведет к расположению абзаца посередине страницы.



Выровнять по правому краю

Выбор этой опции (или нажатие кнопки панели инструментов) приведет к выравниванию абзаца по правому краю страницы.

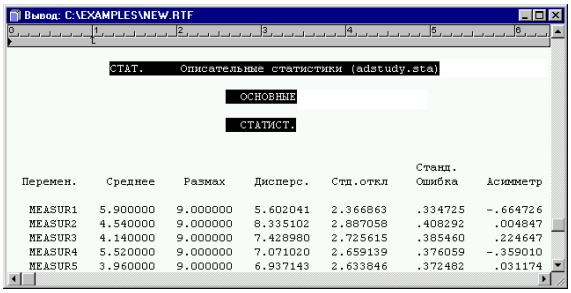


Одиночный интервал

Выбор этой опции (или нажатие кнопки панели инструментов) приведет к установлению одиночного интервала между строками абзаца.

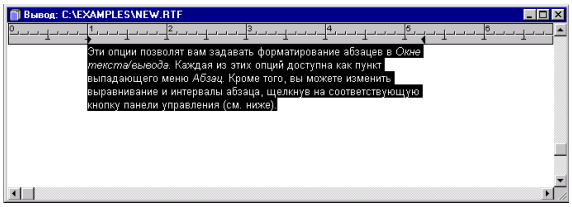
Двойной интервал

Выбор этой опции (или нажатие кнопки панели инструментов) приведет к установлению двойного интервала между строками абзаца.



Отступ слева

Этот пункт позволяет задать для абзаца величину отступа от левого края страницы (см. выделенный текст на следующем рисунке).



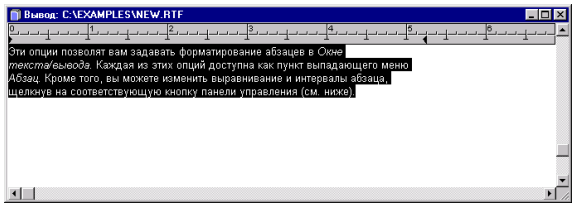
Каждый выбор этого пункта в выпадающем меню *Абзац* увеличит отступ слева на 0,5 дюйма.

Отступ справа

Этот пункт позволяет задать для абзаца величину отступа от правого края страницы.

Copyright © StatSoft, 1995

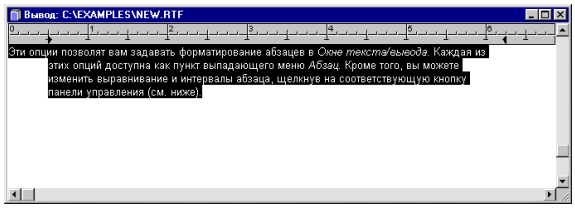
StatSoft®



Каждый выбор этого пункта в выпадающем меню *Абзац* увеличит отступ справа на 0,25 дюйма.

Висячая строка

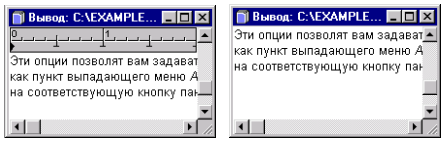
Этот пункт позволяет формировать висячие строки (т.е., отступ второй и последующих строк абзаца).



Каждый выбор этого пункта в выпадающем меню *Абзац* увеличит отступ на 0,5 дюйма.

Линейка

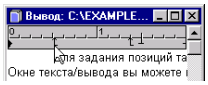
Этот пункт позволяет включать и выключать отображение линейки в *Окне текста/вывода*. Линейка используется как измерительная шкала и отображается в верхней части *Окна текста/вывода*.



Для задания позиций табуляции нужно щелкнуть мышью на линейке (см. ниже).

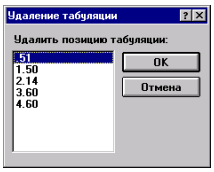
Задание позиций табуляции в *Окне текста/вывода*

Для задания позиций табуляции *Окне текста/вывода* нужно щелкнуть на линейке левой кнопкой мыши. Позиции табуляции будут отмечаться на линейке стрелками (см. ниже).



Удаление позиций табуляции

Чтобы удалить одну позицию табуляции, выберите пункт *Удалить позицию табуляции* из контекстного меню, всплывающего при нажатии правой кнопки мыши. Откроется диалоговое окно *Удалить позицию табуляции*, в котором можно выбрать позицию табуляции для удаления.



Кроме того, можно удалить позицию табуляции, щелкая на стрелке позиции табуляции, показанной на линейке *Окна текста/вывода*.

Чтобы удалить все позиции табуляции в документе, выберите пункт *Удалить все позиции табуляции* из контекстного меню, всплывающего при нажатии правой кнопки мыши.

Печать табуляции при выводе текстовых таблиц

Если в диалоговом окне *Параметры страницы/вывода* выбрана опция *Печать табуляции в Окне текста/вывода* (см. главу 3), программа будет выводить в *Окно текста/вывода* таблицы, содержащие табуляцию.

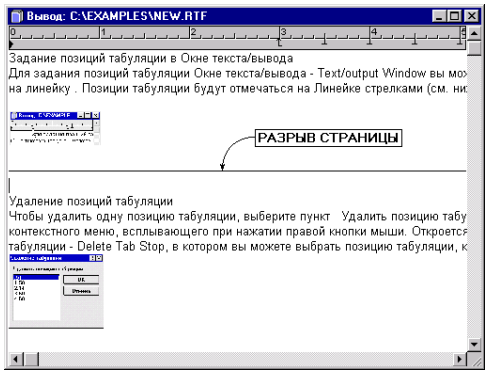
StatSoft®
Copyright © StatSoft, 1995

Обратите внимание, что если в текстовой таблице присутствует табуляция, то при вставке ее в документ текстового редактора (например, MS Word) возможно использование непропорциональных шрифтов (например, Times Roman). Кроме того, такую таблицу можно вставить в таблицу исходных данных так, что все значения будут помещены в различные ячейки.

Подробное описание этой опции см. в разделе *Параметры страницы/вывода,* глава 3.

Разрыв страницы

Чтобы поместить в *Окно текста/вывода* жесткий (неавтоматический) разрыв страницы, воспользуйтесь пунктом *Вставить разрыв страницы* из выпадающего меню *Символ* (или нажмите CTRL+ENTER). Этот разрыв будет отображен в окне как сплошная горизонтальная черта.



Вы можете использовать жесткие разрывы страницы для переноса текста или объектов на новую страницу, сохранения расположения абзацев, размещения объектов или текста на верхней части страницы, и т.д.

ДРУГИЕ ПУНКТЫ ВЫПАДАЮЩИХ МЕНЮ *ОКНА ТЕКСТА/ВЫВОДА*

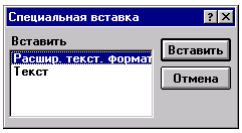
В дополнение к описанным выше пунктам (*Как открыть Окно текста/вывода*, стр. 1277, *Сохранение файлов из Окна текста/вывода*, стр. 1278, *Печать из Окна текста/вывода*, стр. 1280, и *Форматирование текста в Окне текста/вывода*, стр. 1281) из выпадающих меню *Окна текста/вывода* доступны следующие команды.

Новое окно

Этот пункт выпадающего меню *Файл Окна текста/вывода* используется для удаления всего текста из текущего *Окна текста/вывода*.

Специальная вставка

С помощью этой команды вызывается показанное ниже диалоговое окно. В нем задается формат (например, Расширенный текстовый формат, текст или картинка) для вставки объекта из буфера обмена в *Окно текста/вывода*. Список доступных форматов определяется текущим содержанием буфера обмена.



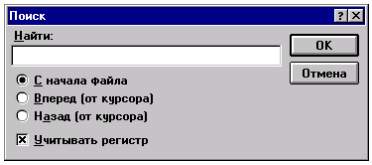
Для сохранения шрифта при отображении текста, вставляемого из другого приложения, в диалоговом окне *Специальная вставка* необходимо выбрать формат *RTF*.

Найти значение

При выборе пункта *Найти* из выпадающего меню *Правка* или нажатии клавиш CTRL+F

Copyright © StatSoft, 1995
StatSoft®

откроется диалоговое окно *Поиск*, в котором можно задать текст или число для поиска в *Окне текста/вывода*.



Все опции этого диалогового окна описаны ниже.

Найти

В этом поле ввода задается текст или число, которое необходимо найти в *Окне текста/вывода*.

Направление

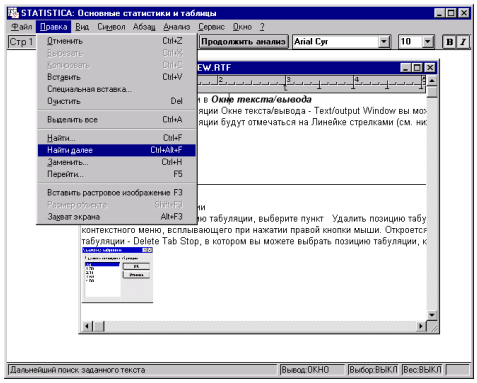
Поиск будет производиться *С начала файла* или от текущей позиции курсора (*Вперед* или *Назад*).

Учитывать регистр

Если эта опция выбрана, система *STATISTICA* будет производить поиск точного совпадения верхнего/нижнего регистра с искомым текстом.

Найти далее

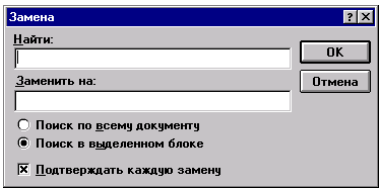
Пункт *Найти далее* доступен из выпадающего меню *Правка.*



После того, как системой *STATISTICA* был найден заданный текст или значение, можно продолжить поиск в *Окне текста/вывода*, нажав кнопку *Найти далее*. Можно продолжить поиск, не вызывая это окно, а выбрав пункт *Найти далее* из выпадающего меню *Правка* (или нажав клавиши CTRL+ALT+F).

Заменить значение

Диалоговое окно *Заменить* (доступное из выпадающего меню *Правка* или по нажатию клавиш CTRL+H) позволяет найти заданный текст или число в *Окне текста/вывода* и заменить его на другой текст или число.



Это диалоговое окно содержит следующие опции.

Найти

В этом поле ввода задается текст или число для поиска в *Окне текста/вывода*.

Заменить на

В этом поле ввода задается текст или число, на которое будет заменяться искомый текст или число.

Поиск

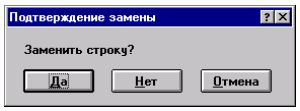
Вы можете выбрать *Поиск по всему документу* или *Поиск в выделенном блоке*.

Подтверждать каждую замену

При выборе этой опции система *STATISTICA* будет выдавать запрос на подтверждение замены.

StatSoft®
Copyright © StatSoft, 1995

Когда будет обнаружен заданный текст или значение, откроется диалоговое окно, в котором предоставлен выбор одной из трех возможностей: заменять (нажмите *Да*), не заменять (нажмите *Нет*) или отменить процедуру замены (кнопка *Отмена*).



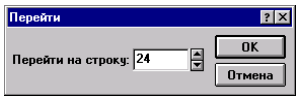
Если эта опция не выбрана, то будет заменен весь текст или числа, заданные в поле ввода *Найти*.

Выделить все

Выбор этого пункта из выпадающего меню *Правка* или нажатие клавиш CTRL+A приведет к выделению всего текста и объектов в *Окне текста/вывода*.

Перейти

Выбор этого пункта из выпадающего меню *Правка* или нажатие клавиши F5 приведет к переходу на заданную позицию текста (*Перейти на строку*).

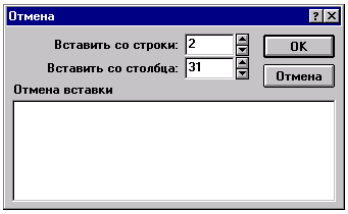


После ввода нужного номера строки надо нажать *OK*.

Отмена в *Окне текста/вывода*

Выбор этого пункта из выпадающего меню *Правка,* нажатие кнопки панели инструментов или клавиш CTRL+Z (или ALT+BACKSPACE) приводит к отмене последней команды или действия. Эта опция действует для следующих типов команд и действий:

- Вырезание, вставка, очистка;
- Последний ввод текста или числа в *Окне текста/вывода*.



Диалоговое окно *Отмена* отображает тип операции отмены и позволяет задать место (номер строки и столбца), в котором ее нужно произвести.

Добавлять результаты

Если этот пункт выбран (выпадающее меню *Сервис*), то результаты анализа добавляются к концу любого текста в *Окне текста/вывода*. Это полезно для составления непрерывного отчета о результатах анализа. В то же время при использовании *Окна текста/вывода* в качестве “блокнота“ для записи примечаний режим *Добавления* можно временно выключить, отменив выбор пункта *Добавлять результаты* (чтобы снова вернуться в этот режим, выберите этот пункт еще раз).

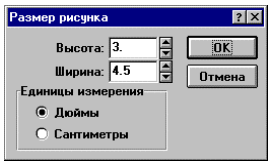
Вставить рисунок из файла

Выбор этого пункта из выпадающего меню *Правка* или нажатие клавиши F3 позволяет вставить объект из файла в текстовый документ в *Окне текста/вывода*. При выборе этого пункта откроется диалоговое окно *Открытие /Сохранение файла,* в котором вы можете выбрать нужный файл. Если нажать *OK* в этом диалоговом окне, то объект будет вставлен на текущую позицию курсора.

Copyright © StatSoft, 1995

Размер объекта

Это диалоговое окно вызывается из выпадающего меню *Правка*, с помощью клавиш SHIFT+F3 или из пункта *Пользовательский размер* контекстного всплывающего меню. В этом окне можно изменить размер объекта, выбранного в *Окне текста/вывода*.



Данное диалоговое окно содержит описанные ниже опции.

Высота и ширина картинки

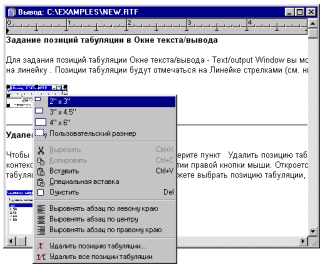
В этом поле задается высота и ширина выбранной картинки (в соответствии с заданными единицами измерения, см. ниже).

Единицы измерения

В этом поле выбираются единицы измерения размеров картинки (высота и ширина): дюймы или сантиметры.

Изменение размеров объекта

Кроме описанных выше способов задания размеров, существует возможность выбора некоторых размеров из контекстного меню для объекта, который выделен в *Окне текста/вывода*.



В этом списке можно быстро выбрать различные размеры объекта.

Объект 2x3

После выбора этого варианта выбранный объект будет иметь размер 2x3 (высота x ширина) дюйма. См. ниже замечание об установках минимальной ширины линии.

Объект 3x4,5

После выбора этого варианта выбранный объект будет иметь размер 3x4,5 (высота x ширина) дюйма. См. ниже замечание об установках минимальной ширины линии.

Объект 4x6

После выбора этого варианта выбранный объект будет иметь размер 4x6 (высота x ширина) дюйма. См. ниже замечание об установках минимальной ширины линии.

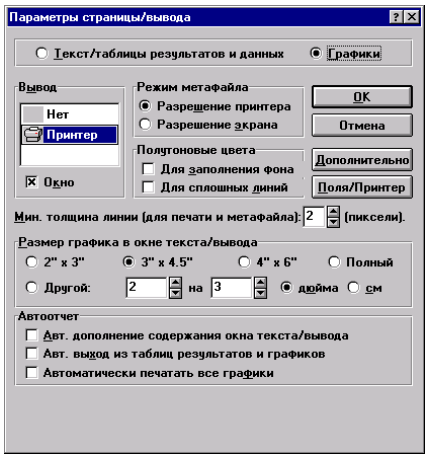
Пользовательский размер

Когда вы выберете этот пункт, откроется диалоговое окно *Размер рисунка,* в котором вы можете ввести необходимый размер объекта (см. стр. 1288). См. ниже замечание об установках минимальной ширины линии.

StatSoft®
Copyright © StatSoft, 1995

Размер графического объекта и минимальная толщина линии

Обратите внимание, что при уменьшении графика системы *STATISTICA* в *Окне текста/вывода* (например, до 2x3) некоторые линии при печати могут оказаться слишком тонкими. В этом случае для нормальной печати может потребоваться изменение значений минимальной толщины линии в диалоговом окне *Параметры страницы/вывода* (см. главу 3).



Copyright © StatSoft, 1995

StatSoft®

ПРЕДМЕТНЫЙ УКАЗАТЕЛЬ

А

Абзац, форматирование, 1282
  висячая строка, 1284
  выровнять по левому краю, 1282
  выровнять по правому краю, 1283
  выровнять по центру, 1283
  двойной интервал, 1283
  одиночный интервал, 1283
  отступ слева, 1283
  отступ справа, 1283
Автоматическое дополнение содержания Окна текста/вывода, 1278
Альбомная ориентация, 1281

В

Вставка рисунка из файла, 1287
Вывод таблиц результатов и графиков, 1276
Выделить все в Окне текста/вывода, 1287

Д

Добавление результатов в Окне текста/вывода, 1287

З

Задание позиций табуляции в Окне текста/вывода, 1284
Заменить в Окне текста/вывода, 1286
  подтверждение, 1286

К

Книжная ориентация, 1281

Н

Найти далее в Окне текста/вывода, 1286

О

Окно блокнота, 1277
Окно текста/вывода, 1276
  вставка рисунка, 1287
  вывод таблиц результатов и графиков, 1276
  выделить все, 1287
  добавление результатов, 1287
  замена, 1286
  линейка, 1284
  найти далее, 1286
  найти значение, 1285
  новое окно, 1285
  окно блокнота, 1277
  открыть по умолчанию, 1277
  открыть текстовый файл, 1277
  открыть файл, 1277
  отмена, 1287
  перейти, 1287
  печать выделенного текста, 1280
  печать файлов, 1280
  подтверждать замену, 1286
  позиции табуляции, 1284
  поля, 1280
  поля вывода текста, 1280
  размер объекта, 1288
  разрыв страницы, 1285
  рисунок, 1287
  сохранить как текст, 1279
  сохранить как…, 1279
  сохранить файлы, 1278
  специальная вставка, 1285
  стиль шрифта, 1282
  текст, содержащий табуляцию, 1284
  текстовые файлы, 1277, 1279
  удаление позиций табуляции, 1284
  формат RTF, 1279
  форматирование абзацев, 1282
  форматирование символов, 1281
  форматирование текста, 1281
  цвет шрифта, 1282
  шрифт, 1281
  шрифт по умолчанию, 1281
Отмена в Окне текста/вывода, 1287

П

Параметры страницы/вывода
  автоматическое дополнение содержания Окна текста/вывода, 1278
Перейти в Окне текста/вывода, 1287
Печать
  альбомная ориентация, 1281
  книжная ориентация, 1281
  поля вывода текста, 1280
Печать табуляции в Окне текста/вывода, 1284
Печать файлов из Окна текста/вывода, 1280
Подтверждать замену в Окне текста/вывода, 1286
Позиции табуляции в Окне текста/вывода
  задание, 1284
  удаление, 1284
Поиск в Окне текста/вывода, 1285
Поля в Окне текста/вывода, 1280

Р

Результаты, отображение окна текста/вывода, 1276
Рисунок в Окне текста/вывода, 1287

С

Символы, форматирование, 1281
Скрытый текст в Окне текста/вывода, 1282
Специальная вставка, 1285
Стиль шрифта в Окне текста/вывода
  верхний индекс, 1282
  двойное подчеркивание, 1282
  зачеркивание, 1282
  курсив, 1282
  нижний индекс, 1282
  обычный, 1282
  подчеркивание, 1282
  полужирный, 1282
  скрытый, 1282

Т

Табуляция в Окне текста/вывода
  печать, 1284
Текстовые файлы, 1277

Ф

Формат RTF (расширенный текстовый формат), 1279

StatSoft®

Copyright © StatSoft, 1995

Форматирование текста в Окне текста/вывода, 1281

Ц

Цвет шрифта в Окне текста/вывода, 1282

Ш

Шрифт, форматирование, 1281
Шрифт, цвет в Окне текста/вывода, 1282

Copyright © StatSoft, 1995

StatSoft®

Copyright © StatSoft, 1995

Глава 6:

Кнопки Автозадач

Оглавление

Вводный обзор 1296
Изменение *кнопок автозадач* 1298
Вид и описание *кнопок автозадач* 1298
Панель инструментов 1299
Изменить/задать кнопку 1299
Удалить кнопки 1299
Добавить кнопки 1299
Файлы автозадач 1300
Сохранять изменения 1301
Значки на кнопках 1301
“Горячие” клавиши 1301
Изменение *кнопок автозадач* 1302
Кнопка 1302
Тип 1302
Описание 1302
Имя файла 1302
Клавиша 1302
Изменить программу 1303
Запись или ввод клавиатурной макрокоманды 1303
Название/Тип кнопки 1304
Описание 1304
Клавиша 1304
Последовательность нажатий клавиш 1304
Кнопки быстрого ввода 1305
Открыть 1306
Сохранить 1306
Описание синтаксиса клавиатурных макрокоманд 1306
Макрокоманда записи действий 1310
Расположение панели инструментов *Кнопки автозадач* 1311
Закрепить панель инструментов *Кнопки автозадач* 1312
Изменить панель инструментов *кнопки автозадач* 1312
Использование “Горячих” клавиш 1312
Предметный указатель 1313
Подробное оглавление см. на следующей странице.

Copyright © StatSoft, 1995

Подробное оглавление

ВВОДНЫЙ ОБЗОР 1296
Кнопки автозадач: задаваемый пользователем “быстрый вызов одним щелчком” 1296
Пользовательские дополнения к системе *STATISTICA* 1297
Большие библиотеки *кнопок автозадач* 1297

ИЗМЕНЕНИЕ *КНОПОК АВТОЗАДАЧ* 1298
Вид и описание *кнопок автозадач* 1298
Вид 1298
Изменение порядка кнопок 1298
Изменение кнопки 1298
Панель инструментов 1299
Изменить/задать кнопку 1299
Удалить кнопки 1299
Добавить кнопки 1299
Программа на Командном языке (*SCL*) 1299
Программа на языке *STATISTICA BASIC* 1299
Быстрый вызов файла (данные, график) 1300
Клавиатурная макрокоманда (запись и/или изменение) 1300
Запись действий (с движениями мыши) 1300
Файлы автозадач 1300
Открыть 1300
Сохранить как 1301
По умолчанию 1301
Сохранять изменения 1301
Значки на кнопках 1301
“Горячие” клавиши 1301

ИЗМЕНЕНИЕ *КНОПОК АВТОЗАДАЧ* 1302
Кнопка 1302
Тип 1302
Описание 1302
Имя файла 1302
Клавиша 1302
Глобальные системные “горячие” клавиши (с CTRL) 1303
Локальные системные “горячие” клавиши (с CTRL) 1303
Изменить программу 1303

ЗАПИСЬ ИЛИ ВВОД КЛАВИАТУРНОЙ МАКРОКОМАНДЫ 1303
Название/Тип кнопки 1304
Описание 1304
Клавиша 1304

StatSoft®
Copyright © StatSoft, 1995

Последовательность нажатий клавиш ... 1304
Ввод/изменение последовательности нажатий клавиш ... 1305
Запись ... 1305
Кнопки быстрого ввода ... 1305
Открыть ... 1306
Сохранить ... 1306
Операции с буфером обмена (вырезать, копировать, вставить) ... 1306
Описание синтаксиса клавиатурных макрокоманд ... 1306
Ввод обычных символов ... 1306
Ввод специальных символов ... 1306
Порядок нажатий клавиш ... 1307
Повторяющиеся нажатия клавиш ... 1307
Команда WAIT ... 1307
Советы по вводу клавиатурных макрокоманд ... 1307
Пример: Ввод последовательности нажатий клавиш в макрокоманду ... 1308

МАКРОКОМАНДА ЗАПИСИ ДЕЙСТВИЙ ... 1310
Название/тип кнопки ... 1310
Описание ... 1310
Имя файла ... 1310
Клавиша ... 1311
Быстрое воспроизведение ... 1311
Запись ... 1311

РАСПОЛОЖЕНИЕ ПАНЕЛИ ИНСТРУМЕНТОВ *КНОПКИ АВТОЗАДАЧ* ... 1311
Закрепить панель инструментов *Кнопки автозадач* ... 1312

ИЗМЕНИТЬ ПАНЕЛЬ ИНСТРУМЕНТОВ *КНОПКИ АВТОЗАДАЧ* ... 1312

ИСПОЛЬЗОВАНИЕ “ГОРЯЧИХ” КЛАВИШ ... 1312

ПРЕДМЕТНЫЙ УКАЗАТЕЛЬ ... 1313

Copyright © StatSoft, 1995

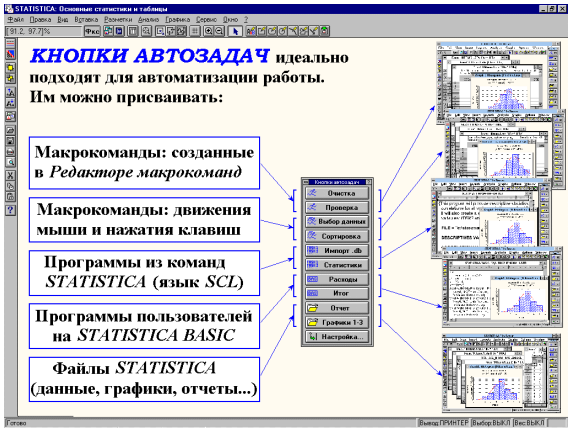
Глава 6:

Кнопки автозадач

ВВОДНЫЙ ОБЗОР

Кнопки автозадач: задаваемый пользователем “быстрый вызов одним щелчком”

Кнопки автозадач – это всплывающая (или фиксированная), задаваемая пользователем панель инструментов, которая позволяет быстро вызывать большую часть функций системы *STATISTICA*.



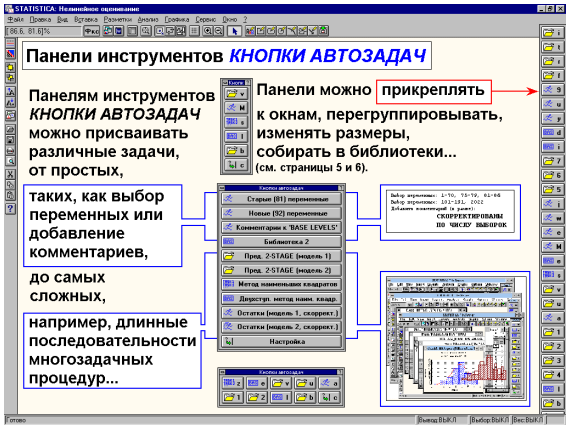
Для упрощения и ускорения доступа к заданным пользователем процедурам формируется панель инструментов *Кнопки автозадач.* Каждой кнопке на этой панели можно поставить в соответствие любую из перечисленных ниже операций:

- макрокоманды, созданные путем записи нажатий клавиш и движений мыши;
- макрокоманды, созданные путем ввода последовательностей нажатий клавиш или изменения ранее записанных макрокоманд;
- наборы процедур, созданные с помощью *SCL* (Командного языка системы *STATISTICA*);
- заданные пользователем процедуры и программы, написанные на языке *STATISTICA BASIC* (например, новые графики, новые процедуры вычисления, преобразования данных и т.д.);
- определенные файлы системы *STATISTICA* (данные, графики, отчеты и пр.), к которым необходим быстрый доступ.

Процедуры, присвоенные отдельным *кнопкам автозадач,* могут также быть вызваны нажатием заданных пользователем “горячих” клавиш. Кроме того, кнопки, которым присвоены макрокоманды, могут использоваться, когда открыто диалоговое окно.

Например, можно создать клавиатурную макрокоманду, которая будет выбирать определенные параметры в часто используемом диалоговом окне (например, *Статистические графики*). После этого при открытии данного окна достаточно нажать соответствующую *кнопку автозадач* (или, возможно, заданную комбинацию “горячих” клавиш), чтобы запустить процедуру выбора параметров. (Полезные советы о введении клавиатурных макрокоманд можно найти в разделе *Советы* на стр. 1307.)

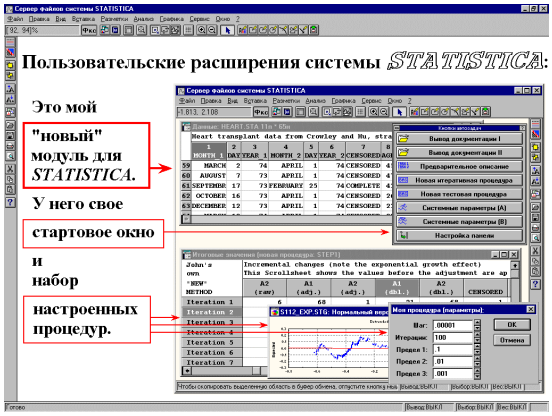
StatSoft
Copyright © StatSoft, 1995



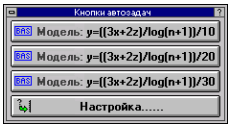
Обратите внимание, что отображение панели инструментов *Кнопки автозадач* можно включать и выключать нажатием клавиш CTRL+M.

Пользовательские дополнения к системе *STATISTICA*

Всплывающие панели инструментов *Кнопки автозадач* представляют собой эффективный интерфейс к пользовательским дополнениям системы *STATISTICA*, поскольку им могут быть присвоены созданные пользователем процедуры любого типа.



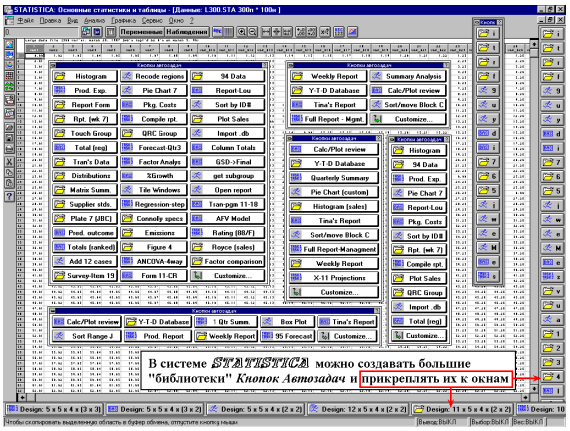
Например, они могут содержать набор заданных пользователем моделей регрессии,



различные типы графиков с заданными пользователем параметрами (определенные с помощью макрокоманд или языка *STATISTICA BASIC*).

Большие библиотеки *кнопок автозадач*

Число создаваемых панелей инструментов *Кнопки автозадач* не ограничено. Каждая из них, в свою очередь, может содержать любое число кнопок.



Панели инструментов могут быть *глобальными* (доступны из любого модуля) или *локальными* (доступны только в определенном модуле или проекте), см. стр. 1301.

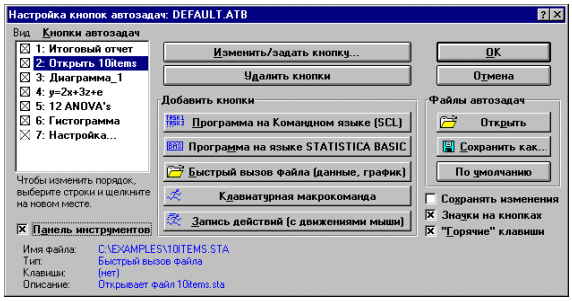
Библиотеки таких списков процедур пользователя можно сохранять. Например, они могут содержать создаваемые по шаблону отчеты, длинные последовательности анализа или ряд программ, используемых для преобразования или очистки данных. Кроме того, в них можно включать “меню” заданных пользователем проверок,

Copyright © StatSoft, 1995
StatSoft

определенных пользовательских графиков, диаграмм или рисунков (например, динамически связанных со значениями данных в наборе данных системы *STATISTICA*).

ИЗМЕНЕНИЕ *КНОПОК АВТОЗАДАЧ*

Содержание панели инструментов *Кнопки автозадач* может быть изменено путем редактирования (например, переназначения), добавления или удаления кнопок, а также изменения размеров меню (перетаскиванием границ панели инструментов с помощью левой кнопки мыши), а затем сохранено в файле. После нажатия на кнопку *Настройка* панели инструментов *Кнопки автозадач*, откроется диалоговое окно *Настройка кнопок автозадач,* в котором можно добавить, удалить, изменить или просмотреть содержание отдельных *кнопок автозадач*.



Ниже подробно описаны опции этого диалогового окна.

Вид и описание *кнопок автозадач*

В этой области диалогового окна отображается список названий кнопок текущей панели инструментов *Кнопки автозадач*. На экране в панели инструментов отображаются только те кнопки, которые отмечены в этом списке (см. ниже пункт *Вид*). Когда кнопка в списке выделена, под списком отображается имя соответствующего файла, тип кнопки, “горячие” клавиши (если они заданы) и введенное ранее описание кнопки (см. рисунок выше).

Вид

Чтобы выбрать кнопки, которые будут отображены в панели инструментов *Кнопки автозадач*, нужно отметить соответствующие им поля в этой области диалогового окна. Если поле не отмечено, кнопка не будет отображаться в панели инструментов, но останется в списке, и ее действия будут вызываться по нажатию соответствующих “горячих” клавиш (если они были присвоены). Таким образом, панель инструментов *Кнопки автозадач* можно настроить так, чтобы в ней отображались только нужные кнопки.

Изменение порядка кнопок

Чтобы изменить порядок кнопок на панели инструментов, нужно сначала выделить в списке одно или несколько названий кнопок, затем переместить курсор мыши на нужную позицию в списке, и, когда курсор изменится на ⇳, нажать левую кнопку мыши.

Изменение кнопки

Чтобы изменить название кнопки, ее описание, “горячие” клавиши или действие, нужно выполнить одно из перечисленных ниже действий.

StatSoft®
Copyright © StatSoft, 1995

(1) Выделить нужное название кнопки в списке диалогового окна *Настройка кнопок автозадач* и нажать кнопку *Изменить/задать кнопку* (см. ниже), которая откроет соответствующее диалоговое окно.

(2) Дважды щелкнуть на нужном названии кнопки в списке кнопок в диалоговом окне *Настройка кнопок автозадач.*

(3) Вне этого диалогового окна, нажать на нужную кнопку на панели инструментов *Кнопки автозадач,* удерживая при этом нажатой клавишу CTRL.

В результате применения каждого из этих способов откроется соответствующее диалоговое окно, в котором можно изменить кнопку.

Панель инструментов

Чтобы выключить отображение панели инструментов *Кнопки автозадач,* нужно убрать галочку из этого поля. Кроме того, включить/выключить панель инструментов можно, нажав клавиши CTRL+M, или с помощью опции *Кнопки автозадач* выпадающего меню *Вид*.

Обратите внимание, что если некоторым кнопкам присвоены “горячие” клавиши, то соответствующие им действия по-прежнему доступны с клавиатуры, даже если отображение панели инструментов *Кнопки автозадач* выключено.

Изменить/задать кнопку

Эту кнопку нужно нажать, чтобы изменить название выделенной кнопки, ее описание, имя файла, “горячие” клавиши и другие параметры.

Удалить кнопки

Нажатие на эту кнопку приведет к тому, что все кнопки, выделенные в области *Кнопки автозадач* этого диалогового окна, будут удалены из списка. Обратите внимание, что эта команда удаляет только выделенные названия из списка, но не связанные с ними файлы.

Добавить кнопки

В этом поле диалогового окна отображаются опции для добавления новых кнопок в панель инструментов *Кнопки автозадач*. Каждая из этих кнопок открывает соответствующее диалоговое окно (см. ниже), в котором можно задать новую кнопку и выбрать действие, которое будет ей присвоено.

Программа на Командном языке (*SCL*)

Эта кнопка открывает диалоговое окно *Настройка кнопок автозадач,* в котором соответствующей кнопке можно поставить в соответствие файл с программой на языке *SCL* (расширение **.scl*). Подробную информацию о других опциях этого диалогового окна см. в пункте *Изменение кнопки автозадач* на стр. 1302.

Программа на языке *STATISTICA BASIC*

Эта кнопка открывает диалоговое окно *Изменение кнопки автозадач,* в котором можно выбрать сохраненный ранее файл с программой на языке *STATISTICA BASIC* (расширение **.stb*). Подробную информацию о других опциях этого диалогового окна см. в пункте *Изменение кнопки автозадач* на стр. 1302.

Copyright © StatSoft, 1995

Быстрый вызов файла (данные, график)

Эта кнопка открывает диалоговое окно *Изменение кнопки автозадач* (см. стр. 1302), в котором можно выбрать файл, принадлежащий одному из типов файлов системы *STATISTICA* (например, файл данных, графический файл, файл таблицы результатов, текстовый файл, файл в формате *RTF* и т.д.); выбранный файл будет быстро вызываться нажатием соответствующей *Кнопки автозадач*.

Клавиатурная макрокоманда (запись и/или изменение)

Эта кнопка открывает диалоговое окно *Запись или ввод клавиатурной макрокоманды,* в котором можно открыть, ввести или записать клавиатурную макрокоманду. Она представляет собой изменяемую макрокоманду, которая либо вводится непосредственно (учитывая стандартные соглашения инструкции Microsoft *Sendkeys*), либо записывается. Также здесь можно открыть текстовый файл, содержащий ранее сохраненную последовательность нажатий клавиш. Подробную информацию об изменяемых макрокомандах и синтаксисе *Sendkeys* см. в пункте *Запись или ввод клавиатурной макрокоманды* (стр. 1303).

Запись действий (с движениями мыши)

Эта кнопка открывает диалоговое окно *Макрокоманда записи действий*, в котором можно включить и начать запись движений мыши и нажатий клавиш. При воспроизведении макрокоманды с *Записью действий* будут воспроизводиться также и записанные движения мыши и/или нажатия клавиш. Предусмотрено два варианта воспроизведения записанных макрокоманд: со скоростью записи или в режиме *Быстрое воспроизведение*.

Обратите внимание, что макрокоманды, содержащие записанные движения мыши, "контекстно-зависимы" и могут привести к другим результатам, если размер окна или его положение на экране изменились. Поэтому для воспроизведения лучше использовать клавиатурные макрокоманды (все меню и диалоговые окна системы *STATISTICA* имеют собственные комбинации клавиш).

Также обратите внимание, что записанные в макрокомандах движения мыши не могут быть изменены. Однако если с помощью операции *Запись действий* были записаны нажатия клавиш, они могут быть импортированы в диалоговом окне *Запись или ввод клавиатурной макрокоманды* (см. стр. 1303).

Файлы автозадач

Следующие опции позволяют изменять или сохранять файл конфигурации панели *Кнопки автозадач*. Имя текущего файла *кнопок автозадач* отображено в верхней части этого диалогового окна. Обратите внимание, что можно также быстро переключаться между файлами конфигурации с помощью опции *Изменить панель кнопок автозадач* контекстного меню, вызываемого правой кнопкой мыши (см. стр. 1312).

Открыть

Эта кнопка открывает диалоговое окно *Открытие файла кнопок автозадач,* в котором можно выбрать ранее сохраненный файл конфигурации *кнопок автозадач*. Когда выбранный файл (с расширением **.atb*) открыт, он ставится в соответствие текущему открытому модулю в файле *Statist.ini* (о файле *Statist.ini* см. в *Электронном руководстве*), а имя текущего модуля убирается из предыдущей конфигурации (если она была).

Обратите внимание, что сохраненный ранее файл конфигурации *кнопок автозадач* можно также открыть, щелкнув правой кнопкой мыши на

StatSoft®

Copyright © StatSoft, 1995

панели инструментов и выбрав из контекстного меню пункт *Изменить панель кнопок автозадач* (см. стр. 1312).

Сохранить как

Кнопка *Сохранить как* открывает диалоговое окно *Сохранение файла кнопок автозадач* и сохраняет конфигурацию *Кнопок автозадач* под именем по умолчанию (*Default.atb*) или под новым именем (расширение **.atb*). При записи конфигурации панели *Кнопки автозадач*, имя файла, путь и связанный с ним модуль будут указаны в файле *Statist.ini* (о файле *Statist.ini* см. в *Электронном руководстве*).

Глобальные и локальные *кнопки автозадач*. Файл конфигурации *кнопок автозадач* может быть поставлен в соответствие одновременно нескольким модулям (т.е. глобально) Это соответствие для каждого модуля записано в файле *Statist.ini*. Однако каждый модуль может иметь только одну связанную с ним (т.е. локальную) конфигурацию *Кнопок автозадач*. Чтобы присвоить файл конфигурации конкретному модулю, нужно либо сохранить, либо открыть его в этом модуле.

Например, файл конфигурации *Default.atb* будет открываться в каждом модуле, которому он поставлен в соответствие в файле *Statist.ini* [по умолчанию, в каждом модуле системы *STATISTICA*]. Если внести изменения в это файл и сохранить его с тем же путем и именем, то он по-прежнему будет связан с каждым из модулей, с которыми был связан ранее.

Если внести изменения в файл конфигурации *Default.atb* и сохранить его под другим именем или в другом каталоге (например, *New.atb*), то новый файл конфигурации будет соответствовать (локально) текущему модулю, а имя текущего модуля будет удалено из списка модулей, связанных с файлом *Default.atb*. Тем не менее, файл *Default.atb* по-прежнему будет присвоен оставшимся модулям системы *STATISTICA*.

По умолчанию

Эта кнопка открывает и ставит в соответствие текущему модулю файл конфигурации *кнопок автозадач Default.atb*.

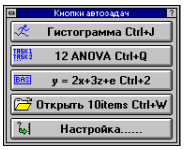
Сохранять изменения

Эта опция предназначена для того, чтобы автоматически сохранять изменения в текущем файле конфигурации *кнопок автозадач,* не выводя запроса на подтверждение (выдаваемого системой *STATISTICA*).

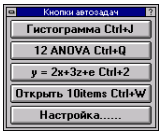
Значки на кнопках

Если убрать галочку напротив этой опции, значки, обозначающие тип *Кнопки автозадач,* не будут отображаться в левой части кнопок (см. ниже).

Значки: Вкл.



Значки: Выкл.

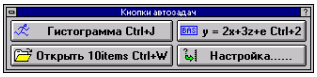


Обратите внимание, что эту опцию можно использовать, если необходимо уменьшить размер панели инструментов *Кнопки автозадач* или задать кнопки только буквами.

“Горячие” клавиши

Эта опция предназначена для того, чтобы включать после названия кнопки отображение “горячих” клавиш, которые могут быть присвоены *Кнопкам автозадач*.

'Горячие' клавиши: Вкл



‘Горячие' клавиши: Выкл

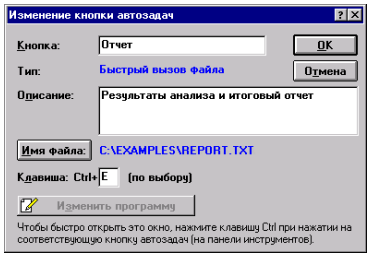
Copyright © StatSoft, 1995

StatSoft

Обратите внимание, что “горячие” клавиши можно использовать вне зависимости от их отображения.

ИЗМЕНЕНИЕ *КНОПОК АВТОЗАДАЧ*

Это диалоговое окно предназначено для создания новых *кнопок автозадач* (см. *Настройка кнопок автозадач*, стр. 1298) или изменения существующих. Здесь можно поставить в соответствие ранее сохраненный файл (например, с программой на языке *SCL* или *STATISTICA BASIC*, с данными, таблицей результатов, графиком или отчетом) *кнопке автозадач*.



Программы на языке *STATISTICA BASIC* и *SCL* могут также быть изменены в этом диалоговом окне (см. опцию *Изменить программу* ниже).

Быстрый вызов. Это диалоговое окно также доступно по двойному щелчку на названии соответствующей кнопки в списке *кнопок автозадач* или по щелчку на соответствующей *кнопке автозадач* на панели инструментов при нажатой клавише CTRL.

Кнопка

Здесь можно изменить или ввести название *кнопки автозадач* (до 30 символов).

Тип

Здесь показан тип выделенной *кнопки автозадач*. В этом диалоговом окне можно вводить или изменять название кнопки, описание, “горячие” клавиши и другие параметры для следующих трех типов *Кнопок автозадач*:

- язык *SCL* (Командный язык системы *STATISTICA*);
- файлы программ на языке *STATISTICA BASIC*;
- быстро вызываемые файлы (например, данные, таблицы результатов, графики, текст в формате *RTF* и т.п.).

Описание

В этом поле ввода можно изменить описание кнопки (до 128 символов). Введенная здесь запись отображается в области описания диалогового окна *Настройка кнопки автозадач* (см. стр. 1298), если выделено название соответствующей кнопки.

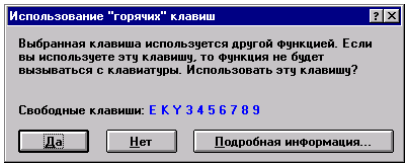
Имя файла

Чтобы изменить файл, связанный с этой *кнопкой автозадач,* нужно нажать на кнопку *Имя файла*.

Клавиша

В случае необходимости *кнопке автозадач* могут быть присвоены “горячие” клавиши. Если здесь указать занятые “горячие” клавиши [т.е., системные (см. ниже) или уже присвоенные другим *кнопкам автозадач*], система *STATISTICA* обнаружит это

StatSoft
Copyright © StatSoft, 1995



и предоставит возможность либо изменить “горячие” клавиши, либо заменить текущее присвоение.

Глобальные системные “горячие” клавиши (с CTRL)

Следующие “горячие” клавиши (комбинации клавиш CTRL+символ) являются глобальными и доступны в каждом окне (таблицы исходных данных, таблицы результатов, графика и т.д.) системы *STATISTICA*.

Открыть, Сохранить, Печать, Создать данные	CTRL+O, S, P, N
Буфер обмена	CTRL+C, V, X
Отменить	CTRL+Z
Стартовая панель, Продолжить анализ	CTRL+T, R
Кнопки автозадач (вкл/выкл)	CTRL+M
Выделить все (таблицу, текст) или объект (графики)	CTRL+A
Закрыть все окна вывода	CTRL+L

Локальные системные “горячие” клавиши (с CTRL)

Следующие “горячие” клавиши (комбинации клавиш CTRL+символ) являются локальными и доступны в определенных окнах системы *STATISTICA*.

Найти, Заменить (данные, текст)	CTRL+F, H
Параметры шрифта (в текстовых редакторах)	CTRL+0 (zero), B, I, U, D
Направляющая сетка графика (вкл./выкл.)	CTRL+G

Подробный список глобальных и локальных “горячих” клавиш см. в разделе *Клавиатурный интерфейс* (глава 1).

Изменить программу

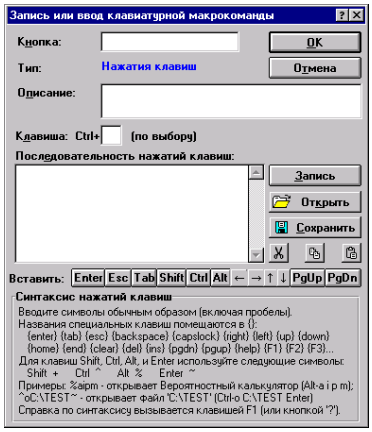
Нажав на эту кнопку, можно изменить файл программы на языке *SCL* или *STATISTICA BASIC* (эта опция затенена, если текущей кнопке присвоено быстрое открытие файла). Эта кнопка открывает соответствующее *Окно текста/вывода* с изменяемым файлом (см. главу 5).

ЗАПИСЬ ИЛИ ВВОД КЛАВИАТУРНОЙ МАКРОКОМАНДЫ

Это диалоговое окно позволяет вводить, записывать, изменять и сохранять (в текстовом файле) последовательность нажатий клавиш (изменяемая макрокоманда), которую позже можно исполнить, нажав на соответствующую кнопку панели инструментов *Кнопки автозадач*. Клавиатурная макрокоманда является изменяемой, она либо вводится непосредственно (с помощью стандартных соглашений инструкции Microsoft *Sendkeys*), либо записывается.

Также здесь можно открыть для изменения сохраненный ранее файл макрокоманды с записанными действиями (расширение **.stm*, см. стр. 1310) или текстовый файл, содержащий записанную последовательность нажатий клавиш.

Copyright © StatSoft, 1995
StatSoft®



В этом диалоговом окне можно вводить или записывать макрокоманды, нажав на кнопку *Клавиатурная макрокоманда*. Для изменения существующей клавиатурной макрокоманды нужно выделить название соответствующей кнопки и нажать кнопку *Изменить/задать* в диалоговом окне *Настройка кнопок автозадач*. Это диалоговое окно также можно использовать для того, чтобы открыть для изменения сохраненную ранее введенную или записанную макрокоманду.

Быстрый вызов. Это диалоговое окно также вызывается двойным щелчком на названии нужной кнопки в списке *кнопок автозадач* или щелчком на нужной кнопке на панели инструментов *Кнопки автозадач* при нажатой клавише CTRL.

Название/Тип кнопки

Поле ввода *Кнопка* позволяет ввести или изменить (до 30 символов) название этой *Кнопки автозадач*. Непосредственно под этим полем отображается тип *кнопки автозадач*.

Описание

В области *Описание* этого диалогового окна можно ввести описание кнопки (до 128 символов). Оно будет показано в области описания диалогового окна *Настройка кнопок автозадач* (см. стр. 1298), когда выделена соответствующая кнопка.

Клавиша

При необходимости *кнопке автозадач* могут быть поставлены в соответствие “горячие” клавиши. Если здесь указать уже присвоенные “горячие” клавиши (т.е., системные или соответствующие другим *кнопкам автозадач*), система *STATISTICA* обнаружит это и предоставит возможность либо изменить “горячие” клавиши, либо заменить текущее присвоение.

Список глобальных и локальных “горячих” клавиш см. на стр. 1303 и 1303 соответственно.

Последовательность нажатий клавиш

В этой области диалогового окна можно вводить, записывать или изменять последовательность нажатий клавиш (изменяемую макрокоманду), которая будет выполняться при нажатии на соответствующую кнопку панели инструментов *Кнопки автозадач*. Для представления нажатий клавиш в макрокомандах используются соглашения инструкции Microsoft *Sendkeys*. Нажатия клавиш могут быть введены непосредственно или записаны с помощью кнопки *Запись*. Оба этих метода описаны ниже.

StatSoft
Copyright © StatSoft, 1995

Ввод/изменение последовательности нажатий клавиш

В этом поле ввода можно непосредственно ввести или изменить нажатия клавиш, используя синтаксис инструкции *Sendkeys*. Для быстрого добавления часто используемых нажатий клавиш можно использовать некоторые *Кнопки быстрого ввода* (см. ниже). В этом диалоговом окне приведено краткое описание синтаксиса клавиш. Подробное его описание (по соглашениям инструкции Microsoft *Sendkeys*) можно найти в разделе *Описание синтаксиса клавиатурных макрокоманд* на стр. 1306.

Обратите внимание, что при изменении последовательности нажатий клавиш в этом поле ввода все дополнительные нажатия клавиш будут вставлены с текущей позиции курсора. Если выделить ряд нажатий клавиш, а затем ввести, записать или вставить из файла дополнительные нажатия клавиш, то выделенные элементы будут заменены вставленными.

Запись

Вместо того, чтобы непосредственно вводить нажатия клавиш, можно нажать на кнопку *Запись* для записи последовательности нажатий клавиш (чтобы записать движения мыши, можно воспользоваться опцией *Макрокоманда с записью действий* в диалоговом окне *Настройка кнопок автозадач*, см. стр. 1310). После нажатия на эту кнопку на *Строке состояния* замигает надпись *ЗАП*, означающая, что идет процесс записи.

Чтобы остановить запись, нужно нажать клавиши CTRL+F3. Записанная таким образом макрокоманда может быть изменена описанным выше способом (нажатия клавиш будут отображаться в поле ввода). Обратите внимание, что если запись производится для того, чтобы в будущем можно было изменить последовательность нажатий клавиш, то во время записи не стоит пользоваться мышью (так как записанные движения мыши не могут быть изменены).

Кнопки быстрого ввода

Кнопки быстрого ввода в диалоговом окне *Запись или ввод клавиатурной макрокоманды* могут использоваться для быстрого ввода в макрокоманду соответствующих нажатий клавиш. Ниже рядом с соответствующим символом клавиатуры показаны кнопки быстрого ввода и коды, используемые для представления определенных символов (обратите внимание, что фигурные скобки **{}** являются частью кода).

Кнопки быстрого ввода	Символы клавиатуры	Код
Alt	ALT	**%**
Ctrl	CTRL	**^**
Shift	SHIFT	**+**
Enter	ENTER	**~** или **{Enter}**
Esc	ESCape	**{Esc}**
PgDn	PAGE DOWN	**{Pgdn}**
PgUp	PAGE UP	**{Pgup}**
Tab	TAB	**{Tab}**
↓	Вниз	**{Down}**
↑	Вверх	**{Up}**
←	Влево	**{Left}**
→	Вправо	**{Right}**

Copyright © StatSoft, 1995

Открыть

Эта кнопка открывает диалоговое окно *Открытие файла кнопок автозадач,* в котором можно выбрать сохраненный ранее текстовый файл (**.txt*) или файл записанной макрокоманды (**.stm*), содержащий последовательность нажатий клавиш (изменяемую макрокоманду). Если открыть файл *Макрокоманды с записью действий*, то в редакторе будут отображаться только нажатия клавиш (движения мыши будут опущены).

Сохранить

Эта кнопка открывает диалоговое окно *Сохранить файл кнопок автозадач,* в котором можно сохранить последовательность нажатия клавиш (изменяемую макрокоманду), записав ее в текстовый файл (**.txt*).

Операции с буфером обмена (вырезать, копировать, вставить)

В этом редакторе для операций с буфером обмена можно использовать кнопки вырезать, копировать и вставить.

Описание синтаксиса клавиатурных макрокоманд

Описанный здесь синтаксис следует тем же правилам, которые используются в других приложениях Windows (например, Excel, Word для Windows). Обратите внимание, что для быстрого ввода некоторых часто используемых нажатий клавиш можно использовать *Кнопки быстрого ввода* (см. стр. 1305).

Ввод обычных символов

Каждая клавиша клавиатуры представляется одним или несколькими символами. Чтобы ввести один символ клавиатуры, нужно нажать этот символ. Например, чтобы ввести текст “ABC”, надо набрать **`ABC`** в поле ввода. Символы клавиатуры **`+`** (знак “плюс”), **`^`** (“крышка”), **`%`** (процент), **`~`** (тильда), и **`()`** (скобки) используются для быстрого вызова определенных клавиш (см. ниже). Чтобы использовать эти символы как части последовательностей нажатий символов, нужно заключить символ в фигурные скобки **`{}`**. Например, чтобы ввести знак плюс как часть текста “ABC”, надо набрать **`ABC{+}`**. Чтобы использовать фигурные скобки (**`{}`**) как часть последовательности нажатий, их тоже надо заключить в фигурные скобки (т.е., использовать **`{{}`** и **`{}}`**).

Ввод специальных символов

Синтаксис клавиатурных макрокоманд позволяет вводить символы, которые не отображаются при нажатии на них (например, ENTER или TAB) и клавиши, которые представляют собой действия, а не символы (например, ВВЕРХ или ВПРАВО). Для таких клавиш нужно использовать код *инструкции Sendkeys,* приведенный ниже (обратите внимание, что фигурные скобки являются частью кода).

Символ клавиатуры	Используемый код:
ALT	**`%`**
BACKSPACE	**`{Backspace}`**
BREAK	**`{Break}`**
CAPS LOCK	**`{Capslock}`**
Clear	**`{Clear}`**
CTRL	**`^`**
DELETE	**`{Del}`**

StatSoft®

Copyright © StatSoft, 1995

Двойные кавычки	**" "**
Вниз	**{Down}**
END	**{End}**
ENTER	**~** или **{Enter}**
ESCape	**{Esc}**
F1, ..., F12	**{F**номер**}** (например, **{F11}**)
Help	**{Help}**
HOME	**{Home}**
INSERT	**{Insert}**
Символ клавиатуры	нажмите эту клавишу (например, **abc@!**)
Влево	**{Left}**
PAGE DOWN	**{Pgdn}**
PAGE UP	**{Pgup}**
PRINT SCREEN	**{Prtsc}**
Вправо	**{Right}**
SHIFT	**+**
Одиночные кавычки	**' '**
TAB	**{Tab}**
Вверх	**{Up}**

Порядок нажатий клавиш

Нажатия клавиш будут обрабатываться в том порядке, в котором они вводятся в редактор. Например, если ввести нажатия клавиш **+EC**, то система *STATISTICA* сначала "нажмет" клавишу SHIFT, а затем буквы *E* и *C*. Для задания порядка нажатий клавиш можно использовать скобки, квадратные скобки и фигурные скобки.

Например, если нужно удерживать нажатой клавишу CTRL и нажимать клавиши SHIFT и C, то эта "строка" указывается так: **^(+C)**. Чтобы удерживать клавиши SHIFT *и* CTRL, нажимая при этом две другие клавиши, нужно задать строку так: **+(^(AB))**.

Повторяющиеся нажатия клавиш

Можно сделать так, чтобы система *STATISTICA* повторяла определенные последовательности нажатий клавиш нужное число раз. Для этого используется формат **{**клавиша число**}**. Например, **{LEFT 42}** означает нажатие левой стрелки 42 раза; **{h 10}** означает нажатие клавиши *H* 10 раз. (Замечание: между ключевым словом и числом обязательно должен стоять пробел.)

Команда WAIT

Можно воспользоваться командой **{WAIT}** для того, чтобы приостановить выполнение клавиатурной макрокоманды на указанное число миллисекунд (1 секунда = 1000 миллисекунд).

Например, если ввести **{WAIT 15000}** , то макрокоманда "подождет" 15 секунд, а потом продолжит работу. Если значение не указано, макрокоманда по умолчанию будет "ждать" 1 секунду.

Советы по вводу клавиатурных макрокоманд

При вводе клавиатурных макрокоманд могут оказаться полезными следующие советы (см. далее пример, иллюстрирующий большую часть этих советов).

- Если нужно закрыть диалоговое окно, в начале макрокоманды можно поставить **{Esc}**. (Обратите внимание, что для работы клавиатурной макрокоманды из открытого окна эту комбинацию включать в макрокоманду не нужно.)

Copyright © StatSoft, 1995

- Чтобы сделать текущим окно таблицы исходных данных или другое окно, можно воспользоваться выпадающим меню *Окно*.
- Чтобы выбрать опцию в диалоговом окне, можно воспользоваться подчеркнутыми буквами (т.е., так называемыми ускорителями), вместо того, чтобы "перемещаться" вниз, вправо и т.д.
- В списках можно использовать **{Home},** для перевода курсора на начало списка, а затем указать число переходов вниз **{Down},** чтобы выбрать нужную опцию.
- Можно использовать **%o** (ALT+O) для того, чтобы подтвердить и закрыть диалоговое окно (т.е. нажать *OK*).

Пример: Ввод последовательности нажатий клавиш в макрокоманду

Для получения двумерной диаграммы рассеяния можно ввести перечисленные ниже нажатия клавиш в диалоговом окне *Запись или ввод клавиатурной макрокоманды* (стр. 1303), причем для ввода некоторых из них можно воспользоваться *Кнопками быстрого ввода* (см. стр. 1305).

Обратите внимание, что часто наиболее простым методом создания макрокоманды оказывается ее запись (с помощью кнопки *Запись* в диалоговом окне *Запись или ввод клавиатурной макрокоманды*). См. полезную информацию о вводе макрокоманд выше в разделе *Советы*.

Шаг 1. Ввод показанной ниже последовательности в области *Последовательность нажатий клавиш* диалогового окна *Запись или ввод клавиатурной макрокоманды*:

```
{ESC 3}%w1%g2{DOWN}~v1{TAB}2%of{HOME}
{DOWN 5}%0
```

где

{ESC 3}	Нажатие клавиши ESC в начале макрокоманды закрывает открытые диалоговые окна.
%w1	ALT+W, а затем 1 открывает выпадающее меню *Окно* и выбирает текущий файл данных (чтобы поместить таблицу исходных данных на передний план; обратите внимание, что таблица исходных данных - это всегда первое окно в этом меню).
%g	ALT+G открывает меню *Графика*.
2	Выбирает из этого меню опцию *Статистические 2М графики*.
{DOWN}	Перемещается на 1 строку ВНИЗ и выбирает пункт *2М диаграммы рассеяния*.
~	Нажимает клавишу ENTER, чтобы открыть диалоговое окно *2М диаграммы рассеяния*.
v	Открывает диалоговое окно *Переменные* для выбора переменных для построения графика.
1	Выбирает переменную 1, вводя *1* в поле выбора переменной в этом диалоговом окне.
{TAB}	Перемещается в другой список клавишей TAB.
2	Выбирает переменную 2, вводя *2* поле выбора переменной в этом диалоговом окне.
%o	Нажимает ALT+O, чтобы закрыть это диалоговое окно (т.е. нажимает *OK*) и возвращается в

StatSoft®

Copyright © StatSoft, 1995

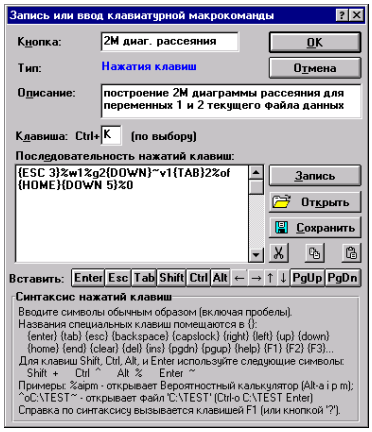
диалоговое окно *2M диаграммы рассеяния*.

f{HOME} Перемещает курсор в область *Подгонка* диалогового окна и ставит курсор в начало списка,

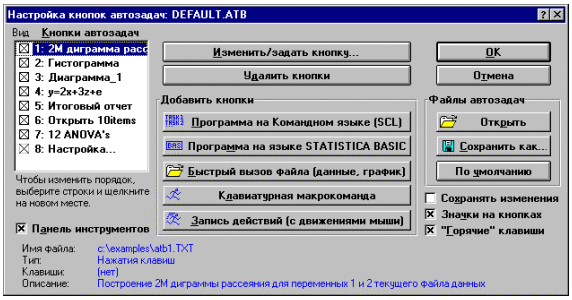
{DOWN 5} перемещается на 5 строк ВНИЗ и выбирает *Полиномиальную* подгонку.

%o Нажимает ALT+O, чтобы закрыть это диалоговое окно (т.е. нажимает *OK*), и строит график.

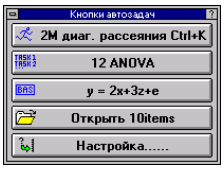
Шаг 2. Если для этой кнопки еще не введено название и описание, нужно их ввести и сохранить эти нажатия клавиш в текстовом файле (с расширением **.txt*).



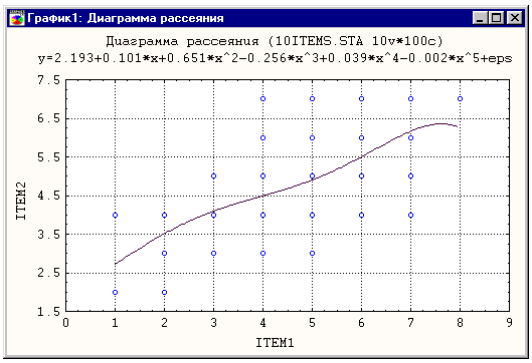
Шаг 3. Нажатие *OK* для возвращения в диалоговое окно *Настройка кнопок автозадач* (стр. 1298), где название кнопки будет добавлено в список *Кнопок автозадач*.



Нажмите *OK* в этом диалоговом окне, чтобы кнопка появилась на панели инструментов *Кнопки автозадач*.



Теперь чтобы построить график (например, используя файл данных *10items.sta*), достаточно нажать эту кнопку.

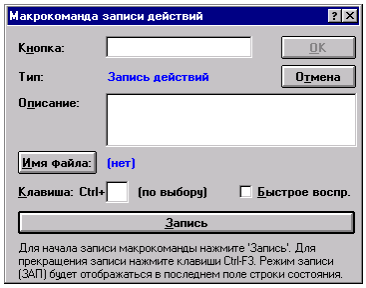


Как было упомянуто ранее, запись нажатий клавиш является, как правило, наиболее эффективным способом создания клавиатурных макрокоманд, а редактор обычно используется для их дополнения или изменения (например, для их объединения).

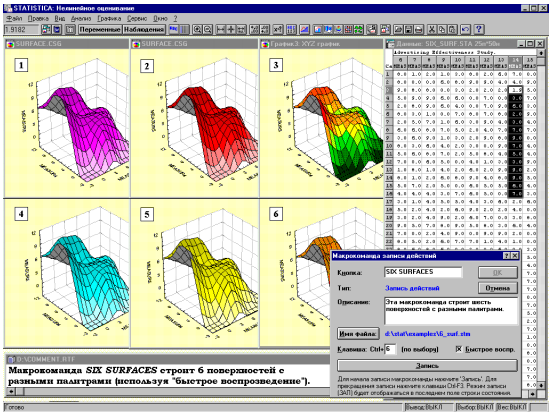
Copyright © StatSoft, 1995
StatSoft

МАКРОКОМАНДА ЗАПИСИ ДЕЙСТВИЙ

Это диалоговое окно доступно по нажатию кнопки *Запись действий (с движениями мыши)* (для записи новой макрокоманды) или кнопки *Изменить/задать* (если выделено название *Кнопки автозадач* данного типа) в диалоговом окне *Настройка кнопок автозадач* (см. стр. 1298).



Здесь можно указать название кнопки, описание, и, в качестве необязательного параметра, “горячие” клавиши для макрокоманды. Этот тип макрокоманды может сохранять движения мыши (которые можно воспроизвести с той же скоростью, с которой они были сохранены, или в ускоренном режиме) и нажатия клавиш (которые позднее могут быть изменены в диалоговом окне *Запись или ввод клавиатурной макрокоманды*, стр. 1303). Однако обратите внимание, что макрокоманды, содержащие движения мыши, “контекстно-зависимы” и они могут привести к другим результатам, если размер окна или его положения на экране изменились. Поэтому для воспроизведения там, где это возможно, лучше использовать макрокоманды с нажатиями клавиш (все меню и диалоговые окна системы *STATISTICA* имеют свои клавиатурные эквиваленты).



Быстрый вызов. Это диалоговое окно также доступно по двойному щелчку на названии кнопки в списке *кнопок автозадач* или с помощью нажатия на *кнопку автозадач* панели инструментов при нажатой клавише CTRL.

Название/тип кнопки

Поле ввода *Кнопки* позволяет вводить или изменять (до 30 символов) имя этой *кнопки автозадач*. Под этим полем отображается тип соответствующей кнопки.

Описание

В этом поле можно ввести или изменить описание *Кнопки автозадач* (до 128 символов). Оно отображается в области описания диалогового окна *Настройка кнопок автозадач* (стр. 1298), если выделено название соответствующей кнопки.

Имя файла

Эта кнопка открывает диалоговое окно *Добавить макрокоманду с записью действий,* в котором можно выбрать имя файла (**.stm*) для сохранения макрокоманды.

StatSoft
Copyright © StatSoft, 1995

Клавиша

При необходимости *кнопке автозадач* можно поставить в соответствие “горячие” клавиши. Если указать здесь уже занятые “горячие” клавиши (т.е. системные или присвоенные другим *кнопкам автозадач*), система *STATISTICA* обнаружит это и предоставит возможность изменить “горячие” клавиши или текущее присвоение. См. список глобальных и локальных “горячих” клавиш на стр. 1303 и 1303 соответственно. Полный список системных “горячих” клавиш см. в главе 1.

Быстрое воспроизведение

Здесь можно задавать скорость воспроизведения макрокоманд. Если эта опция выбрана, макрокоманда будет воспроизводиться очень быстро (без записанных задержек); в противном случае она будет воспроизводиться с той же скоростью, с которой она была записана. Обратите внимание, что *Макрокоманды записи действий,* которые выполняются с той скоростью, с какой они были записаны (режим *Быстрое воспроизведение* не выбран), могут служить для демонстраций.

Прервать воспроизведение. Чтобы остановить воспроизведение макрокоманды, можно нажать кнопку мыши или клавиши ESC или CTRL+BREAK.

Запись

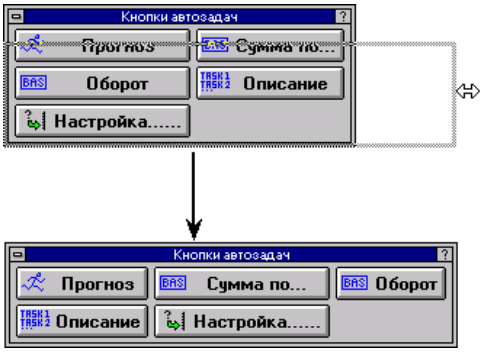
После нажатия на кнопку *Запись* программа начнет записывать движения мыши (и нажатия клавиш). Во время записи в последнем поле *Строки состояния* будет мигать надпись *ЗАП*.

Чтобы остановить запись, нужно нажать клавиши CTRL+F3.

Обратите внимание, что макрокоманды, содержащие движения мыши, “контекстно-зависимы” и могут привести к другим результатам, если размер окна или его положения на экране изменились. Для воспроизведения там, где это возможно, лучше использовать макрокоманды с нажатиями клавиш (см. пункт *Запись или ввод клавиатурной макрокоманды* на стр. 1303). Заметим, что все меню и диалоговые окна системы *STATISTICA* имеют свои клавиатурные эквиваленты.

РАСПОЛОЖЕНИЕ ПАНЕЛИ ИНСТРУМЕНТОВ *КНОПКИ АВТОЗАДАЧ*

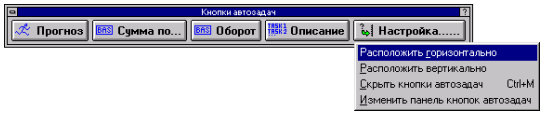
Следующие опции контекстного меню позволяют изменять отображение кнопок на панели инструментов *Кнопки автозадач*. Кроме того, можно изменить формат отображения, перетаскивая границы панели инструментов



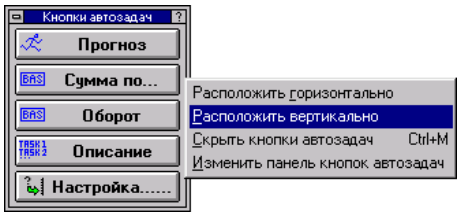
или дважды щелкая на заголовке панели инструментов *Кнопки автозадач*.

Расположить горизонтально. Эта опция контекстного меню располагает *кнопки автозадач* на панели инструментов горизонтально.

Copyright © StatSoft, 1995
StatSoft

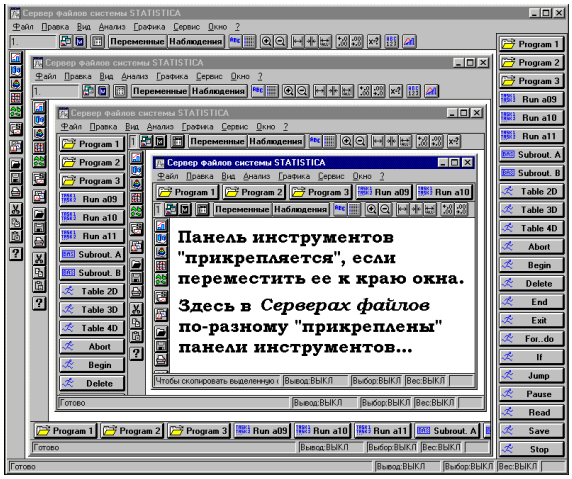


Расположить вертикально. Эта опция контекстного меню располагает *кнопки автозадач* на панели инструментов горизонтально.



Закрепить панель инструментов *Кнопки автозадач*

Панель инструментов можно “закрепить”, перетащив ее к границе окна системы *STATISTICA* (над панелью инструментов, под строкой состояния или с любого бокового края окна).



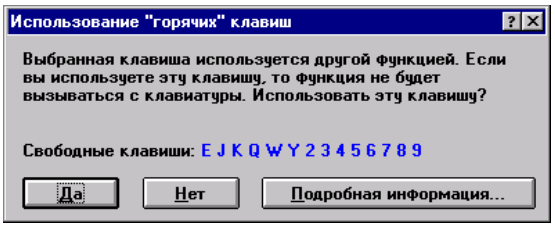
Здесь эта панель инструментов изменит свой вид, и будет оставаться такой до тех пор, пока не будет перенесена в другую область окна.

ИЗМЕНИТЬ ПАНЕЛЬ ИНСТРУМЕНТОВ *КНОПКИ АВТОЗАДАЧ*

Если выбрать эту опцию из контекстного меню, откроется диалоговое окно *Открытие файла кнопок автозадач* (с расширением **.atb*), в котором записана другая панель инструментов *Кнопки автозадач*. См. также раздел *Файлы автозадач* на стр. 1300.

ИСПОЛЬЗОВАНИЕ “ГОРЯЧИХ” КЛАВИШ

Это диалоговое окно открывается, когда выбранные “горячие” клавиши уже присвоены системным операциям (см. ниже) или другой *кнопке автозадач*.



Чтобы переназначить указанные “горячие” клавиши на текущую *кнопку автозадач,* нужно нажать *Да* (предыдущее присвоение будет при этом отменено). Чтобы не отменять существующее присвоение, нужно нажать *Нет – No,* в этом случае откроется предыдущее диалоговое окно, в котором можно изменить “горячие” клавиши. Список глобальных и локальных “горячих” клавиш см. на стр. 1303 и 1303 соответственно. Полный список системных “горячих” клавиш можно найти в главе 1.

StatSoft
Copyright © StatSoft, 1995

ПРЕДМЕТНЫЙ УКАЗАТЕЛЬ

А

Автозадачи
обзор кнопок, 1296

Б

Библиотеки кнопок автозадач, 1297
Быстрый вызов файла, 1300

В

Воспроизведение записанной макрокоманды с записью действий, 1311
Восстановить значения по умолчанию, 1301

Г

Глобальные 'горячие' клавиши, 1303
Горячие клавиши
отображать на кнопках автозадач, 1301
присвоение кнопкам автозадач, 1302

'

'Горячие' клавиши
глобальные, 1303
использование 'горячих' клавиш, 1312
локальные, 1303

Д

Движения мыши в макрокомандах, 1310
Добавить Кнопки автозадач, 1299

З

Задать кнопки автозадач, 1299
Закрепить панель инструментов Кнопки автозадач, 1312
Запись действий в макрокомандах
'горячие' клавиши, 1311
быстрое воспроизведение, 1311
Значки
на кнопках автозадач, 1301

И

Изменение
имя файла кнопки автозадач, 1302
названия кнопок автозадач, 1302
описание кнопок автозадач, 1302
параметры кнопок автозадач, 1302
программа, 1303
тип кнопки автозадач, 1302
Изменение кнопок автозадач, 1298
Изменение порядка кнопок автозадач, 1298
Инструкция Sendkeys, 1304
синтаксис, 1306

К

Клавиатурные макрокоманды, 1300
ввод обычных символов, 1306
ввод специальных символов, 1306
ввод/изменение, 1303
запись, 1303
кнопки быстрого ввода, 1305
команда wait, 1307
описание синтаксиса, 1306
повторяющиеся нажатия клавиш, 1307
порядок нажатий клавиш, 1307
пример, 1308
советы по вводу, 1307
Кнопки автозадач, 1296
библиотеки, 1297
быстрое воспроизведение, 1311
ввод обычных символов, 1306
ввод специальных символов, 1306
восстановить значения по умолчанию, 1301
вызов файла, 1300
глобальные 'горячие' клавиши, 1303
глобальные и локальные, 1301
горячие клавиши, 1301
добавить кнопки, 1299
задать кнопки, 1299
закрепить панель инструментов, 1312
запись движений мыши, 1311
запись клавиатурной макрокоманды, 1303
запись нажатий клавиш, 1305
значки, 1301
изменение, 1298, 1302
изменение кнопки, 1298
изменить имя файла, 1302
изменить кнопки, 1299
изменить название, 1302
изменить описание кнопок, 1302
изменить присвоение 'горячих' клавиш, 1302
изменить программу, 1303
изменить тип, 1302
использование 'горячих' клавиш, 1312
клавиатурные макрокоманды, 1300, 1303
кнопки быстрого ввода, 1305
команда wait, 1307
командный язык, 1299
конфигурация панели инструментов, 1311
локальные 'горячие' клавиши, 1303
макрокоманда последовательности нажатий клавиш, пример, 1308
макрокоманды с движениями мыши, 1300
макрокоманды с записью действий, 1300, 1310
операции с буфером обмена, 1306
открыть файл, 1300
отображать панель инструментов, 1299
повторяющиеся нажатия клавиш, 1307
порядок кнопок, 1298
порядок нажатий клавиш, 1307
последовательность нажатий клавиш, 1304
расположить вертикально, 1312
расположить горизонтально, 1311
синтаксис инструкции Sendkeys, 1306

Copyright © StatSoft, 1995

советы по вводу клавиатурных макрокоманд, 1307
сохранение изменений, 1301
сохранить, 1301
список, 1298
удалить кнопки, 1299
файлы автозадач, 1300
язык SCL, 1299
язык STATISTICA BASIC, 1299
Кнопки быстрого ввода, 1305
Командный язык
присвоение Кнопке автозадач, 1299
Конфигурация панели инструментов Кнопки автозадач, 1311

Л

Локальные 'горячие' клавиши, 1303

М

Макрокоманда записи действий, 1310
Макрокоманда с движениями мыши, 1300
быстрое воспроизведение, 1311
горячие клавиши, 1311
запись, 1311
Макрокоманда с записью движений, 1311
Макрокоманда с записью действий, 1300
Макрокоманды
движения мыши, 1300, 1310
запись действий, 1300
клавиатурные, 1300, 1306, 1308
командный язык, 1299
объединение, 1309
советы по вводу клавиатурных макрокоманд, 1307
язык STATISTICA BASIC, 1299

О

Обзор кнопок автозадач, 1296
Объединение клавиатурных макрокоманд, 1309
Отображение 'горячих' клавиш на кнопках автозадач, 1301
Отображение панели инструментов Кнопки автозадач, 1299

П

Панель инструментов Кнопки автозадач
закрепить панель инструментов, 1312
изменить панель инструментов, 1312
конфигурация, 1311
расположить вертикально, 1312
расположить горизонтально, 1311
Пользовательские дополнения к системе *STATISTICA*, 1297
Присвоение кнопкам автозадач ‘горячих’ клавиш, 1302
Программа на языке SCL
присвоение кнопке автозадач, 1299

С

Сохранение изменений, 1301

У

Удалить Кнопки автозадач, 1299

Я

Язык STATISTICA BASIC
присвоение кнопке автозадач, 1299

StatSoft®
Copyright © StatSoft, 1995

Глава 7:
УПРАВЛЕНИЕ ДАННЫМИ

Оглавление

Вводный обзор 1322
Обзор программных модулей 1336
Что такое *Менеджер мегафайлов*? 1339
Формат файлов данных системы *STATISTICA* 1343
Общий формат 1343
Спецификации переменных 1345
Формат матричных файлов данных системы *STATISTICA* 1359
Пример 1362
Диалоговые окна и меню 1370
Стартовая панель 1370
Создание нового файла данных 1370
Объединение двух файлов данных 1372
Объединение переменных 1373
Объединение наблюдений 1375
Объединение текстовых значений 1375
Подмножество 1376
Сортировка 1376
Изменение переменных 1378
Изменение наблюдений 1378
Проверка имен и формата переменных 1378
Повторяющиеся имена 1379
Недопустимое имя переменной 1380
Проверка данных (условия) 1380
Проверка данных 1383
Стандартизация значений 1383
Замена пропущенных данных 1385
Средства *Менеджера мегафайлов* модуля управление данными 1386
Создание новых файлов данных MM 1386
Открытие данных *Менеджера мегафайлов* 1388
Преобразование файлов *STATISTICA* в формат MM 1388
Импорт и экспорт данных 1389
Импорт данных 1389
Экспорт данных 1408
Предметный указатель 1421

Подробное оглавление приведено на следующей странице.

Copyright © StatSoft, 1995

Подробное оглавление

ВВОДНЫЙ ОБЗОР 1322
Общая организация файлов данных в системе *STATISTICA* 1322
Значения переменных 1323
Числовые и текстовые (алфавитно-цифровые) значения 1325
Способ представления *дат* в файлах данных системы *STATISTICA* 1326
Способ представления *времени* в файлах данных системы *STATISTICA* 1327
Использование значений *дат* в качестве кодов (например, значений группирующих переменных) 1328
Прямое и обратное преобразование значений *дат* в текст 1328
Создание файлов данных и ввод данных 1328
Импорт данных из других приложений 1328
Экспорт данных в другие приложения 1329
Средства динамического обмена данными (*DDE*) 1329
Проверка данных 1330
Редактирование данных и изменение их структуры 1331
Операции над таблицей исходных данных с использованием буфера обмена 1332
Преобразование данных 1332
Перекодировка данных 1333
Разделение и объединение файлов данных 1335

ОБЗОР ПРОГРАММНЫХ МОДУЛЕЙ 1336
Модуль *Управление данными* 1336
Обычные свойства таблиц исходных данных и модуля *Управление данными* 1338
Управление данными с помощью *Командного языка* системы *STATISTICA* 1339
Операции управления данными и пользовательские процедуры на языке *STATISTICA BASIC* 1339
Что такое *Менеджер мегафайлов*? 1339
Обзор 1339
Менеджер мегафайлов и другие средства системы *STATISTICA* 1340
Уникальные возможности *Менеджера мегафайлов* 1341

ФОРМАТ ФАЙЛОВ ДАННЫХ СИСТЕМЫ *STATISTICA* 1343
Общий формат 1343
Переменные и наблюдения 1343
Заголовок файла и комментарий 1343
Список файлов Рабочей книги 1343
Имена наблюдений 1344
Код пропущенных данных 1344
Спецификации переменных 1345
Имена переменных 1345
Формат отображения: Число 1345
Формат отображения: Дата 1346

StatSoft®

Copyright © StatSoft, 1995

Формат отображения: Время ... 1347
Формат отображения: Научный ... 1348
Формат отображения: Денежный ... 1348
Формат отображения: Проценты ... 1349
Длинные имена переменных, формулы в таблице исходных данных и связи динамического обмена данными (*DDE*) ... 1349
Числовые и текстовые значения ... 1354
Метки значений ... 1356
Другие файлы данных *STATISTICA* ... 1358
Формат матричных файлов данных системы *STATISTICA* ... 1359
Типы матриц ... 1360
Сохранение матричного файла ... 1361
Файлы, содержащие множественные корреляционные матрицы ... 1361

ПРИМЕР ... 1362
Обзор ... 1362
Запуск модуля *Управление данными* ... 1362
Создание нового файла данных ... 1362
Использование языка программирования *STATISTICA BASIC* для преобразования данных ... 1367

ДИАЛОГОВЫЕ ОКНА И МЕНЮ ... 1370
Стартовая панель ... 1370
Создание нового файла данных ... 1370
Имя нового файла ... 1370
Число переменных ... 1371
Число наблюдений ... 1371
Длина имен наблюдений ... 1371
Формат значений ... 1371
Код пропущенных данных ... 1371
Префикс имен переменных ... 1371
Начальный номер имени переменной ... 1371
Строка заголовка файла данных ... 1371
Взять из… ... 1372
Объединение двух файлов данных ... 1372
Объединить ... 1373
Второй файл ... 1373
Объединение переменных ... 1373
Режим ... 1373
Параметры реляционного объединения ... 1374
При несовместимости наблюдений ... 1375
Объединение наблюдений ... 1375
Объединение текстовых значений ... 1375
Режим ... 1376
Подмножество ... 1376
Переменные ... 1376

Copyright © StatSoft, 1995

Наблюдения ... 1376
Сортировка ... 1376
Переменная ... 1377
Имя наблюдения ... 1377
Порядок сортировки ... 1377
Режим сортировки ... 1377
Дополнительно ... 1377
Изменение переменных ... 1378
Изменение наблюдений ... 1378
Проверка имен и формата переменных ... 1378
Имена переменных ... 1378
Создание стандартных имен ... 1379
Форматы ... 1379
Повторяющиеся имена ... 1379
Переменная 2 – Переменная 1 ... 1380
Принять ... 1380
Прервать ... 1380
Отменить все ... 1380
Недопустимое имя переменной ... 1380
Переменная ... 1380
Пропустить ... 1380
Принять ... 1380
Прервать ... 1380
Отменить все ... 1380
Проверка данных (условия) ... 1380
Наблюдение считается допустимым, если: ... 1381
Условие ... 1381
Верно/Неверно, если ... 1382
Открыть/сохранить ... 1382
Поле ввода ... 1382
Просмотр переменных ... 1382
Открыть все ... 1382
Сохранить все ... 1382
Диапазон ... 1383
Проверка данных ... 1383
Просмотр/изменение данных ... 1383
Пропустить и продолжить ... 1383
Пропустить и завершить ... 1383
Стандартизация значений ... 1383
Переменные ... 1384
Наблюдения ... 1384
Вес ... 1384
Замена пропущенных данных ... 1385
Переменные ... 1385

StatSoft®
Copyright © StatSoft, 1995

Наблюдения ... 1385
Вес ... 1385

СРЕДСТВА *МЕНЕДЖЕРА МЕГАФАЙЛОВ* МОДУЛЯ УПРАВЛЕНИЕ ДАННЫМИ ... 1386
Создание новых файлов данных ММ ... 1386
Имя файла ... 1386
Число столбцов ... 1387
Число строк ... 1387
Префикс имен столбцов ... 1387
Типы и ширина столбцов ... 1387
Спецификации типов столбцов ... 1387
Строка заголовка файла ... 1388
Открытие данных *Менеджера мегафайлов* ... 1388
Преобразование файлов *STATISTICA* в формат ММ ... 1388

ИМПОРТ И ЭКСПОРТ ДАННЫХ ... 1389
Импорт данных ... 1389
Импорт графических файлов *STATISTICA* (STG) ... 1390
Импорт файлов, созданных *Менеджером мегафайлов* ... 1390
Импорт текстовых значений ... 1390
Импорт имен наблюдений ... 1391
Автоматическое согласование чисел и текста ... 1391
Диапазон импортируемых значений ... 1391
Импорт SPSS-файлов ... 1392
Импорт текстовых значений ... 1392
Импорт имен наблюдений ... 1392
Импорт меток значений ... 1392
Пропущенные значения ... 1392
Импорт файлов Excel ... 1393
Импорт текстовых значений ... 1393
Импорт имен наблюдений ... 1393
Автоматическое согласование чисел и текста ... 1393
Диапазон импортируемых значений ... 1394
Имена переменных ... 1394
Импорт файлов Lotus ... 1395
Импорт текстовых значений ... 1395
Импорт имен наблюдений ... 1395
Автоматическое согласование чисел и текста ... 1395
Диапазон импортируемых значений ... 1396
Имена переменных ... 1397
Импорт данных Lotus – Просмотр диапазонов ... 1397
Импорт файлов Quattro ... 1397
Импорт текстовых значений ... 1398
Импорт имен наблюдений ... 1398
Автоматическое согласование чисел и текста ... 1398

Copyright © StatSoft, 1995

Диапазон импортируемых значений ... 1399
Имена переменных ... 1399
Импорт файлов Paradox ... 1399
Импорт текстовых значений ... 1399
Импорт имен наблюдений ... 1400
Автоматическое согласование чисел и текста ... 1400
Диапазон импортируемых значений ... 1400
Импорт файлов dBASE ... 1400
Импорт текстовых значений ... 1401
Импорт имен наблюдений ... 1401
Автоматическое согласование чисел и текста ... 1401
Диапазон импортируемых значений ... 1402
Импорт ASCII-файлов ... 1402
Импорт файлов в свободном ASCII-формате ... 1402
Размер файла ... 1402
Импорт текстовых значений ... 1403
Импорт имен наблюдений ... 1403
Формат ... 1403
Спецификации типов столбцов ... 1403
Разделители ... 1404
Разделители в свободном ASCII-формате ... 1405
Импорт файлов в фиксированном ASCII-формате ... 1406
Размер файла ... 1406
Импорт текстовых значений ... 1406
Импорт имен наблюдений ... 1406
Формат ... 1407
Спецификации типов столбцов ... 1407
Экспорт данных ... 1408
Экспорт данных в файлы *Менеджера мегафайлов* ... 1408
Имена наблюдений ... 1409
Переменные с текстовыми значениями ... 1409
Экспорт файлов в Excel ... 1410
Имена наблюдений ... 1410
Имена переменных ... 1410
Переменные с текстовыми значениями ... 1410
Экспорт файлов в Lotus ... 1411
Листы (только для Lotus 1-2-3 Версия 3 и Lotus 1-2-3/w) ... 1411
Имена наблюдений ... 1412
Имена переменных ... 1412
Переменные с текстовыми значениями ... 1412
Экспорт файлов в Quattro ... 1413
Имена наблюдений ... 1413
Имена переменных ... 1413
Переменные с текстовыми значениями ... 1413

StatSoft®
Copyright © StatSoft, 1995

Экспорт файлов в dBASE ... 1414
Имена наблюдений ... 1414
Переменные с текстовыми значениями ... 1414
Замечания ... 1415
Экспорт файлов в Paradox ... 1415
Имена наблюдений ... 1415
Переменные с текстовыми значениями ... 1415
Экспорт файлов в SPSS ... 1416
Имена наблюдений ... 1417
Переменные с текстовыми значениями ... 1417
Экспорт файлов в свободном ASCII-формате ... 1417
Имена наблюдений ... 1418
Переменные с текстовыми значениями ... 1418
Длина строки ... 1418
Разделители ... 1418
Экспорт в свободный ASCII-формат - Разделители ... 1419
Экспорт файлов в фиксированном ASCII-формате ... 1419
Имена наблюдений ... 1419
Переменные с текстовыми значениями ... 1420
Длина строки ... 1420
Экспорт файлов в CSS/3 ... 1420
ПРЕДМЕТНЫЙ УКАЗАТЕЛЬ ... 1421

Copyright © StatSoft, 1995

Глава 7:
УПРАВЛЕНИЕ ДАННЫМИ

ВВОДНЫЙ ОБЗОР

В данном разделе описания программного модуля *Управление данными* рассматриваются основные принципы организации файлов данных (общие как для системы *STATISTICA*, так и многих других программ, в том числе большинства приложений для управления базами данных), а также некоторые уникальные возможности, предоставляемые системой *STATISTICA*.

Общая организация файлов данных в системе *STATISTICA*

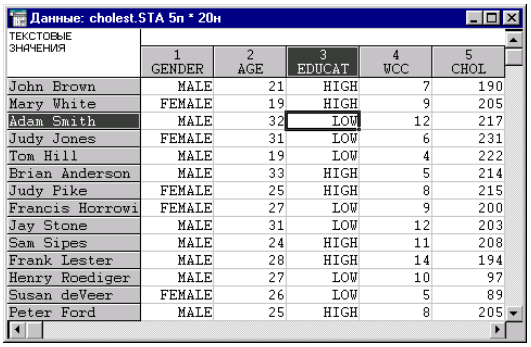
Данные в системе *STATISTICA* хранятся в виде файлов, где записи оптимизированы для быстрого доступа и экономии места на диске. Общая организация файлов совпадает с организацией файлов в программах управления базами данных. Каждый такой файл можно представить себе в виде "таблицы", в строках которой находятся *записи.* (например, информация о пациентах), причем все записи состоят из одинакового числа *полей*. Так, в приведенном ниже примере для каждого из пациентов представлены данные пяти типов.

Наблюдения (записи). При статистическом анализе данных записи часто называют *наблюдениями* (строки в таблице исходных данных). Они представляют собой элементарные "единицы" для анализа (например, пациенты, участники эксперимента, дни при анализе временных рядов, результаты измерений при проведении контроля качества и т.п.).

Переменные (поля). Поля, представляющие собой столбцы в таблице исходных данных, называются *переменными.* В каждой переменной хранится информация определенного типа, соответствующая каждому из наблюдений. Другими словами, каждое наблюдение включает ряд величин, которые соответствуют определенным переменным в наборе данных. Например, если в качестве наблюдений фигурируют пациенты, для каждого из которых регистрируют данные пяти типов: пол (*Gender*), возраст (*Age*), уровень образования (*Educat*), количество лейкоцитов (*WCC*) и уровень холестерина (*Chol*), в файле данных, соответствующем этому примеру, будет пять переменных, а сам файл данных будет выглядеть следующим образом:

Имя набл.	GENDER	AGE	EDUCAT	WCC	CHOL
John Brown	MALE	21	HIGH	7	190
Mary White	FEMALE	19	HIGH	9	205
Adam Smith	MALE	32	LOW	12	217
Judy Jones	FEMALE	31	LOW	6	231
Tom Hill	MALE	19	LOW	4	222
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.

А так этот файл будет выглядеть в виде таблицы исходных данных в окне системы *STATISTICA.*



StatSoft®
Copyright © StatSoft, 1995

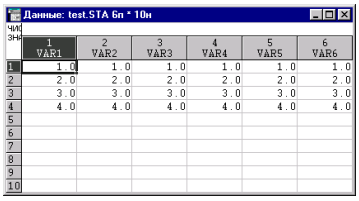
Преобразование наблюдений в переменные (транспонирование файлов данных). В большинстве методов статистического анализа различие между наблюдениями и переменными является чрезвычайно важным. Практически во всех методах анализа в системе *STATISTICA* вычисляются статистические параметры для набора переменных *по всем наблюдениям*, например, сравниваются средние значения *переменных*, полученных для различных *групп наблюдений*, и т.п. Однако для некоторых целей требуется прямо противоположное, а именно, необходимо рассматривать переменные как наблюдения.

Например, вместо вычисления значимости различия между двумя группами объектов исследования (например, группами наблюдений) при одном испытании (то есть по одной переменной), можно вычислить разность между средними значениями в двух группах испытаний (то есть группах переменных) для одного объекта (то есть для одного наблюдения). Для этого необходимо *транспонировать* файл, другими словами, "повернуть его на +90 градусов", чтобы наблюдения стали играть роль переменных, а переменные - наблюдений.

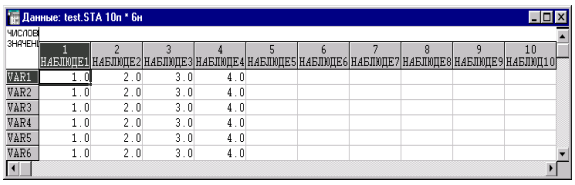
1	2	3	После	1	11	111
11	12	13	Транспони-рования	2	12	112
111	112	113	Переходит в	3	13	113

Пункт "транспонирование" имеется в выпадающем меню *Правка*. То же самое можно сделать с помощью средств *Транспонирование блока* или *Транспонирование файла данных*, как описано в главе 3 *Окно таблицы исходных данных*. Транспонирование часто приходится применять также при импорте файла с таблицей данных, организованной не стандартным для баз данных образом, а так, что наблюдения или "случаи" представляются в нем столбцами, а не строками. После импорта такие данных необходимо транспонировать, чтобы они правильно интерпретировались при последующем анализе.

До транспонирования:



После транспонирования:



Заметьте, что после транспонирования файла имена наблюдений стали именами переменных, а имена переменных стали именами наблюдений. Для получения более детальной информации нажмите клавишу F1 и войдите в *Электронное руководство*.

Значения переменных

На примере файла данных *Cholest.sta*, приведенного на предыдущей странице, проиллюстрированы различные типы данных, которые могут храниться в переменных. Другими словами, переменные могут различаться в зависимости от характера измерений, в которых были получены соответствующие данные (см. главу 8 *Элементарные понятия статистики*).

Категориальные (группирующие) переменные. В приведенном примере переменная *Gender (пол)* не содержит результатов физических измерений, таких,

Copyright © StatSoft, 1995

StatSoft®

например, как уровень холестерина, а содержит только *коды* или названия категорий (*MALE*, *FEMALE) (Мужской, Женский*), обозначающих групповую принадлежность отдельных наблюдений. Например, первое наблюдение принадлежит к группе *MALE (Мужской)*, а второе наблюдение – к группе *FEMALE (Женский)*. То же самое можно сказать и о переменной *Educat (Уровень образования).* Переменные такого рода часто называют *группирующими*, поскольку они позволяют определять принадлежность наблюдений к определенной группе (например, принадлежность наблюдений к группам с *высоким - HIGH* или *низким - LOW* уровнями образования), и впоследствии при проведении анализов производить сравнение этих групп.

В нашем примере обе группирующие переменные содержат текстовые (алфавитно-цифровые) данные, однако их значения могут быть также и числовыми (например, *1=MALE*, *2=FEMALE*). На самом деле благодаря такой гибкости определений в системе *STATISTICA* может использоваться так называемое *двойное представление* переменных, при котором для одной и той же переменной могут взаимозаменяемым образом использоваться как числовое, так и текстовое представления (см. стр. 1325 настоящего обзора).

Непрерывные переменные. Остальные переменные в приведенном файле данных (например, возраст или уровень холестерина) содержат значения, которые изменяются непрерывным образом (см. главу 8 *Элементарные понятия статистики*). В отличие от категориальных переменных, рассмотренных в предыдущем параграфе, это “обычные” переменные, которые могут использоваться для вычисления средних значений, стандартных отклонений и других статистических параметров.

Например, при простом анализе данных можно сравнивать уровни холестерина у мужчин и у женщин. Для этого программа вначале находит всех мужчин и всех женщин в наборе данных на основе значений группирующей переменной *Gender (Пол*), а затем вычисляет среднее значение переменной *Chol* для каждой группы.

Использование непрерывных переменных в качестве группирующих. Различие между двумя упомянутыми типами переменных состоит только в том, какие указания по их использованию и интерпретации применяются в программе. В системе *STATISTICA* эти типы не приписываются переменным “жестким” образом. Более того, непрерывные переменные можно использовать для определения принадлежности к группе. Для этого их перекодируют. Таблица организована таким образом, что позволяет производить перекодирование в интерактивном режиме для преобразования непрерывных переменных в группирующие (см. ниже).

Например, на основании значений уровня холестерина можно создать новую переменную и назвать ее *Ch_level* или каким-либо другим именем. Этой переменной будет присваиваться значение *HIGHCHOL* или числовое значение 2 в каждом случае, когда уровень холестерина превышает 200, и значение *LOWCHOL* или числовое значение 1, если уровень холестерина не превышает 200.

Пропущенные данные (ПД). Часто набор данных оказывается неполным из-за того, что для отдельных наблюдений данные для некоторых переменных отсутствуют. Например, для одного из пациентов могут отсутствовать данные по количеству лейкоцитов. В таком случае некоторые клетки таблицы оказываются пустыми. Такие *пропущенные данные* маркируются в файле данных заранее зарезервированным для этой цели значением (по умолчанию равным *-9999*). При проведении

StatSoft®

Copyright © StatSoft, 1995

любого анализа значения переменных, которым присвоены такие *коды пропущенных данных*, игнорируются, а клетки таблицы данных в системе *STATISTICA*, содержащие такие значения, отображаются как пустые. При создании нового файла данных все его значения первоначально являются пустыми, но в действительности они заполнены кодами пропущенных данных.

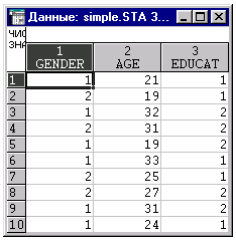
Формат отображения и точность хранения данных. Одним из атрибутов, который может устанавливаться отдельно для каждой переменной, является *формат отображения* (см. главу 3). Например, *1.005* может отображаться в таблице как *1.005000000000* или как *1.* В последнем случае дробная часть числа на экран не выводится. Форматы отображения определяют только способы отображения значений переменных в таблице на экране и никак не влияют на внутреннее представление (то есть точность хранения) данных, благодаря чему изменения формата не вносят изменений в реальные значения, хранящиеся в файле.

По умолчанию внутренней точностью хранения данных для стандартных файлов данных в системе *STATISTICA* является *двойная точность*, позволяющая хранить 15 значащих цифр (например, 1.23456789012345), независимо от установленного пользователем формата отображения. Однако в целях экономии места на жестком диске можно изменять точность хранения данных в системе управления файлами данных *Менеджер мегафайлов* (см. стр. 1338 и Том III).

Числовые и текстовые (алфавитно-цифровые) значения

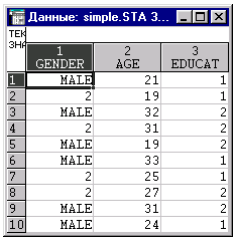
Обычно в программах управления базами данных тип каждого поля (переменной) определяется либо как *числовой*, либо как *текстовый*. Однако в системе *STATISTICA* любая переменная в наборе данных может быть *одновременно* и числовой, и текстовой, а при просмотре данных в таблице можно переключаться с одного типа на другой. Для этого нужно использовать пункт *Текстовые значения* выпадающего меню *Вид* таблицы исходных данных или нажать кнопку на панели инструментов этой таблицы.

Пример. Этот простой пример иллюстрирует так называемое двойное представление значений переменных в системе *STATISTICA*. Приведенная ниже таблица соответствует простому файлу данных, содержащему только числовые данные:



	1 GENDER	2 AGE	3 EDUCAT
1	1	21	1
2	2	19	1
3	1	32	2
4	2	31	2
5	1	19	2
6	1	33	1
7	2	25	1
8	2	27	2
9	1	31	2
10	1	24	1

Эти переменные не имеют текстовых значений. Чтобы их добавить, выделите щелчком мыши первое значение первого наблюдения, равное *1*, введите вместо него *MALE (Мужской)* и нажмите *Enter (Ввод)*. Таким образом, числовому значению *1* ставится в соответствие текстовое значение *MALE*. При этом не только замененное значение, но и все остальные значения, равные до введения изменений *1*, заменятся на *MALE*



	1 GENDER	2 AGE	3 EDUCAT
1	MALE	21	1
2	2	19	1
3	MALE	32	2
4	2	31	2
5	MALE	19	2
6	MALE	33	1
7	2	25	1
8	2	27	2
9	MALE	31	2
10	MALE	24	1

Что произойдет, если ввести текстовое значение в поле, не содержащее никаких числовых

Copyright © StatSoft, 1995

StatSoft®

значений? Будет ли это значение иметь какой-либо числовой эквивалент? Да, поскольку каждое текстовое значение в файлах данных системы *STATISTICA* имеет числовой эквивалент, и в нашем случае программа создаст его автоматически. Такие числовые эквиваленты берутся из последовательного ряда целых чисел, начиная с числа *100*. Заметим, что вводить новые текстовые значения и редактировать уже существующие можно в диалоговом окне *Диспетчер текстовых значений* (см. главу 3 *Окно таблицы исходных данных*).

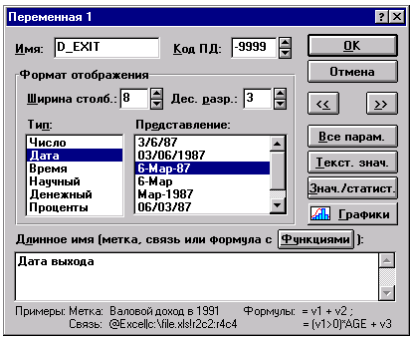
Преимущества двойного представления значений переменных. Эта уникальная особенность файлов данных системы *STATISTICA* упрощает ввод данных и помогает хранить данные и управлять ими, используя наиболее информативные названия, представляющие собой обычные слова, и сохраняя при этом их числовые значения, необходимые для проведения статистических вычислений. Однако необходимо отметить, что текстовые значения являются “настоящими” текстовыми значениями, а не только метками, удобными для вывода значений на экран. Их можно использовать, например, для определения уровней переменных, групповых кодов, условий выбора (например, выбрать, если *v1 = ‘MALE’*) и т.п.

Способ представления *дат* в файлах данных системы *STATISTICA*

Переменные, соответствующие календарным датам, хранятся в формате юлианского календаря, то есть каждая дата представляет собой одно целое число, равное количеству дней, прошедших с 1 января 1900 года. Например, дата, введенная и отображаемая на экране в виде *1/21/1968*, хранится в виде *24858*. Десятичные разряды, которые также могут присутствовать в записи числа, интерпретируются как время (см. следующий раздел). Хранящиеся в таком формате даты могут использоваться при последующем анализе, например, в модуле *Анализ выживаемости* для вычисления времени жизни (см. ниже), и преобразовываться с помощью арифметических операций. При этом в отчетах и на графиках, например, при разметке осей, даты выводятся в общепринятом формате.

Значения дат могут отображаться в таблице как в числовом (юлианском) формате, так и в одном из заранее определенных форматов отображения дат (например, *1/6/64*, *6-Янв-64*, *Янв-1964*, *01/06/64*, *01/06/1964*, *6-Янв*).

Для изменения формата отображения дат выберите пункт *Дата* в диалоговом окне *Переменная…(текущие спецификации)* (оно вызывается с помощью двойного щелчка мышью на имени переменной в таблице или из выпадающего меню *Правка*), а затем выберите один из предлагаемых форматов отображения.

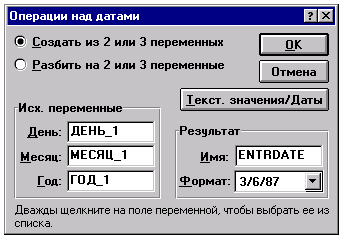


При вводе дат в новую переменную необходимо сначала заменить в диалоговом окне *Переменная…(текущие спецификации)* формат *Число*, который используется по умолчанию, на формат *Дата*, а затем выбрать нужный формат отображения дат. Таким способом можно вводить даты в любом заданном формате, например, в наиболее простом для ввода, даже если он отличается от желаемого формата отображения. Система *STATISTICA* распознает эти форматы, переводит введенные даты в нужный формат отображения и сохраняет их в

StatSoft®
Copyright © StatSoft, 1995

файле данных в юлианском формате. Для получения подробной информации о вводе новых и вставке скопированных значений дат в файл данных обратитесь к интерактивному *Электронному руководству*.

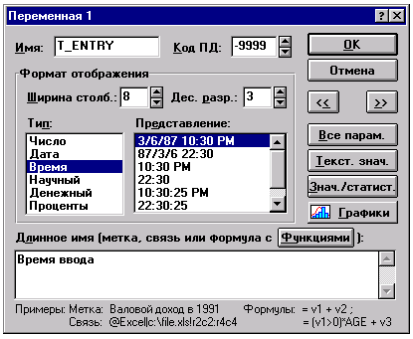
С помощью диалогового окна *Операции над датами* можно создать одну переменную-дату из двух (месяц и год или день и месяц) или трех (день, месяц, год) переменных. Можно также разделить одну переменную-дату на две или три переменные. Указанное диалоговое окно открывается при выборе пункта *Значения дат* выпадающего меню кнопки **Переменные** таблицы исходных данных (см. ниже).



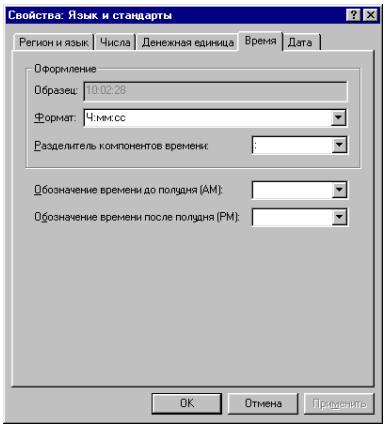
Для демонстрации примеров создания переменных-дат из числовых переменных и разделения переменных-дат на числовые переменные обратитесь к интерактивному *Электронному руководству*, для чего нажмите клавишу F1 или кнопку ? в показанном выше диалоговом окне.

Способ представления *времени* в файлах данных системы *STATISTICA*

Значения времени в переменных хранятся в виде десятичных значений, эквивалентных части дня, прошедшей к данному моменту после полуночи. Например, *6:00AM (6 ч. утра)* хранится в виде *0.25*. Временные значения, хранящиеся в таком формате, могут использоваться при последующем анализе данных и преобразовываться с помощью арифметических операций. В отчетах и на графиках, например, при разметке осей, значения времени выводятся в общепринятом формате.



Отображение времени в системе *STATISTICA* производится в соответствии с текущими установками параметра *Язык и Стандарты – Время* на *Панели управления* Windows.



Для демонстрации примеров разделения переменных времени или их создания из числовых переменных обратитесь к интерактивному *Электронному руководству* (для этого нажмите кнопку ? в диалоговом окне *Переменная...* или клавишу F1).

Copyright © StatSoft, 1995

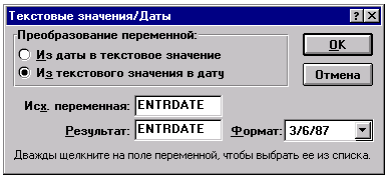
StatSoft®

Использование значений *дат* в качестве кодов (например, значений группирующих переменных)

В некоторых процедурах системы *STATISTICA* требуется, чтобы значения группирующих переменных (кодов) были меньше 32000, т.е. меньше, чем значения некоторых дат, хранящихся в юлианском формате. Это, однако, не относится к группирующим кодам, используемым во всех процедурах при построении *статистических* (и *быстрых статистических*) *графиков*, а также в модулях *Основные статистики*, *Контроль качества*, *Планирование эксперимента*, *Анализ процессов* и других. Для того чтобы можно было использовать значения дат в качестве кодов, величины которых не могут превышать 32000, их нужно преобразовать в текстовые значения с помощью приведенной ниже процедуры.

Прямое и обратное преобразование значений *дат* в текст

Иногда бывает полезно создавать текстовые значения, содержащие информацию о датах, например, при использовании переменных-дат в качестве переменных-кодов, значения которых не превышают 32000 (см. предыдущий абзац). В таком случае можно преобразовать дату в текстовую переменную, имеющую числовой эквивалент, который находится в требуемом диапазоне значений. Это позволяет использовать такие переменные в качестве кодов во всех процедурах системы *STATISTICA*, требующих, чтобы числовое значение было меньше 32000. Для преобразования дат в коды используется диалоговое окно *Текстовые значения/Даты*.



Для входа в это диалоговое окно нажмите кнопку *Текстовые значения/Даты* в диалоговом окне *Операции над датами* (см. стр. 1327).

Создание файлов данных и ввод данных

Файлы данных можно *создавать* путем *импорта* из других форматов или приложений (см. ниже) или непосредственно вводя данные в таблицу. Каждый файл данных содержит фиксированное количество переменных и наблюдений, которое, однако, можно легко изменить непосредственно в окне таблицы. Кроме того, он имеет другие общие характеристики, например, имена переменных, строку комментария и т.п. (см. главу 3 *Окно таблицы исходных данных*).

При импорте данных в систему *STATISTICA* из файла какого-либо другого формата указанные выше характеристики нового файла данных автоматически определяются на основании свойств импортируемого набора данных. Если же файл данных приходится создавать “вручную”, то упомянутые характеристики необходимо ввести непосредственно или принять установки по умолчанию, предлагаемые программой. Доступ к данным в системе *STATISTICA* может также осуществляться с помощью средств *Динамического обмена данными* (см. ниже).

Импорт данных из других приложений

Буфер обмена. Самым быстрым и во многих случаях самым простым способом получения доступа к файлам данных других приложений Windows, например электронных таблиц,

StatSoft®
Copyright © StatSoft, 1995

является использование буфера обмена. В системе *STATISTICA* он поддерживает форматы приложений MS Excel или Lotus 1-2-3 для Windows. Система *STATISTICA* может, например, правильно интерпретировать форматированные ячейки (типа *1000000* или *$10*), а также текстовые значения.

Импорт файлов. Использовании возможностей импорта файлов, предоставляемых модулем *Управление данными*, обеспечивает доступ к файлам данных ряда приложений Windows, а также приложений, не работающих под Windows. Эти файлы данных могут быть преобразованы в формат системы *STATISTICA*. Возможности импорта файлов включают доступ как к форматированным, так и неформатированным (ASCII) файлам.

Основные преимущества использования средств импорта файлов по сравнению с применением буфера обмена заключаются в следующем:

- Пользователь может определять конкретный способ преобразования данных, например, выбирать именованные блоки во внешних файлах данных, решать, нужно ли импортировать имена переменных, текстовые значения, имена наблюдений, а также определять способ их интерпретации.
- Пользователь имеет доступ к типам данных, с которыми нельзя или сложно работать непосредственно через буфер обмена (например, таким как длинные метки или специальные коды пропущенных данных, а также к компонентам таких файлов данных, как трехмерные рабочие листы пакета Lotus 1-2-3 для Windows).

Связи *DDE*. Наконец, система *STATISTICA* поддерживает соглашения *динамического обмена данными (DDE)*, что позволяет динамически связать блок данных в ее таблице исходных данных с подмножеством данных другого Windows-приложения. Эта процедура намного проще, чем могло бы показаться, и ее реализация не требует глубокого знания принципов организации *DDE*. В частности, это относится к использованию команды *Установить связь* (которая используется вместо метода сценариев (script-entry)). Подробное описание см. в главе 3 или *Электронном руководстве*.

Форматы отображения *dat*. В файлах данных системы *STATISTICA,* которые организованы аналогично файлам баз данных, форматы отображения дат задаются для переменных в целом, а не для отдельных ячеек, как в таблицах Excel. Вследствие этого переменные, форматированные в Excel как даты, если для них в системе *STATISTICA* не установлен формат *Дата*, будут представлены в юлианском (целочисленном) формате (например, *34092* вместо *3 мая 1993*) (см. главу 3*Окно таблицы исходных данных*).

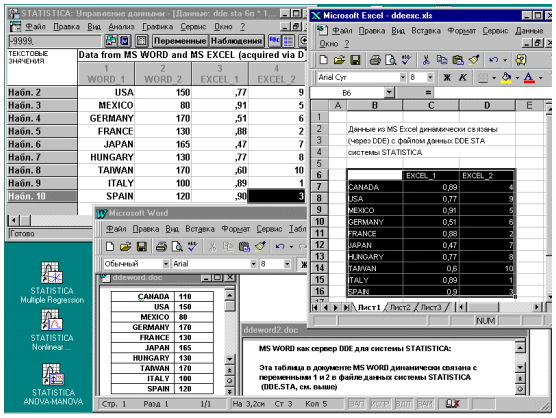
Экспорт данных в другие приложения

Буфер обмена и средства трансляции файлов данных, описанные в предыдущем разделе применительно к импорту файлов в формате других приложений, могут также быть использованы для экспорта данных системы *STATISTICA* в другие форматы. При этом поддерживаются те же форматы файлов, что и при импорте (см. предыдущий раздел).

Средства динамического обмена данными (*DDE*)

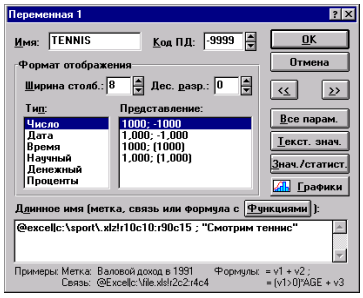
Пользователь может организовать *динамический обмен данными (DDE)* между файлом-источником (или “сервером”), например электронной таблицей MS Excel или документом MS Word, и файлом данных системы *STATISTICA* (файлом-клиентом), так что при изменении данных в файле-источнике данные в файле-клиенте будут автоматически обновляться.

Copyright © StatSoft, 1995



Например, приведенный выше файл данных системы *STATISTICA* динамически связан с данными рабочего листа MS Excel и таблицы MS Word.

Типичная область применения динамически связанных файлов – производственные системы, где файл данных системы *STATISTICA* может, в частности, быть связан с выходом измерительного устройства (например, для автоматического ежечасного обновления результатов определенных измерений). Организовать связь *DDE* можно, используя удобную команду *Установить связь* выпадающего меню *Правка* таблицы исходных данных. Кроме того, можно ввести определение связи в поле *Длинное имя (метка, формула, связь)* диалогового окна *Переменная… (Изменение (текущих) спецификаций)*. Более подробная информация и примеры приведены на стр. 1352.



После того как связь создана, ее характеристиками можно управлять в окне *Диспетчер связей* , доступ к которому осуществляется через команду *Связи…* выпадающего меню *Правка* таблицы исходных данных.

Средства динамической интеграции данных *DDE* имеют и другие применения в системе *STATISTICA*. Более подробную информацию можно найти в разделе интерактивного электронного руководства, посвященном *DDE*.

Проверка данных

После того как данные введены в файл вручную или импортированы, может возникнуть необходимость проверки их целостности (логической непротиворечивости) и полноты. Хотя существуют некоторые типы ошибок данных, которые выявить невозможно, имеется множество способов проверки (верификации) данных, основанных на общих критериях, которым эти данные должны удовлетворять.

К примеру, простейшие процедуры сводятся к проверке того, является ли значение переменной допустимым (например, переменная *Gender (Пол)* может принимать лишь значения 1 и 2) или попадает ли оно в диапазон допустимых значений (скажем, переменная *Age (Возраст)* должна быть больше *0* и меньше *200*). Эти процедуры могут быть также весьма сложными и предусматривать проверку множества логических условий, в которых каждая величина может быть связана с другими величинами того же или другого файла данных. Вот относительно простой пример проверки с условиями: если некоторое лицо является мужчиной или не достигло определенного возраста, количество беременностей для него не должно быть больше нуля. Система *STATISTICA* содержит гибкие средства проверки данных и позволяет работать с критериями практически неограниченной

StatSoft®
Copyright © StatSoft, 1995

сложности. Более подробная информация об этих средствах приведена на стр. 1380.

Редактирование данных и изменение их структуры

Редактирование данных включает в себя не только ввод и изменение отдельных значений. Часто возникает необходимость выполнения более глобальных операций, таких как переупорядочивание, перемещение, копирование переменных или наблюдений или изменения общего размера файла данных путем добавления или удаления наблюдений или переменных.

В системе *STATISTICA* имеется два типа операций реструктуризации данных: (1) - традиционные операции над таблицей исходных данных с использованием буфера обмена, и (2) - глобальные операции реструктуризации данных.

(1) Операции над наблюдениями и переменными с использованием буфера обмена. В отличие от *глобальных* операций, производимых над блоками наблюдений или переменных, рассматриваемыми в качестве цельных логических единиц (см. ниже), операции с использованием буфера обмена выполняются по стандартным для таблиц правилам и затрагивают только содержание выделенных ячеек. Результат операции полностью определяется выделенным блоком и положением курсора.

Следует отметить, что поскольку *глобальные* операции удаления или перемещения переменных (см. ниже) оказывают влияние не только на содержание соответствующих столбцов, но и на сами столбцы, то они изменяют и саму структуру файлов данных. Такие операции всегда выполняются над переменными в целом, независимо от того, все или только часть наблюдений выделены в данный момент в выбранных переменных.

При операциях с буфером обмена, напротив, вырезается только выделенный сегмент данных (как обычно при редактировании электронных таблиц). Вставка данных всегда производится, начиная с текущего положения курсора и далее вниз.

Таким образом, если вы, например, для перемещения переменной (a) выделили и вырезали целый столбец, (b) выделили другой целый столбец и затем (c) вставили содержимое буфера обмена в новое место для замены предыдущих значений, то процедура достигнет цели только в том случае, если курсор помещен в верхнюю ячейку нового столбца. Если курсор находится где-то в середине столбца, то вставка блока будет выполнена, начиная с этого места и далее вниз.

(2) Глобальные операции над наблюдениями и переменными. В отличие от операций, использующих буфер обмена для вырезания, копирования и вставки блоков данных (или содержания целых столбцов или строк таблицы), операции, производимые с помощью меню *Правка-Переменные* или *Правка-Наблюдения* таблицы исходных данных, а также с помощью кнопок панелей инструментов *Переменные* или *Наблюдения* (в том числе *Добавить*, *Переместить*, *Копировать* и т. п.), влияют не только на содержание столбцов или строк данных, но и на сами столбцы и строки (когда это допустимо). Например, операция *Удалить* удаляет не только содержание выбранной группы столбцов (строк), но и сами эти столбцы (строки).

Следует также отметить различие между глобальными операциями над переменными или наблюдениями, которые рассматриваются как цельные логические единицы файлов данных системы

Copyright © StatSoft, 1995

STATISTICA, и всеми операциями над таблицей исходных данных с использованием буфера обмена, которые производятся по тем же правилам, что и для любых других электронных таблиц (например, MS Excel). Например, глобальные операции копирования, удаления и перемещения переменных (или наблюдений), производимые через эти меню, не зависят от текущего положения курсора или выделенного блока, за исключением того, что имена переменных (или наблюдений), соответствующих выделенному блоку, будут предложены в качестве выбора по умолчанию в соответствующем диалоговом окне (это дает простой способ выбора переменных или наблюдений, над которыми должна быть произведена операция). Глобальные операции над переменными всегда затрагивают все наблюдения для выбранных переменных независимо от того, выделены ли для этих переменных все наблюдения или только их часть.

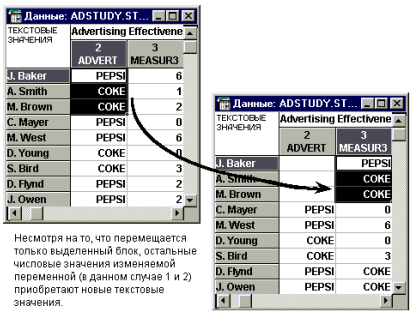
При операциях с использованием буфера обмена копируется только выделенный сегмент данных, а вставка блока данных всегда выполняется, начиная с текущего положения курсора и далее вниз.

Операции над таблицей исходных данных с использованием буфера обмена

При копировании или перемещении блока в таблице исходных данных (например, методом *Перетащить и отпустить*, см. главу 3), копируемые значения зависят от режима отображения информации в таблице исходных данных.

Числовые значения или метки значений. Если при копировании блока в буфер обмена значения в таблице исходных данных отображены в числовом виде (или в виде меток значений), то будут скопированы только эти числовые значения (или метки значений).

Текстовые значения. Если при копировании блока в буфер обмена в таблице исходных данных отображены текстовые значения, то в буфер обмена копируются не только эти текстовые значения, но и соответствующие числовые значения и метки (если они определены). Это может привести к тому, что текстовые значения будут приписаны числовым значениям, которые ранее не имели текстовых эквивалентов (см. ниже).



Преобразование данных

К самым распространенным операциям обработки данных в статистическом анализе относятся операции преобразования и перекодировки данных. Например, вместо анализа исходных значений переменных, полученных в результате ввода или импорта, может потребоваться вычисление новых значений, представляющих собой некоторые комбинации исходных результатов измерений.

Пример. Предположим, необходимо создать переменную, которая должна содержать итоговый результат испытаний для каждого лица, равный среднему его/ее результатов по трем испытаниям. Если данные о результатах этих испытаний хранятся в переменных *v1* - *v3*,

StatSoft
Copyright © StatSoft, 1995

то необходимо вычислить значения новой переменной (например, *v4*), равные

```
(v1+v2+v3)/3
```

Это типичный пример формулы преобразования данных, применяемой к каждому наблюдению в файле данных.

Преобразования данных в системе *STATISTICA*. Как уже говорилось, файлы данных системы *STATISTICA* организованы в виде совокупности наблюдений. Если пользователь не задал иных условий, то всякая формула преобразования (см. вышеприведенный пример) будет применена ко всем наблюдениям в файле данных. В системе *STATISTICA* существует два способа задания преобразований:

(1) Простые однострочные формулы преобразования могут быть введены непосредственно в диалоговом окне *Переменная… (Изменение (текущих) спецификаций* (см. главу 3*Окно таблицы исходных данных*). Простейший способ открыть его – это дважды щелкнуть на имени переменной в таблице исходных данных. При этом формула становится в некотором смысле “определением” переменной. Например, переменная *Testmean* в вышеприведенном примере определяется как среднее трех других переменных. Таким образом, действие таких формул отличается от действия формул в типичных приложениях по обработке электронных таблиц, поскольку формула, задающая переменную в системе *STATISTICA*, влияет не на одну ячейку, а на все значения этой переменной (столбца), то есть на все наблюдения, если иное не оговорено специально.

(2) Для преобразований можно использовать и язык программирования *STATISTICA BASIC*, доступ к которому осуществляется через выпадающее меню *Анализ* любого модуля системы *STATISTICA* (см. стр.1367). Язык *STATISTICA BASIC* представляет собой простой в использовании, но мощный язык программирования, который может применяться для разнообразных целей - от простых преобразований переменных до создания специализированных постоянных дополнений к вычислительным или графическим процедурам и средствам управления данными в системе *STATISTICA*. В частности, он используется в тех наблюдениях, когда требуемое преобразование нельзя осуществить средствами таблицы исходных данных, например при необходимости использовать циклические процедуры или функции перекодировки.

Другие преобразования. Кроме формул в таблицах исходных данных и программ на языке *STATISTICA BASIC*, в системе *STATISTICA* имеется ряд заранее определенных специализированных процедур преобразования, доступ к которым осуществляется через меню и диалоговые окна конкретных модулей (например, в модуле *Временные ряды* имеется большой выбор специализированных функций сглаживания и фильтрации данных).

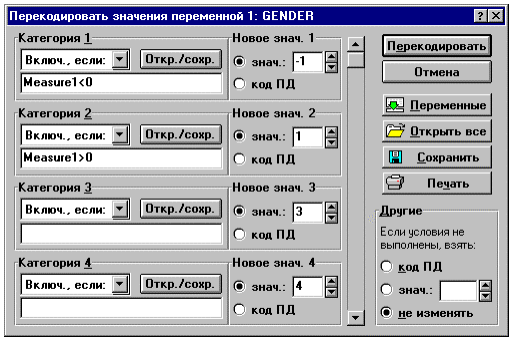
Перекодировка данных

Перекодировка данных означает переход от исходных к некоторым новым значениям. Например, можно перекодировать уровни холестерина, значения которых расположены в интервалах *менее 170*, *от 171 до 200* и *более 200* в значения *1*, *2* и *3* (или *НИЗКИЙ*, *СРЕДНИЙ* и *ВЫСОКИЙ*) соответственно. Собственные функции перекодировки практически любой сложности, многократно используемые в программах преобразования данных, можно создавать средствами языка *STATISTICA BASIC* (полное и наиболее современное описание его синтаксиса можно найти в *Электронном руководстве*). Кроме того, в таблице исходных

Copyright © StatSoft, 1995

данных в любой момент доступны удобные интерактивные средства перекодировки (см. описание кнопки **Переменные** панели инструментов таблицы исходных данных в главе 2). Вызываемое этой кнопкой прокручиваемое диалоговое окно *Перекодировать значения* позволяет задавать новые значения текущей переменной (см. ниже поля *Новое знач. 1*, *Новое знач. 2*, *Новое знач. 3* и т.п.). Эти значения вычисляются в соответствии с конкретными условиями, задаваемыми пользователем (см. поля *Категория 1*, *Категория 2*, *Категория 3*).

При задании условий используются стандартные синтаксические соглашения, общие для всех процедур системы *STATISTICA*, где производятся какие-либо операции по "выбору наблюдений" в зависимости от их значений (см. раздел *Строка состояния* в главе 2).



Например, описанные выше условия перекодировки преобразуют отрицательные значения текущей переменной (*Measur1*) в *–1*, а положительные – в *+1*. Обратите внимание, что нулевое значение (единственное не упомянутое в двух условиях перекодировки) останется неизменным в соответствии с установкой переключателя в состояние *не изменять* (см. поле *Другие* в нижнем правом углу окна).

Здесь могут быть также использованы *условия выбора наблюдений*, определенные в другом диалоговом окне и сохраненные в виде файлов, содержащих отдельные условия или наборы условий. Следует отметить, что условия перекодировки могут быть гораздо более сложными (обратитесь к электронной документации, нажав F1 в соответствующем окне). Они могут быть сформулированы таким образом, что новые значения переменной будут зависеть не от предыдущих ее значений, а только от значений некоторых других переменных совокупности данных. Таким образом, эта операция может быть использована не только для перекодировки существующих данных, но и для задания значений новой переменной в зависимости от условий, которым удовлетворяют другие переменные (см. приведенный ниже пример).

Использование перекодировки для создания новых переменных. Любые средства перекодировки данных - формулы в таблице исходных данных или программы на языке *STATISTICA BASIC* – могут использоваться для задания значений новой переменной в зависимости от условий, которым удовлетворяют другие переменные. Однако наиболее просто определить новую переменную с помощью интерактивной процедуры перекодировки, описанной в предыдущем параграфе. Эта вспомогательная процедура может быть в любой момент вызвана из окна таблицы исходных данных (см. описание кнопки **Переменные** панели инструментов таблицы исходных данных в главе 2). Как уже говорилось в предыдущем разделе, выделенная в настоящий момент переменная может даже не использоваться в условиях перекодирования, а функция для задания ее значений может включать условия, определенные для других переменных.

Например, можно добавлять к файлу данных новую (пустую) переменную, а затем с помощью функции перекодировки определять ее новые значения. Например, таким образом можно задать переменную, имеющую значение *1* для всех лиц мужского пола в возрасте от 18 до 25 лет с уровнем холестерина ниже 200; значение *2*

StatSoft®
Copyright © StatSoft, 1995

- для всех лиц мужского пола в возрасте от 18 до 25 лет с уровнем холестерина более 200 и значение “пропущенные данные” для всех остальных лиц.

Разделение и объединение файлов данных

Разделение файлов данных. Все процедуры анализа данных в системе *STATISTICA* требуют выбора переменных, подлежащих анализу. Однако все они также позволяют использовать условный выбор наблюдений для обработки данных (то есть фильтры, позволяющие учитывать при анализе только наблюдения, удовлетворяющие определенному условию, например, исключать наблюдения, для которых *v1='MALE'*; см. раздел *Строка состояния* в главе 2). Хотя каждую процедуру анализа можно выполнять только для определенного подмножества данных, тем не менее, иногда удобно разделить файл данных на такие подмножества, например, для увеличения скорости обработки, т.к. короткие файлы обрабатываются быстрее. Кроме того, для выбора всех переменных можно просто использовать кнопку *Выбрать все*, поэтому целесообразно сохранить все представляющие интерес переменные в отдельном файле с новым именем. Для выбора подмножеств данных можно использовать пункт *Подмножество* выпадающего меню *Анализ* модуля *Управление данными*.

Объединение файлов данных. Существует множество ситуаций, в которых необходимо использовать объединение файлов. К примеру, может потребоваться *добавить* к концу (“низу”) существующего файла данные, собранные для новой совокупности лиц, то есть “слить” этот файл с новым файлом, содержащим те же переменные. Можно также провести новые испытания (создать новые переменные) для наблюдений, входящих в существующий файл данных.

Файл, содержащий новые переменные, будет присоединен к концу (“правой стороне”) существующего файла данных, то есть будет объединен с другим файлом, содержащим те же наблюдения. Команда *Объединить* включена в выпадающее меню *Анализ* модуля *Управление данными*.

Относительное объединение. При слиянии переменных (см. второй из вышеприведенных примеров) часто возникает проблема, связанная с тем, что количество и порядок наблюдений в обоих файлах данных должны точно совпадать. Если во втором файле данных хотя бы одно наблюдение отсутствует или не имеет эквивалента в первом файле, то правильное соответствие следующих за ним наблюдений будет нарушено. Решение этой проблемы состоит в использовании *реляционного объединения*. Для реляционного объединения наблюдений из двух файлов необходимо задать в каждом файле *переменную-ключ*, или так называемый идентификатор наблюдения. При слиянии файлов программа для каждого наблюдения проверяет значение переменной-ключа в каждом из файлов и объединяет наблюдения только тогда, когда соответствующие ключи совпадают. Система *STATISTICA* располагает богатым набором средств реляционного объединения файлов.

Copyright © StatSoft, 1995

ОБЗОР ПРОГРАММНЫХ МОДУЛЕЙ

Модуль *Управление данными*

Модуль *Управление данными* – это полноценная система управления базами данных, располагающая разнообразными средствами редактирования и преобразования данных и реструктуризации файлов.

Таблица исходных данных. *Окно таблицы исходных данных* системы *STATISTICA*, доступное не только в модуле *Управление данными*, но и в любом другом месте программы, содержит большой набор возможностей для редактирования данных и реструктуризации файлов. В их число входят: средства ввода данных; полный набор операций метода *Перетащить и отпустить*, встроенные формулы для преобразования значений переменных, встроенные средства перекодировки данных, поиска и замены. Кроме того, пользователь имеет возможность быстрого перемещения курсора в заданные позиции файла, выполнения операций с блоками (копирование, перемещение, удаление, транспонирование, стандартизация) и операций быстрого заполнения случайными значениями, а также располагает средствами для выполнения глобальных операций реструктуризации файлов (например, добавления, удаления, копирования и перемещения групп наблюдений и переменных, и т. п.) и для ранжирования значений, управления текстовыми значениями и метками и т.п.

Файлы данных системы *STATISTICA*. Файлы данных системы *STATISTICA* оптимизированы с точки зрения времени доступа и эффективности использования дискового пространства. Их можно рассматривать как "Рабочие книги", в которых хранится информация о всех вспомогательных файлах (графиках, таблицах результатов, отчетах, программах и т. п.), используемых при работе с этим набором данных. Все эти файлы могут открываться автоматически. Файлы данных системы *STATISTICA* могут содержать имена наблюдений, текстовые и числовые значения данных, значения в формате дат и времени, короткие и длинные метки переменных с формулами и связями *DDE*, длинные метки значений, а также другую информацию, в том числе заголовки, примечания, описания шрифтов, цветов и т.п. (см. *Изменение (текущих) спецификаций таблицы исходных данных*, глава 3). Поскольку переменные могут определяться формулами таблицы исходных данных, то для их пересчета достаточно одного нажатия клавиши (см. *Изменение (текущих) спецификаций таблицы исходных данных*, глава 3). Один файл может содержать практически неограниченное количество наблюдений и до 4092 двойных (текстовых и числовых) переменных. Программа *Менеджер мегафайлов*, доступная из модуля *Управление данными*, позволяет обрабатывать файлы, содержащие до 32000 переменных (см. ниже соответствующий раздел этой главы, а также том III).

Двойное (числовое и текстовое) представление значений. В системе *STATISTICA* реализована уникальная возможность *двойного представления* переменных, благодаря которому каждое значение переменной может иметь две эквивалентных формы - числовую и текстовую. В любом из модулей можно использовать как числовое, так и текстовое представление значений переменной. Например, если *1* эквивалентно значению *MALE* в файле данных, то в диалоговом окне *Условия выбора* можно задать как условие "исключить, если *v1=1*", так и более понятное условие "исключить, если *v1='MALE'*". Диалоговое окно *Диспетчер текстовых значений* работает как с текстовыми

StatSoft®

Copyright © StatSoft, 1995

значениями, так и с длинными метками значений, позволяя полностью реализовать преимущества гибкой структуры данных при минимальных затратах сил на работу с клавиатурой. Двойное представление значений поддерживается всеми модулями системы *STATISTICA*, что облегчает создание процедур обработки и описание результатов (создание таблиц результатов и графиков) даже в тех ситуациях, когда обрабатываются только числовые данные. При просмотре данных в таблице исходных данных можно переключать способ представления переменных с помощью кнопки панели инструментов. Программа поддерживает также длинные метки значений. Для их просмотра или редактирования необходимо одновременно нажать клавишу CTRL и кнопку панели инструментов.

Импорт и экспорт данных. Кроме разнообразных возможностей обмена данными с помощью буфера обмена (программа поддерживает специальные форматы представления данных в буфере обмена, что позволяет правильно интерпретировать данные других приложений) и поддержки способов *динамического обмена данными* (*DDE*), система *STATISTICA* имеет ряд мощных средств для импорта и экспорта данных (см. *Импорт файлов* и *Экспорт файлов*, стр. 1389 и 1408 соответственно). Здесь предусмотрена возможность импорта и экспорта файлов различных форматов, в том числе применяемых в последних версиях пакетов Microsoft Excel, Lotus 1-2-3, а также в любых версиях Symphony, Quattro, dBASE III+, IV и V, Paradox, файлов формата STG (формат графических файлов системы *STATISTICA*), переносимого формата файлов больших вычислительных машин (совместимого с SPSS), а также ASCII-файлов в фиксированном и свободном форматах. Развитые средства поддержки протокола *ODBC* (Open Data Base Connectivity) (см. главу 3) обеспечивают удобную связь практически со всеми распространенными системами управления базами данных (СУБД), включая Oracle и Sybase. Кроме того, можно использовать ряд произвольных текстовых форматов файлов (последовательный, матричный, неполный матричный и т.п.). При импорте форматированных данных система *STATISTICA* распознает структуру файла, метки значений, текстовые и логические переменные, пустые ячейки и поля и преобразует их в эквивалентные элементы собственного файлового формата. Некоторые функции диалоговых окон импорта и экспорта данных (например, операции с именованными блоками в таблицах) упрощают процесс преобразования данных.

Преобразование данных. Средства преобразования данных (см. *Изменение (текущих) спецификаций таблицы исходных данных*, глава 3) встроены в таблицу исходных данных системы *STATISTICA* и позволяют задавать переменные с помощью формул. Результаты вычислений по всем или некоторым формулам пересчитываются при нажатии кнопки *Пересчитать* панели инструментов (имеется функция автоматического пересчета, позволяющая пересчитывать результаты вычислений по всем формулам, на которые влияют изменения в файле данных). Формулы таблицы исходных данных могут содержать логические операции и использоваться не только для вычисления новых значений, но и для простых операций перекодировки и проверки данных. Имеются также специализированные средства перекодирования, которые могут быть вызваны непосредственно из таблицы исходных данных в любом месте программы. Кроме того, в каждом модуле имеется доступ к средствам языка программирования *STATISTICA BASIC*, с помощью которого легко создаются процедуры преобразования и перекодировки данных любой степени сложности.

Copyright © StatSoft, 1995

Язык *STATISTICA BASIC*. Язык *STATISTICA BASIC* представляет собой простой в использовании, но мощный язык программирования, который может применяться для разнообразных целей - от простых преобразований переменных до создания сложных специализированных постоянных дополнений к вычислительным или графическим процедурам и средствам управления данными в системе *STATISTICA*. В частности, его обычно применяют, если требуемое преобразование нельзя осуществить средствами таблицы исходных данных, например, при необходимости использовать циклы или функции перекодировки.

Проверка данных. Интерактивные средства проверки и очистки данных модуля *Управление данными* позволяют задавать критерии целостности и внутренней непротиворечивости данных практически любого уровня сложности. Проверка данных может проводиться в диалоговом режиме, благодаря чему пользователь имеет возможность найти, отметить и/или откорректировать каждое наблюдение, не удовлетворяющее заданным условиям.

Другие операции. В системе *STATISTICA* предусмотрены средства для разделения файлов данных на *подмножества* в соответствии с выбранным критерием, а также возможность выбора операций *объединения* файлов, в том числе реляционного и иерархического реляционного объединения. Файлы данных можно также *сортировать* различными способами, в том числе с использованием нескольких вложенных условий сортировки. Заголовки файлов, содержащие метки, формулы и текстовые определения допускается объединять и преобразовывать. Можно *сдвигать* переменные, то есть задавать их смещение относительно друг друга, ранжировать их различными способами, *стандартизировать*, а *пропущенные данные* - заменять средними значениями. Форматированные списки спецификаций данных, файлы данных и подмножества файлов данных можно выводить на печать в соответствии с параметрами и установками, заданными пользователем.

Обычные свойства таблиц исходных данных и модуля *Управление данными*

Как уже указывалось, большинство совместно используемых средств управления данными встроены в меню таблицы исходных данных и могут вызываться во всех модулях с помощью панели инструментов, выпадающего меню *Правка* или контекстного меню. Эти средства включают в себя все операции в таблицах над наблюдениями и переменными, преобразование данных с помощью формул, язык программирования *STATISTICA BASIC*, возможности интерактивной перекодировки, ранжирования, заполнения диапазонов, сдвига переменных, *динамический обмен данными (DDE)*, работу с текстовыми значениями, временными значениями, датами, длинными метками, а также многие другие возможности, в том числе прямой импорт и экспорт данных и импорт с использованием протокола *ODBC*.

Модуль *Управление данными* обеспечивает все вышеперечисленные возможности и помимо этого предоставляет дополнительные специальные средства управления данными, такие как интерактивная проверка данных и функция очистки, а также специальные средства импорта и экспорта. В состав последних входит возможность импорта или экспорта двойного представления значений в форматы файлов, которые в обычном режиме это представление не поддерживают.

Доступ к возможностям, предоставляемым модулем *Менеджер мегафайлов*. В модуле *Управление данными* системы *STATISTICA* (за исключением версии *Quick STATISTICA*) имеется

StatSoft®

Copyright © StatSoft, 1995

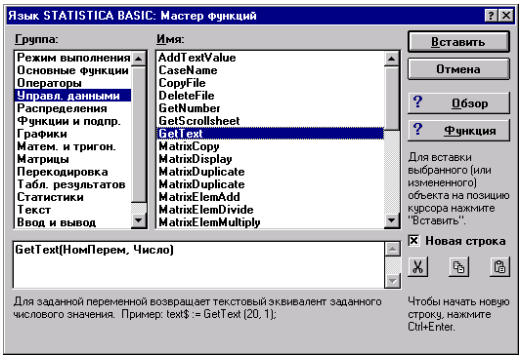
доступ ко всем возможностям, предоставляемым системой баз данных *Менеджер мегафайлов*. Здесь имеются средства для обработки данных, записанных в нестандартных форматах, в том числе к данным, организованным в виде очень длинных записей, и данным с очень длинными текстовыми значениями. Уникальной особенностью *Менеджера мегафайлов* является возможность обработки данных с предельно длинными записями (до 8 Мбайт на запись). Можно также использовать *Менеджер мегафайлов* в качестве архивной системы баз данных, в которой совместно могут храниться данные различных источников в исходных форматах. Модуль *Управление данными* предоставляет простые в использовании и гибкие средства обмена данными между файлами данных *STATISTICA* и архивными базами данных *Менеджера мегафайлов*. Широкие возможности модуля *Менеджер мегафайлов* подробно обсуждаются в отдельной главе (см. том III).

Управление данными с помощью *Командного языка* системы *STATISTICA*

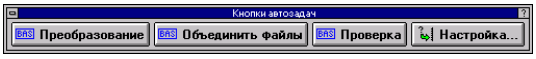
Операции управления данными могут также выполняться с помощью соответствующих команд *Командного языка* системы *STATISTICA,* доступ к которому имеется из выпадающего меню *Анализ* каждого модуля.

Операции управления данными и пользовательские процедуры на языке *STATISTICA BASIC*

В языке *STATISTICA BASIC*, доступ к которому при работе в любом модуле возможен из выпадающего меню *Анализ*, имеются разнообразные специализированные функции для управления базами данных.



Заметим, что программы, написанные на языке *STATISTICA BASIC*, можно включать в пакетные программы, написанные на *Командном языке* системы *STATISTICA* (*SCL*) (см. выше).



Пользователь может создавать собственные программы для управления базами данных, написанные на языке *STATISTICA BASIC*, вызываемые с помощью кнопок панели инструментов *Кнопки автозадач* и используемые как постоянные расширения системы *STATISTICA*.

Что такое *Менеджер мегафайлов*?

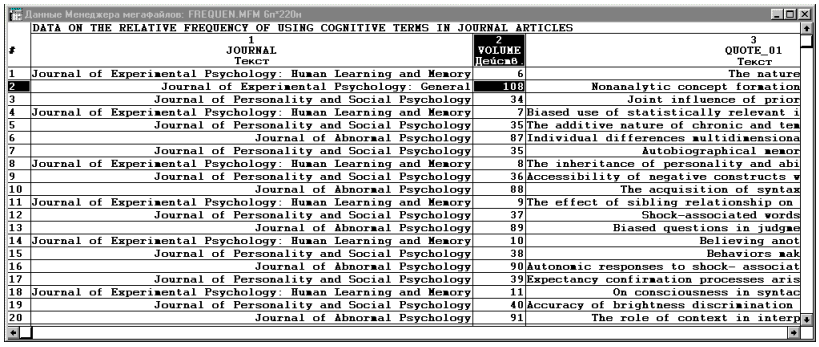
Обзор

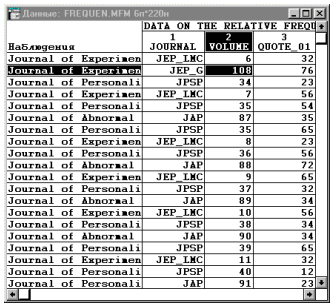
Менеджер мегафайлов представляет собой специализированную систему управления базами данных, доступную из модуля *Управление данными* системы *STATISTICA*. Ее особенностью является возможность управлять и непосредственно обрабатывать данные, которые необходимо преобразовать, агрегировать, выделить или очистить перед их использованием в статистических или графических процедурах. Примерами таких данных могут служить данные, организованные в записи очень большой длины, а также данные, являющиеся составной частью очень длинных текстовых значений.

Copyright © StatSoft, 1995

StatSoft®

Менеджер мегафайлов может обрабатывать записи данных (строки), максимальный размер которых 32000 столбцов, содержащих до 255 знаков, то есть максимальная длина каждой такой строки может достигать 8 Мбайт. Данные, состоящие из таких длинных записей, могут формироваться автоматизированными устройствами контроля качества, устройствами сбора данных или контроля технологических процессов. Иногда такие файлы используются в интегрированных архивных банках числовых данных, состоящих из файлов, полученных посредством объединения или конкатенации.





Исходные данные Менеджера мегафайлов (с длинными текстовыми значениями)

Готовый для обработки файл данных STATISTICA

Менеджер мегафайлов и другие средства системы *STATISTICA*

Как правило, большинство процедур обработки данных могут быть легко реализованы с помощью стандартных средств, доступ к которым осуществляется из меню любого модуля системы *STATISTICA*. Поэтому *Менеджер мегафайлов* обычно используют только в том случае, когда возникает необходимость облегчить обработку данных нестандартного формата (импорт, управление и предварительную обработку), например, при поддержке очень больших баз данных, проведении предварительной обработки импортированных текстовых значений большой длины и т.п. Заметим, что стандартные процедуры, имеющиеся в каждом модуле системы *STATISTICA*, тоже могут решать очень большие и сложные задачи, например:

- вычислять за один проход количественные показатели распределения или корреляционные матрицы для очень большого количества переменных, (порядка 1000). Для этого используется функция *Быстрые основные статистики*, доступ к которой осуществляется в каждом модуле с помощью выпадающего меню *Анализ* или кнопки (для получения подробной информации см. главу 10);
- осуществлять сложную перекодировку и агрегирование данных, сложные операции над текстовыми значениями или комплексную очистку и проверку данных с использованием средств языка *STATISTICA BASIC*, доступных в каждом модуле системы *STATISTICA* через выпадающее меню *Анализ*;

StatSoft®

Copyright © StatSoft, 1995

- импортировать файлы большого размера из других приложений, например, Excel, Lotus, Quattro Pro, Paradox и т.д. Для этого используется команда *Импорт*, доступная в каждом модуле *STATISTICA* из выпадающего меню *Файл* (см. главу 3);
- осуществлять последовательное, иерархическое или реляционное объединение файлов с помощью команды *Объединить* выпадающего меню *Анализ* в модуле *Управление данными*. Для этого также используются средства языка *STATISTICA BASIC* для доступа к внешним файлам и функции слияния данных с использованием простого или сложного алгоритма, задаваемого пользователем.

Уникальные возможности *Менеджера мегафайлов*

Поддержка больших архивных банков данных; иерархические связи между базами данных . В модуле *Менеджер мегафайлов* предусмотрены возможности для агрегирования наборов данных других приложений и создания очень больших (размером до 8 Мбайт на запись), эффективных архивных баз данных. Средства *Менеджера мегафайлов* также позволяют поддерживать связь между связанными и иерархически организованными наборами данных. Из любого такого архивного банка данных можно извлекать подмножества столбцов, чтобы затем использовать их в других приложениях, таких как система *STATISTICA,* электронная таблица Excel или Paradox.

Предварительная обработка длинных записей исходных данных. Еще одно уникальное приложение *Менеджера мегафайлов* – предварительный анализ, в процессе которого исходные данные подвергают агрегированию или предварительной обработке для последующего использования в аналитических процедурах. Такие наборы данных (получаемые, в частности, на автоматических установках контроля качества или на других устройствах) могут содержать записи, длина которых слишком велика для применения в стандартном приложении. Такая запись может содержать, например, до 32000 измерений в строке. С помощью *Менеджера мегафайлов* можно получить доступ к таким данным, преобразовать их в значимые величины и записать для последующего анализа в формате, совместимом с другим приложением, например, таким как система *STATISTICA* или таблицы Excel. Часто такие данные перед предварительной обработкой требуется подвергать очистке и проверке. Пользователь может создавать собственные процедуры, используя встроенный в *Менеджер мегафайлов* язык программирования, в котором имеются специализированные функции для интерактивного редактирования и проверки данных.

Обработка и анализ данных. Как показано выше, с помощью *Менеджера мегафайлов* можно не только агрегировать, хранить и поддерживать файлы с длинными записями, но и эффективно их обрабатывать. Во встроенном языке программирования *MM* имеется множество средств для проведения анализа данных и библиотека функций. Основные средства статистической обработки в модуле *Менеджер мегафайлов* позволяют работать с записями любой длины. Например, табулировать данные, вычислять количественные показатели распределения и генерировать корреляционные матрицы практически неограниченного размера (размер корреляционной матрицы, которая может быть создана в модуле *Менеджер мегафайлов*, превосходит емкость любого из существующих запоминающих устройств).

Длинные текстовые значения . Еще одна особенность *Менеджера мегафайлов* – это его способность обрабатывать очень длинные текстовые значения. Кроме того, встроенный в

Copyright © StatSoft, 1995

него язык программирования *MM* включает широкий набор функций для операций с текстовыми данными.

Обмен данными с файлами системы *STATISTICA*. В *Менеджере мегафайлов* используется специальный формат файлов, оптимизированный с учетом конкретного назначения модуля (в частности, для работы с данными различных типов из разных программ). В то же время модуль *Управление данными* содержит удобные и гибкие средства для обмена данными между файлами данных системы *STATISTICA* и архивными базами данных *Менеджера мегафайлов*. Более подробные сведения о *Менеджере мегафайлов* приведены в томе III.

StatSoft®

Copyright © StatSoft, 1995

ФОРМАТ ФАЙЛОВ ДАННЫХ СИСТЕМЫ *STATISTICA*

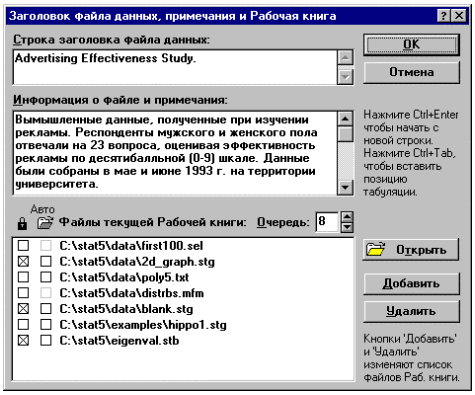
Общий формат

Переменные и наблюдения

Стандартные файлы данных STATISTICA, называемые также файлами исходных данных, образованы *наблюдениями* (записями или строками таблицы исходных данных) и *переменными* (полями или столбцами таблицы исходных данных). Количество наблюдений в файле данных практически не ограничено. В каждом файле может содержаться до 4092 переменных, каждая из которых в свою очередь может иметь "двойное представление" (текстовое и числовое) и длинные метки значений. Когда количество переменных превышает указанное предельно допустимое значение, используется модуль *Менеджер мегафайлов*, допускающий включение в каждый файл до 32000 переменных (см. том III).

Заголовок файла и комментарий

В каждом файле данных системы *STATISTICA* может быть записан однострочный заголовок длиной не более 77 символов. Кроме того, в диалоговом окне *Заголовок файла данных, примечания и Рабочая книга* для каждого файла можно ввести комментарий, длина которого не превышает 512 символов.



Диалоговое окно *Заголовок файла данных, примечания и Рабочая книга* открывается с помощью кнопки на панели инструментов или команды *Заголовок* выпадающего меню *Правка*, а также при двойном щелчке в таблице на области заголовка файла.

Печать комментариев и заголовков.

Строка заголовка файла данных и комментарий (*информация и примечания*) хранятся как часть файла и отображаются на экране вместе с таблицей данных при выборе режима *Инф. и примечания* в диалоговом окне *Печать данных*, открывающемся с помощью команды *Печать* в выпадающем меню *Файл*. Если выбрать в диалоговом окне *Параметры страницы/вывода* (см. главу 3) *краткий, средний* или *полный* тип отчета, то вместе с результатами анализа будет печататься однострочный заголовок соответствующего файла. Кроме того, в диалоговом окне *Параметры страницы/вывода* (см. главу 3) можно задать режим печати *заголовка вывода* на каждой странице.

Список файлов Рабочей книги

Некоторые файлы из списка Рабочей книги можно открывать из указанного диалогового окна или помечать для автоматического открытия (Авто) одновременно с файлом набора данных. При этом в течение всего

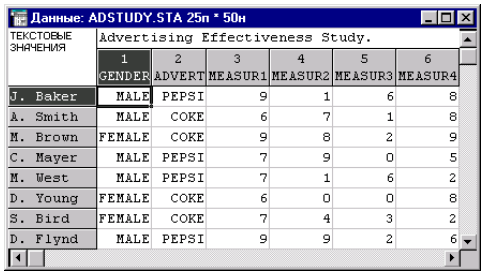
Copyright © StatSoft, 1995

StatSoft

времени, пока открыт набор данных, доступна вся дополнительная информация, например, графики или отчеты, соответствующие предыдущей ступени анализа.

Имена наблюдений

Отдельные наблюдения в файле данных (строки в таблице данных) можно помечать, присваивая им имена длиной до 20 символов.


Данные: ADSTUDY.STA 25п * 50н

ТЕКСТОВЫЕ ЗНАЧЕНИЯ	Advertising Effectiveness Study.					
	1 GENDER	2 ADVERT	3 MEASUR1	4 MEASUR2	5 MEASUR3	6 MEASUR4
J. Baker	MALE	PEPSI	9	1	6	8
A. Smith	MALE	COKE	6	7	1	8
M. Brown	FEMALE	COKE	9	8	2	9
C. Mayer	MALE	PEPSI	7	9	0	5
M. West	MALE	PEPSI	7	1	6	2
D. Young	FEMALE	COKE	6	0	0	8
S. Bird	FEMALE	COKE	7	4	3	2
D. Flynd	MALE	PEPSI	9	9	2	6

Имена наблюдений могут содержать любые печатаемые символы, в том числе пробелы и символы, использование которых недопустимо в именах переменных и текстовых значениях (см. ниже). Имена наблюдений имеют в файлах данных системы *STATISTICA* особый статус, отличный от текстовых значений. Они всегда выводятся в первом столбце таблицы, (то есть перед первой переменной) и остаются на экране при прокрутке данных в окне редактирования в горизонтальном направлении. Имена наблюдений используются по умолчанию в качестве надписей при любом выводе результатов анализа (в виде таблиц результатов или графиков), содержащих перечисление наблюдений, например содержимого кластеров при кластерном анализе, остатков при регрессионном анализе или значений факторов при факторном анализе.

Имена наблюдений вводятся в столбец имен наблюдений таблицы исходных данных в диалоговом окне *Диспетчер имен наблюдений* (см. главу 3 *Окно таблицы исходных данных*). Кроме того, их можно импортировать (например, из первого столбца файла Excel или ASCII-файла) или ввести из заданной пользователем текстовой или числовой переменной текущего файла данных.

Все операции управления именами наблюдений, а именно, создание, копирование, импортирование или изменение длины, производятся в диалоговом окне *Диспетчер имен наблюдений*, которое открывается при двойном щелчке кнопкой мыши в области имени наблюдения таблицы исходных данных. Поскольку имена наблюдений могут занимать большую часть таблицы, их отображение на экране можно включать и отключать путем нажатия кнопки на панели инструментов *Отображение имен наблюдений* окна таблицы исходных данных (см. главу 2 *Панели инструментов*).

Код пропущенных данных

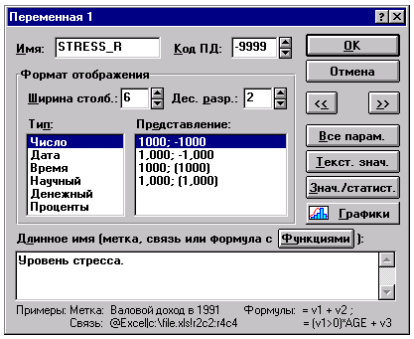
Кодом пропущенных данных является заданное числовое значение, по умолчанию равное *–9999*, указывающее те ячейки таблицы, данные в которых недоступны. В таблице исходных данных таким значениям соответствуют пустые клетки. Любой файл данных изначально содержит только пропущенные данные. Для ввода кода пропущенных данных в определенную ячейку таблицы или замены существующего значения пропущенным введите туда значение *-9999* или просто удалите содержание ячейки и оставьте ее пустой. Пользователь может устанавливать свои коды пропущенных данных в диалоговом окне *Переменная…(Изменение (текущих) спецификаций таблицы исходных данных)* (см. главу 3 *Окно таблицы исходных данных*).

StatSoft®

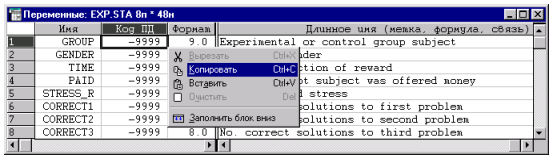
Copyright © StatSoft, 1995

Спецификации переменных

Введение. Как уже указывалось, благодаря “двойному представлению” значений в системе *STATISTICA* не существует различия между текстовыми и числовыми значениями переменных. Каждая переменная может содержать как числовые, так и текстовые значения. Переменная в наборе данных характеризуется (a) - порядковым номером, (b) - именем, длина которого не превышает 8 символов, (c) - форматом отображения и, возможно, (d) - длинным именем переменной, т.е. набором символов, задающим связь *DDE* или формулу. Ввести или изменить спецификации переменной можно в диалоговом окне *текущих спецификаций таблицы исходных данных*, которое вызывается двойным щелчком мыши на имени переменной.



Для просмотра или редактирования таблицы спецификаций всех переменных в текущем наборе данных нажмите в указанном выше диалоговом окне кнопку *Все парам.*



Ниже будут подробно рассмотрены спецификации переменных. Для объяснения понятия *Код пропущенных данных* обратитесь к предыдущему параграфу.

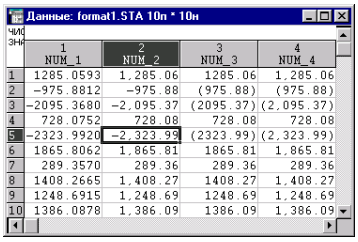
Имена переменных

Имя переменной в системе *STATISTICA* должно быть не длиннее 8 символов. Оно не может начинаться с цифры, содержать пробелы, непечатаемые символы и символы (**" # & ' ( ) * + , - . / ; : < = > [ ] ^ \ { } ~**). Программа не различает прописные и строчные буквы; все буквы в именах переменных преобразуются в прописные.

Формат отображения: Число

Числовой формат отображения следует выбирать в том случае, если значения переменных являются числовыми, в том числе и с текстовым эквивалентом. Используя параметр *Дес. разр.*, можно задавать количество десятичных разрядов, отображаемых в таблице исходных данных.

Числовое представление. Числовые значения представляются в десятичном формате с запятой (американское представление чисел) или без нее (например, *1000* или *1,000*). Отрицательные значения переменных либо имеют знак "минус" перед числом (например, *-1000*), либо заключаются в скобки (например, *(1000)*).



Существуют четыре способа представления чисел:

Copyright © StatSoft, 1995

StatSoft®

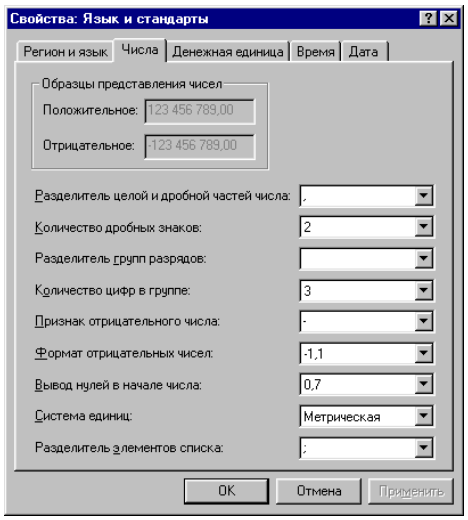
1000.00; -1000.00 (см. выше переменную *NUM_1*)

1,000.00; -1,000.00 (см. выше переменную *NUM_2)*

1000.00; (1000.00) (см. выше переменную *NUM_3*)

1,000.00; (1,000.00) (см. выше переменную *NUM_4*)

Заметим, что система *STATISTICA* поддерживает и европейское представление чисел (например, 1,20 вместо 1.20 или 1.000.000 вместо 1,000,000) в том случае, когда соответствующий стиль выбран в окне *Панель управления* Windows (см. ниже).



Если этот стиль не выбран, а вы ввели запятую вместо десятичной точки в значении переменной, то система *STATISTICA* будет рассматривать такое значение как текстовое.

Формат отображения: Дата

Значения переменных, представляющих даты, хранятся в юлианском формате, то есть в виде целого числа, равного количеству дней, прошедших с 1 января 1900 года. Например, число, введенное и отображаемое как *1/21/1968*, хранится в юлианском формате в виде *24858*; при этом возможные десятичные разряды интерпретируются как время. Записанные таким способом значения дат могут затем использоваться при проведении расчетов, например, в модуле *Анализ выживаемости* для вычисления времени жизни и преобразовываться с помощью арифметических операций. В то же время они могут отображаться на графиках или в отчетах в виде дат в обычном формате, например, использоваться в обозначениях осей. Для того чтобы переменные интерпретировались как даты (например, при импорте или вставке из таблицы Excel), нужно выбрать указанный формат. Заметим, что для отображения правильной даты соответствующие значения переменной должны быть правильными значениями-датами или явно определенными текстовыми значениями.

Способы представления дат. Выберите желаемый способ представления дат из приведенного ниже списка форматов:

3/6/87 (см. ниже переменную *DATE_1*)

03/06/1987 (см. ниже переменную *DATE_2*)

6-Mar-87 (см. ниже переменную *DATE_3*)

6-Mar (см. ниже переменную *DATE_4*)

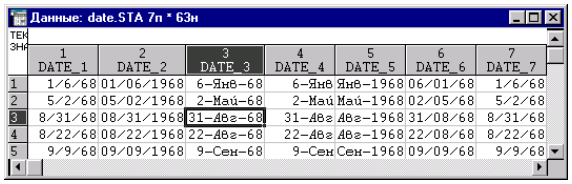
Mar-1987 (см. ниже переменную *DATE_5*)

06/03/87 (см. ниже переменную *DATE_6*)

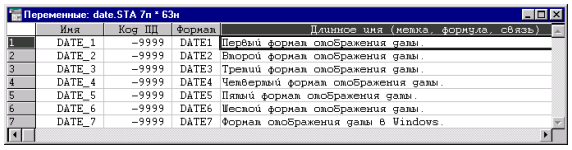
Формат Windows (см. ниже переменную *DATE_7*)

StatSoft®

Copyright © StatSoft, 1995



В окне *Переменные:…(просмотр спецификаций)*



различным способам представления *дат* соответствуют перечисленные ниже ключевые слова, которые могут быть непосредственно введены в столбец *Формат* указанного окна:

DATE1 3/6/87

DATE2 03/06/1987

DATE3 6-Mar-87

DATE4 6-Mar

DATE5 Mar-1987

DATE6 06/03/87

DATE7 Выбранный формат отображения дат на *Панели управления* Windows

Для быстрого копирования формата в заданный диапазон строк используется команда контекстного меню *Заполнить вниз*, вызываемого с помощью правой кнопки мыши.

Формат отображения: Время

Существует несколько форматов отображения времени в значениях переменных. Значения времени хранятся в данных, как правило, в виде десятичных чисел, представляющих часть дня, прошедшую после полуночи. Например, *6:00AM (6 ч утра)* представлено как *0.25*. Такое представление позволяет использовать эти значения при последующем анализе данных и преобразовывать их с помощью арифметических операций. В то же время в отчетах и при построении графиков, в частности, в обозначениях осей, они могут отображаться в общепринятом формате времени.

Способы представления времени. Ниже перечислены возможные форматы представления времени:

3/6/87 10:30 PM (см. ниже переменную *TIME_1*)

87/3/6 22:30 (см. ниже переменную *TIME_2*)

10:30 PM (см. ниже переменную *TIME_3*)

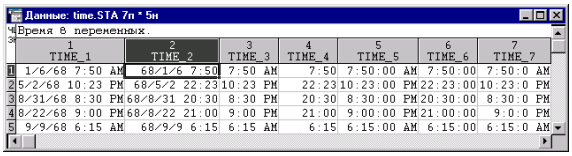
22:30 (см. ниже переменную *TIME_4*)

10:30:25 PM (см. ниже переменную *TIME_5*)

22:30:25 (см. ниже переменную *TIME_6*)

Формат Windows (см. ниже переменную *TIME_7*)

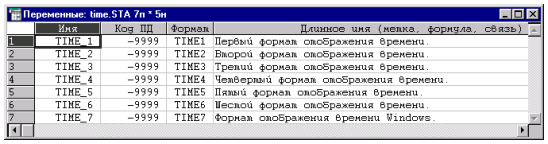
Можно выбрать комбинацию даты и времени с 24-часовым или 12-часовым (AM/PM) форматом отображения или только время с 24-часовым или 12 –часовым (AM/PM) форматом, а также представление в виде часов и минут или часов, минут и секунд.



В окне *Переменные:…(просмотр спецификаций)*

Copyright © StatSoft, 1995

StatSoft®



различным способам представления *времени* соответствуют приведенные ниже ключевые слова, которые могут быть также непосредственно введены в столбец *Формат*:

TIME1 3/6/87 10:30 PM

TIME2 87/3/6 22:30

TIME3 10:30 PM

TIME4 22:30

TIME5 10:30:25 PM

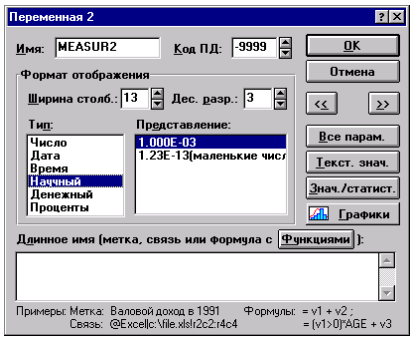
TIME6 22:30:25

TIME7 Текущий формат отображения времени, установленный на *Панели управления* Windows.

Для быстрого копирования формата в диапазон строк используется команда контекстного меню (вызываемого правой кнопкой мыши) *Заполнить вниз*.

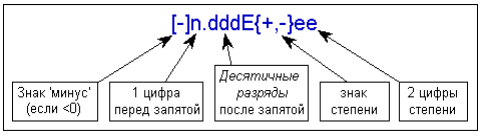
Формат отображения: Научный

Этот формат позволяет отображать значения переменных в научном представлении.



Используя параметр *Дес. разр.*, можно задавать количество десятичных разрядов, отображаемых в таблице исходных данных.

Научное представление. Первый вариант представления (стандартное научное представление, см. выше) отображает все числа в столбце в экспоненциальном виде согласно следующим правилам:



Например, *1.2345E-02* или *-3.2100E+08*.

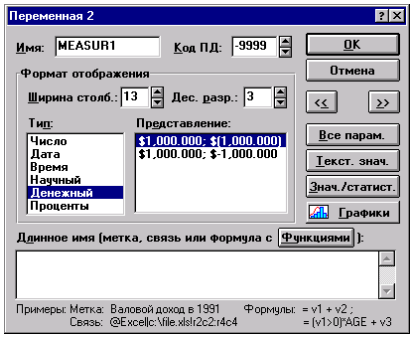
Второй вариант представления предусматривает применение экспоненциального формата только к тем значениям, которые являются достаточно малыми и в числовом формате отображаются как нули. Другими словами, этот формат применяется к числам, находящимся в интервале (-1,1), значащие разряды которых находятся после позиции, определяемой в поле редактирования *Дес. разр.*

Например, если параметр *Дес. разр. = 4*, а представляемое значение равно *0.0000123*, то в числовом формате такое число будет отображаться как *0.0000*, в то время как во втором экспоненциальном представлении оно отобразится как *1.230E-5* (перед *E* отображаются четыре цифры).

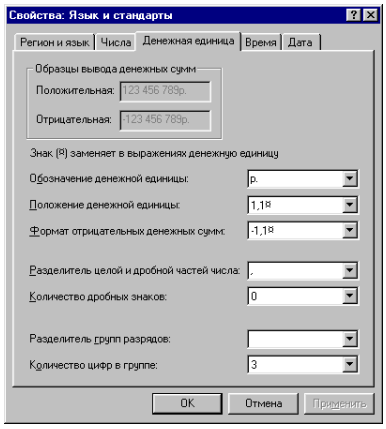
Формат отображения: Денежный

Этот тип представления выбирается для отображения значений переменных, имеющих денежный формат. Выбрав его, можно задать с помощью параметра *Дес. разр.* количество десятичных разрядов, отображаемых в таблице исходных данных.

StatSoft®
Copyright © StatSoft, 1995



Система *STATISTICA* отображает денежные значения в соответствии с установками режима *Язык и стандарты – Денежная единица Панели Управления* Windows (см. ниже).



Представление денежных величин. Перед числовыми значениями переменной, имеющей денежный формат, или после них отображаются символы денежной единицы в соответствии с конфигурацией, установленной пользователем в операционной системе, например, *$*, *DM* и т.п. (см. выше поле ввода *Символ* в диалоговом окне *Язык и стандарты – Денежная единица*). Пользователь может выбирать способ отображения отрицательных значений: либо со знаком "минус" перед числом (например, *$-1,000*), либо с помощью заключения в скобки (например, *($1,000)*)

Формат отображения: Проценты

Этот формат используется для отображения значений переменных в виде процентов. Такое представление основывается на том, что представляемое значение является долей числа 1. Например, значение *0.23* будет представлено в виде *23%*, а значение *23.4* – в виде *2340%*.

Способ представления процентов. После числа следует символ %.

Длинные имена переменных, формулы в таблице исходных данных и связи динамического обмена данными (*DDE*)

Длинные имена переменных. Помимо 8-символьного имени, отображаемого в таблице исходных данных, каждая переменная в файле данных может иметь *длинное имя*, размером до 128 символов, содержащее примечания и комментарии, включающие любые печатаемые символов. Оно может содержать также формулы, используемые для вычисления значений переменной, или спецификации связей *динамического обмена данными (DDE)* с файлами других приложений Windows.

Формулы в таблице исходных данных. Формула преобразования или перекодировки данных может быть введена непосредственно в поле *Длинное имя* диалогового окна *текущих спецификаций таблицы исходных данных*. При этом формула становится, в некотором смысле, “определением” переменной.

Эти формулы работают не так, как в обычных электронных таблицах других приложений. Формула, введенная в переменную, действует не только на одну ячейку таблицы исходных данных, а на все значения переменной (если явным образом не заданы иные условия, см. ниже). Как уже указывалось, файлы данных в системе *STATISTICA* организованы в виде

Copyright © StatSoft, 1995

записей, как в программах управления базами данных, поэтому обработка данных с помощью формул будет производиться последовательно по записям или наблюдениям.

При вводе формулы преобразования в метку переменной начните формулу со знака равенства, например:

```
= (v2+v3)/2
```

(как и в таблице Excel, первым символом в формуле должен быть знак равенства). К формуле можно добавить комментарий, которому предшествует символ ";", например:

```
= (v1+v2+v3)/3 ; это комментарий
```

Имена переменных в формулах. Вместо краткого обозначения *v#* (буква *v* и число) в формулах могут использоваться имена переменных, например:

```
= (profit1+profit2)/2
```

Имена наблюдений в формулах. Предусмотрена возможность создания значений, зависящих от номера наблюдения, обозначаемого как *v0* (*нулевая* переменная, предшествующая переменной 1). Например, формула

```
= v0+99
```

присваивает последовательным наблюдениям текущей переменной последовательные целые значения, начиная с числа 100. Заметим, что номер наблюдения может использоваться в условных выражениях (см. следующий параграф), а также для создания категориальной переменной, зависящей от последовательного расположения наблюдений в файле данных. Например, формула

```
= trunc(((v0-1)/10)+1)
```

присвоит значениям переменной последовательные целые числа, начиная с 1, причем каждое подмножество из 10 последовательных наблюдений будет иметь одно и то же значение, например, *1* будет присвоено наблюдениям от 1 до 10, *2* будет присвоено наблюдениям от 11 до 20 и т.д.

Условные выражения. Формулы в таблице исходных данных могут включать логические выражения, принимающие значение *1* при выполнении условия и *0* при его невыполнении. Таким образом, при введении логических выражений, заключенных в скобки, можно образовывать условные выражения, например:

```
= (v0<=100)*1 + (v0>100)*2
```

Приведенная формула присваивает значение *1* для наблюдений от *1* до*100* и значение *2* для наблюдений больше *100*. (Заметим, что в приведенном примере логические выражения подчеркнуты.) Логическому условию, такому как *v0>100*, присваивается значение *1*, если оно истинно (в приведенном примере - если номер наблюдения превышает *100*), и значение *0,* если оно ложно. Следовательно, для первых ста наблюдений получается выражение *1*1+0*2,* которое равно *1*, а для всех наблюдений с номерами, большими *100*, - выражение *0*1+1*2*, равное *2*.

Операторы, константы и функции. В формулах таблицы исходных данных могут использоваться операторы, константы и функции. Они являются подмножеством операторов, констант и функций языка программирования *STATISTICA BASIC* (самый последний перечень допустимых функций см. в *Электронном руководстве).* Ниже в порядке приоритета (выполнения) от самого низкого до самого высокого приведены операторы, которые могут использоваться в формулах таблицы исходных данных. Для изменения порядка выполнения операций используются скобки.

`=, <, >, <>, <=, >=`	- операторы отношения
`+ - * /`	- арифметические операторы
`or, |`	- включающая дизъюнкция
`and, &`	- конъюнкция

StatSoft®

Copyright © StatSoft, 1995

****, ^** - возведение в степень
not, ~ - отрицание

Ниже приведены константы, используемые в формулах таблицы исходных данных:

Pi - π (3.14159265358979)
Euler - *e* (2.71828182845905)

В формулы таблицы исходных данных могут входить следующие функции:

Abs(x) - Модуль *x*
Arcsin(x) - Арксинус *x*
Cos(x) - Косинус *x*
Exp(x) - Экспоненциальная функция с показателем *x*
Log(x) - Натуральный логарифм *x* (логарифм с основанием *e*)
Log2(x) - Логарифм *x* по основанию 2
Log10(x) - Десятичный логарифм *x*
Rnd(x) - Случайное действительное число, распределенное равномерно в диапазоне от *0* до *x*
Sign(x) - Знак *x* ($x<0 \rightarrow -1$; $x=0 \rightarrow 0$; $x>0 \rightarrow 1$)
Sin(x) - Синус *x*
Sinh(x) - Гиперболический синус *x*
Sqrt(x) - Квадратный корень из *x*
Trunc(x) - Целая часть *x*

Другие специализированные функции можно задать с помощью встроенного языка программирования *STATISTICA BASIC*.

Распределения и их интегралы. В формулах таблицы исходных данных могут присутствовать следующие функции распределения:

Распределение	Плотность распределения	Функция распределения	Обратная функция распределения
Beta	beta(x,ν,ω)	ibeta(x,ν,ω)	vbeta(x,ν,ω)
Биномиал.	binom(x,p,n)	ibinom(x,p,n)	
Коши	cauchy(x,η,θ)	icauchy(x,η,θ)	vcauchy(x,η,θ)

Распределения и их интегралы (продолжение)

Распределение	Плотность функции вероятности	Функция распределения	Обратная функция распределения
Хи-квадрат	chi2(x,ν)	ichi2(x,ν)	vchi2(x,ν)
Экспоненц.	expon(x,λ)	iexpon(x,λ)	vexpon(x,λ)
Экстремал.	extreme(x,a,b)	iextreme(x,a,b)	vextreme(x,a,b)
F	F(x,ν,ω)	iF(x,ν,ω)	vF(x,ν,ω)
Гамма	gamma(x,c)	igamma(x,c)	vgamma(x,c)
Геометрич.	geom(x,p)	igeom(x,p)	
Лапласа	laplace(x,a,b)	ilaplace(x,a,b)	vlaplace(x,a,b)
Логистич.	logis(x,a,b)	ilogis(x,a,b)	vlogis(x,a,b)
Лог-норм.	lognorm(x,μ,σ)	ilognorm(x,μ,σ)	vlognorm(x,μ,σ)
Нормальн.	normal(x,μ,σ)	inormal(x,μ,σ)	vnormal(x,μ,σ)
Парето	pareto(x,c)	ipareto(x,c)	vpareto(x,c)
Пуассона	poisson(x,λ)	ipoisson(x,λ)	
Релея	rayleigh(x,b)	irayleigh(x,b)	vrayleigh(x,b)
Стьюдента	student(x,df)	istudent(x,df)	vstudent(x,df)
Вейбулла	weibull(x,b,c,θ)	iweibull(x,b,c,θ)	vweibull(x,b,c,θ)

Подробная информация об указанных распределениях приведена в разделе *Приложение I*.

Случайные числа. Используя генератор случайных чисел с равномерным распределением и обратные функции распределения, можно генерировать случайные числа с распределениями, приведенными в таблице. Например, по формуле в таблице исходных данных:

```
= vweibull(rnd(1), 1, 2, 0)
```

можно получать случайные числа, удовлетворяющие распределению Вейбулла с параметром масштаба *1*, параметром формы *2* и параметром сдвига 0. Для получения более

Copyright © StatSoft, 1995

детальной информации см. Evans, Hastings, and Peacock (1993, стр. 23-24).

Пропущенные данные. Если значение какой-либо переменной в таблице исходных данных пропущено, то вычисляемой переменной будет присвоено значение пропущенных данных. В качестве примера рассмотрим следующую формулу преобразования:

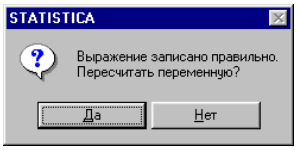
```
= v2 + v3 + v4 + v5
```

Если хотя бы одно из значений переменных *V2*, *V3*, *V4* или *V5* пропущено, то переменной, вычисляемой по этой формуле, для текущего наблюдения будет присвоено значение пропущенных данных. Заметим, что язык программирования *STATISTICA BASIC* (см. *Электронное руководство*) включает специальные функции обработки пропущенных данных, которые, например, автоматически учитывают наличие пропущенных данных при подсчете и суммировании.

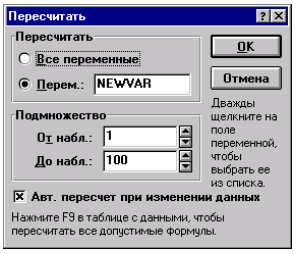
Недопустимые функции. Если нельзя провести вычисления по формуле в таблице исходных данных из-за наличия недопустимых значений, например, рассчитать квадратный корень или логарифм отрицательного числа, то соответствующей переменной присваивается значение пропущенных данных.

Деление на нуль. Практически во всех наблюдениях делению на нуль переменной приписывается значение пропущенных данных, за исключением одного случая: *деление нуля на нуль считается равным нулю* (0/0 = 0).

Пересчет таблиц исходных данных. Можно пересчитать значения соответствующей переменной сразу после ввода или изменения формулы в таблице исходных данных (нажав кнопку *Да*) или позднее (нажав кнопку *Нет*).



Для пересчета значений в соответствии с новой формулой нажмите кнопку панели инструментов, чтобы открыть диалоговое окно *Пересчитать*.



Можно также нажать клавишу F9 для пересчета всех допустимых формул в текущем наборе данных без вызова этого диалогового окна.

Авт. пересчет при изменении данных. Формула в таблице исходных данных часто зависит от значений других переменных таблицы (например, *=v1+v2*). Система *STATISTICA* находит такие взаимозависимости между переменными и в случае выбора режима *Авт. пересчет при изменении данных* в диалоговом окне *Пересчитать* (см. выше), заново вычисляет значения соответствующих переменных.

Если указанный режим не установлен, то для изменения формул в таблице исходных данных нужно нажать клавишу F9 или кнопку .

Связи *DDE* (динамический обмен данными). Связи *DDE* хранятся и устанавливаются (редактируются) с помощью параметра *Длинное имя* диалогового окна *текущих спецификаций таблицы исходных данных* (см. главу 3 *Окно таблицы исходных данных*). Для этого необходимо задать определение связи (путь *DDE*). Заметим, что для установки связи между файлами и

StatSoft®
Copyright © StatSoft, 1995

автоматической записи определений связей в диалоговом окне *текущих спецификаций таблицы исходных данных* можно также использовать команду *Установить связь* из выпадающего меню *Правка.* Дополнительная информация по этому вопросу приведена в главе 3 *Окно таблицы исходных данных*.

Ниже рассмотрены четыре элемента описания связи (например, *@Excel|c:\Excel\File.xls!r2c2:r4c4;c1;r1*).

***@Приложение-источник*|.** Сначала необходимо определить приложение-источник или "сервер" (в приведенном выше примере приложением-источником является таблица Excel). Его имя отделяют от остальной части описания связи символом "|". Необходимо также указать отдельно, является ли связь *DDE* "горячей". Для этого перед именем приложения-источника поместите символ "@", означающий, что файл-клиент должен быть обновлен немедленно после внесения изменений в файл-источник. В противном случае надо указать, что связь является "холодной", для чего поместить перед именем приложения символ "^", означающий, что файл-клиент может обновляться только по требованию пользователя или при замене пользователем символа "^" на "@".

Путь!. Путь представляет собой полное имя файла-источника (с указанием пути) и отделяется от остальной части описания связи символом "!". В приведенном выше примере параметр *Путь* представляет собой последовательность *c:\Excel\File.xls!,* то есть содержит не только имя файла *File.xls*, но также название устройства и каталога, в котором этот файл расположен (в нашем примере *c:\Excel*).

Обозначение файла-источника зависит от приложения-источника, которое вы используете. Для определения полного имени файла надо обратиться к приложению-источнику. При использовании команды *Установить связь* выпадающего меню *Правка* таблицы исходных данных, этот элемент определяется автоматически (см. главу 3).

Элемент;. Здесь определяется ячейка, диапазон ячеек, именованный диапазон ячеек или какое-либо иное имя, соответствующее приложению-источнику, за которым следует символ ";", отделяющий координаты ячеек от остальной части описания связи. Содержание этого пункта зависит от приложения-источника. При использовании команды *Установить связь* выпадающего меню *Правка* таблицы исходных данных, этот параметр определяется автоматически.

В приведенном выше примере задан диапазон ячеек таблицы Excel (например, *r2c2:r4c4* определяет блок "от строки номер 2, столбца номер 2 до строки номер 4, столбца номер 4").

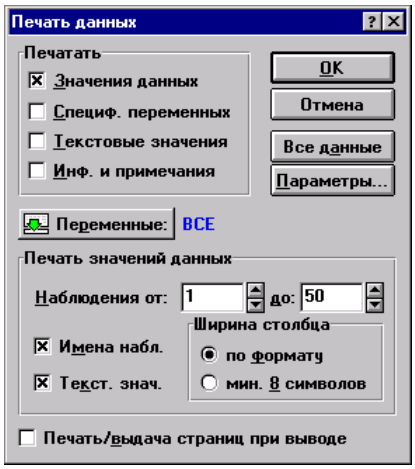
Столбец;строка. Эта часть описания связи является необязательной. По умолчанию, если не заданы начальные столбец и строка, система *STATISTICA* помещает первое значение из указанного диапазона в первую ячейку (строку) выделенной в таблице исходных данных переменной и помещает последующие значения, заполняя блок ячеек слева направо и сверху вниз от начальной точки ("точки привязки"). Если вы хотите поместить начальную точку в другое место таблицы исходных данных, укажите соответствующие номера столбца и строки, разделив их символом ";" и определив таким образом положение начальной ячейки.

Можно также указать только номер столбца (то есть переменной), тогда система *STATISTICA* использует первую ячейку (строку) в этом столбце в качестве "точки

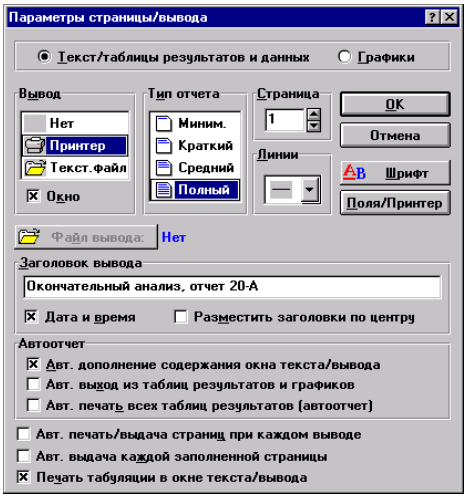
Copyright © StatSoft, 1995

привязки". Поскольку эта часть описания связи является необязательной, она не определяется автоматически при использовании команды *Установка связи*.

Печать и сохранение длинных имен переменных, формул таблиц исходных данных или связей *DDE*. Длинные имена переменных, формулы таблицы исходных данных или связи *DDE* сохраняются как часть файла и выводятся на принтер, в окно вывода или в файл вместе с таблицей, соответствующей файлу данных, в том случае, если в диалоговом окне *Печать данных* - выбран пункт *Специф. переменных*.



Длинные имена, формулы или связи *DDE* также будут выводиться одновременно с результатами анализа, если в диалоговом окне *Параметры страницы/вывода* (которое вызывается с помощью команды *Принтер* выпадающего меню *Сервис* или двойным щелчком мыши на пункте *Вывод* строки состояния) выбран *краткий, средний* или *полный* тип отчета.



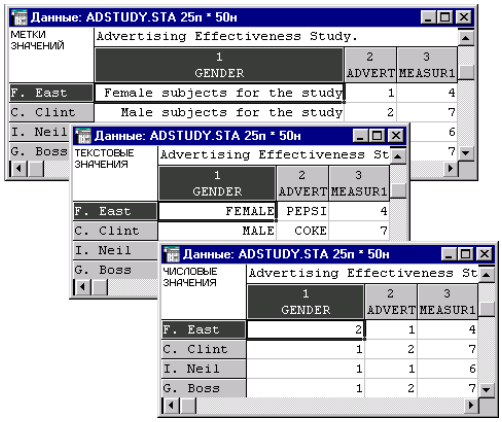
Числовые и текстовые значения

Что означает “двойное представление ” значений (текстовое и числовое)? В системе *STATISTICA* каждое значение может иметь два представления: числовое (например, *1*) и текстовое (например, *Male (Мужской)*). Такое двойное представление упрощает использование текстовых значений. Например, введя значения *1* или *2* в переменную *Gender*, можно определить пол – соответственно мужской или женский. Далее, можно ввести слово *Male (Мужской)* в любую ячейку таблицы исходных данных, содержащую *1*, и в момент окончания ввода содержимое всех ячеек в столбце, в которых было ранее записано *1,* автоматически заменится на *Male (Мужской)*. Другими словами, поскольку значение *1* не имело текстового эквивалента, программа решит, что вы поставили в соответствие значению *1* текстовое значение *Male (Мужской)*. Аналогично можно значению *2* поставить в соответствие текстовое значение *Female (Женский)*. Такое представление упрощает ввод текстовых значений. При этом, однако, сохраняется преимущество числовых значений: их по-прежнему можно использовать в

StatSoft
Copyright © StatSoft, 1995

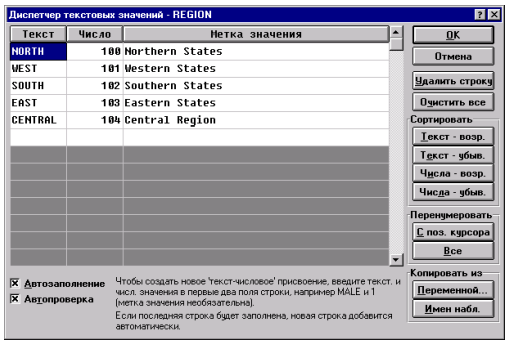
последующих численных аналитических процедурах. Кроме того, каждое текстовое значение может иметь длинную метку (см. диалоговое окно *Диспетчер текстовых значений* в главе 3 *Окно таблицы исходных данных*).

Способ переключения режимов отображения текстовых и числовых значений, а также меток значений в таблице данных. Как уже указывалось выше, *STATISTICA* поддерживает двойное представление значений, то есть каждое значение каждой переменной может иметь одновременно как числовое, так и текстовое представления. Нажимая кнопку панели инструментов таблицы исходных данных, можно переключаться между отображением текстовых и числовых значений; если при этом удерживать нажатой клавишу CTRL, то будут отображаться метки значений, если они имеются.



Способ ввода и редактирования соответствий между текстовыми и числовыми значениями. Вводить и редактировать соответствия между текстовыми и числовыми значениями можно любым из приведенных ниже способов.

- С помощью кнопки панели инструментов откройте диалоговое окно *Диспетчер текстовых значений*, в котором можно присваивать текстовые и числовые значения, а также управлять ими, то есть сортировать, избирательно переносить из других переменных и т.п. Более детальная информация приведена в главе 3.



Используя команду *Копировать из* указанного диалогового окна, можно скопировать в текущую переменную текстовые значения из других переменных или из других наблюдений. Для копирования соответствий между числовыми, текстовыми значениями и длинными метками значений из одного файла в другой используйте метод *объединения файлов* (см. стр. 1372). С помощью одной из операций объединения данных можно осуществить конкатенацию или перемещение текстовых значений из другого файла.

- С помощью кнопки установите режим отображения таблицы исходных данных - текстовый или числовой - и введите или отредактируйте соответствующие значения. Для включения режима отображения меток значений при нажатии кнопки удерживайте нажатой клавишу CTRL.
- Можно также вставить в таблицу исходных данных текстовые или числовые значения из другой переменной или из другого приложения. При этом необходимо убедиться, что режим отображения таблицы

Copyright © StatSoft, 1995

StatSoft®

исходных данных соответствует типу вставляемых значений, например, перед вводом в таблицу текстовых значений необходимо установить режим отображения текстовых значений. Заметим, что при копировании в буфер обмена блока таблицы данных содержимое скопированного блока будет зависеть от текущего режима отображения данных. Например, перед копированием текстовых значений убедитесь, что установлен режим отображения *Текстовые значения*. Единственным исключением являются операции с использованием буфера обмена, производимые внутри таблицы данных *STATISTICA*, когда установлен режим отображения текстовых значений или меток значений. В этом случае скопированный или вырезанный блок, вставляемый затем в другое место таблицы данных *STATISTICA*, всегда будет содержать полное представление выбранных данных, то есть числовые значения, текстовые значения и метки значений, если последние имеются.

- Текстовые значения в таблице данных можно создать путем импорта (см. описание диалогового окна *Быстрый импорт*, глава 3).

Импорт текстовых значений, длина которых превышает 8 символов. Если длина текстовых значений, импортируемых с использованием функций диалогового окна *Быстрый импорт* (см. главу 3), превышает 8 символов, создаются метки значений, с помощью которых длинные текстовые значения сохраняются целиком. При этом первые восемь символов хранятся также в соответствующем текстовом значении *STATISTICA*. Заметим, что при этом могут создаваться текстовые значения, неоднозначно соответствующие своим длинным меткам, то есть, возможна ситуация, когда одному и тому же текстовому значению в разных наблюдениях соответствуют разные метки. В таком случае последний символ(ы) текстовых значений заменяется цифрой, в результате чего создаются текстовые значения, однозначно соответствующие своим меткам. Например, если значения в исходном наборе данных, который нужно импортировать, представляют собой записи

```
Arizona, the south region
Arizona, the north region
Arizona, the central region,
```

то эти значения будут храниться как метки значений, а однозначно соответствующие им текстовые значения длиной 8 символов будут следующими:

```
Arizona1
Arizona2
Arizona3
```

Указанные правила действуют и в том случае, если импорт данных в таблицу данных *STATISTICA* производится с использованием буфера обмена, например, путем копирования блока значений в другую таблицу данных (выделите блок в таблице и нажмите CTRL+C) и вставьте его в таблицу данных *STATISTICA* (поместите курсор в нужное место и нажмите CTRL+V). Если длина текстовых значений превышает 8 символов, эти значения будут интерпретироваться как метки данных *STATISTICA*, а соответствующие текстовые значения будут сформированы из первых восьми символов соответствующей метки, как показано выше.

Метки значений

Метки значений представляют собой комментарии или описания и могут иметь длину до 40 символов. Они могут быть присвоены отдельных текстовым/числовым значениям в наборе данных *STATISTICA,* например:

StatSoft®

Copyright © StatSoft, 1995

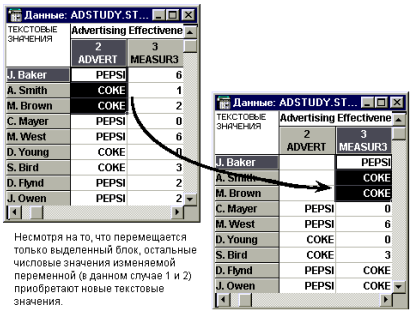
Текст	Число	Метка значения
ARIZONA1	1	Arizona, the south region
ARIZONA2	2	Arizona, the north region
ARIZONA3	3	Arizona, the central region

Метки значений могут отображаться и редактироваться непосредственно в таблице исходных данных аналогично редактированию в диалоговом окне *Диспетчер текстовых значений* (см. главу 3). Если какие-либо значения имеют метки, с помощью кнопки на панели инструментов *Отображение числовых/текстовых значений* можно переключать режим отображения между тремя установками: *Числовые значения, Текстовые значения, Метки значений*. Для отображения меток значений нажмите кнопку, удерживая нажатой клавишу CTRL. Кроме того, можно выбрать режим отображения с помощью выпадающего меню *Вид* таблицы исходных данных. Текущий режим отображения выводится в виде текстовой информации в верхнем левом углу таблицы исходных данных. Заметим, что текстовые значения и метки значений выводятся только для тех значений переменных, которые их имеют. Если в файле данных не содержится текстовых значений, то при нажатии кнопки панели инструментов или выборе пункта *Текстовые значения* выпадающего меню *Вид* появится сообщение: *В файле данных не содержится текстовых значений*.

Текстовые данные как спецификации переменных. Заметим, что сопоставление числовым значениям текстовых эквивалентов или длинных меток является частью спецификаций переменных, следовательно, текстовые значения и метки значений можно копировать, перемещать и удалять вместе с переменной. Если данные перемещаются из одной переменной в другую, например, при перемещении или копировании блока данных, сопоставляемые текстовые значения соответственно также изменяются и остаются согласованными с сопоставлениями той переменной, в которую они копируются. Другими словами, вся текстовая информация является частью переменной и может быть перемещена в другую переменную только посредством копирования текстовой информации (с использованием одного из режимов диалогового окна *Диспетчер текстовых значений*).

Пример. Если вы копируете три числовых значения *1*, *1* и *2* из одного столбца (переменной) в таблице данных в другой столбец, в котором *1=LOW (Низкий)*, а *2=MEDIUM (Средний)*, то в новом столбце значениям (*1, 1, 2*) будут соответствовать текстовые значения *LOW, LOW, MEDIUM*.

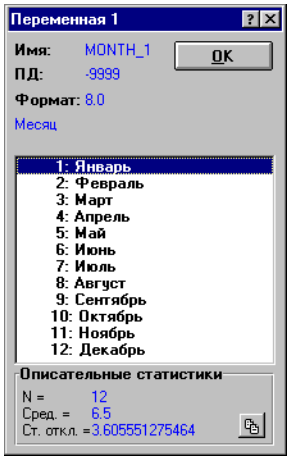
Это же правило проиллюстрировано ниже. Здесь перемещение блока значений из переменной *ADVERT* в новую переменную *MEASUR3* вызывает изменение значений новой переменной вне копируемого блока, а именно, сопоставление им текстовых значений.



Просмотр текстовых значений. Просмотр текстовых значений можно осуществить не только с помощью окна *Диспетчер текстовых значений* (см. выше), но и через диалоговое окно *Переменная…*.

Copyright © StatSoft, 1995

StatSoft®



Это окно открывается при нажатии кнопки *Знач./статист.* в диалоговом окне *текущих спецификаций таблицы исходных данных*, при выборе пункта *Быстрые статистические графики – Значения* из контекстного меню или переменной в диалоговом окне *Просмотр спецификаций переменных* (см. главу 3 *Окно таблицы исходных данных*).

Другие файлы данных *STATISTICA*

Файлы данных *Менеджера мегафайлов*. Программа *Менеджер мегафайлов* вызывается из модуля *Управление данными*. Особенности и способы применения *Менеджера мегафайлов* приведены в томе III. Краткий обзор программы см. на стр. 1338.

Другие файлы. В файлах *STATISTICA* может храниться информация различного рода. Для распознавания типа информации, хранящейся в файле, используется расширение файла. При открытии или сохранении файла вы можете либо принять расширение по умолчанию, либо задать свое расширение.

В системе *STATISTICA* используются следующие расширения по умолчанию.

- Расширение имен файлов данных таблицы исходных данных (эти файлы содержат данные, имена переменных, длинные метки, формулы, связи, имена наблюдений, текстовые значения, спецификации переменных, примечания к файлу и другую информацию): **.sta*.
- Расширение имен файлов данных таблиц результатов (выходных файлов *STATISTICA*, содержащих форматированные таблицы результатов): **.scr*.
- Расширение имен файлов с графической информацией (содержащих графики, настройки графиков, встроенные графические объекты, связи и графические данные): **.stg*.
- Расширение имен файлов с заданными пользователем графиками (представляющих собой статистические графики пользователя, добавленные в выпадающее меню *Графика*): **.sug*.
- Расширение имен файлов в формате RTF (*R*ich *T*ext *F*ormat) (таких, как RTF-файлы, созданные или используемые в *Окне текста/вывода*, см. главу 5): **.rtf*.
- Расширение имен остальных текстовых файлов, не записанных в формате RTF, а сохраненных в виде ASCII-текста (содержащих, например, условия выбора наблюдений, спецификации перекодировки или проверки данных, редактируемые последовательности нажатий клавиш и т.п.): **.txt*.
- Расширение имен файлов, написанных на языке *STATISTICA BASIC*: **.stb*.
- Расширение имен файлов, написанных на Командном языке *STATISTICA* (*SCL*): **.scl*.
- Расширение имен файлов *Менеджера мегафайлов* (см. том III): **.mfm*.

StatSoft®
Copyright © StatSoft, 1995

- Расширение имен файлов, содержащих программы на языке *MM*: **.mml*.
- Расширение имен файлов, содержащих параметры одной или нескольких осей графиков: **.axd*.
- Расширение имен файлов, содержащих условия выбора: **.sel*.
- Расширение имен файлов, содержащих описание панели инструментов *Кнопки автозадач* (содержащих списки программ на языках *SCL* и *STATISTICA BASIC* , файлов для *быстрого открытия* и редактируемые последовательности нажатий клавиш): **.atb*.
- Расширение имен файлов, содержащих макрокоманды и созданных в *STATISTICA*: **.stm*.
- Расширение имен файлов, содержащих программы модуля *SEPATH* (язык *PATH1)*: **.cmd*
- Расширение имен файлов, содержащих *Электронное руководство* системы *STATISTICA*: **.hlp*
- Расширение имен конфигурационных и инициализационных файлов *STATISTICA* (в том числе файл глобальной конфигурации системы и специальные файлы, такие как диаграммы для анализа контроля качества, определения контуров и поверхностей и т.п.): **.ini*
- Расширение имен файлов для выбора цветов палитры: **.pal*

Далее приведен список расширений имен файлов данных и графических файлов для обмена данными с другими приложениями:

- Расширение имен файлов данных CSS (и *STATISTICA/DOS*): **.css*
- Расширение имен графических файлов CSS (и *STATISTICA/DOS*): **.csg*
- Расширение имен графических метафайлов Windows: **.wmf*
- Расширение имен графических растровых аппаратно-независимых файлов Windows: **.bmp*
- Расширение имен файлов рабочих листов Excel: **.xls*
- Расширения имен файлов рабочих листов Lotus, Symphony и QuattroPro (например, **.wk1*, **.wk3*): **.wk?*, **.wr?*, **.wq?*
- Расширение имен файлов dBASE, FoxPro: **.dbf*
- Расширение имен файлов Access: **.mdb*
- Расширение имен файлов Paradox: **.db*
- Расширение имен переносимых файлов SPSS: **.por*
- Расширение имен файлов в формате ASCII (текстовых файлов в свободном формате): **.txt* or **.fre*
- Расширение имен файлов в формате ASCII (текстовых файлов в фиксированном формате): **.fix*

Формат матричных файлов данных системы *STATISTICA*

Стандартные файлы данных *STATISTICA*, рассмотренные в предыдущих разделах, - это основной тип файлов, используемый в системе для хранения данных. Помимо таких файлов в системе *STATISTICA* могут сохраняться и использоваться при последующем анализе матричные данные (например, корреляционные, ковариационные матрицы, матрицы сходства или различия, матрицы расстояний и другие). Во всех модулях *STATISTICA*, где вычисляются какие-либо матрицы (например, *Основные статистики и таблицы*, *Регрессионный анализ*,

Copyright © StatSoft, 1995

Кластерный анализ, *Факторный анализ*, *Канонический анализ* и т.п.), они могут быть сохранены в формате матричных файлов. Корреляционные матрицы могут быть сохранены также в каждом модуле *STATISTICA* с помощью средств *Быстрые основные статистики* (используйте для этого пункт меню *Быстрые основные статистики – Дополнительно* выпадающего меню *Файл* (подробная информация по этому вопросу приведена в главе 10).

Матричный файл можно также создать "вручную", вводя данные в таблицу исходных данных. При этом необходимо выполнять следующие правила:

- Количество наблюдений (строк) = количеству переменных (столбцов) + 4.
- Матрица должна быть квадратной, а имена наблюдений должны совпадать с именами переменных.
- Последние четыре наблюдения содержат следующие имена наблюдений (выделенные ниже курсивом) и следующую информацию:

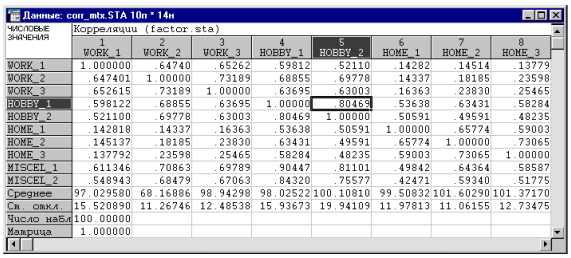
  Средние: в этой строке содержатся средние для всех переменных; для матриц *сходства* и *различия* это наблюдение должно быть оставлено пустым (т.е. в строке не должно содержаться никаких данных).

  Ст. откл.: в этой строке содержатся стандартные отклонения для всех переменных; для матриц *сходства* и *различия* это наблюдение должно быть оставлено пустым (т.е. в строке не должно содержаться никаких данных).

  N: Это число представляет собой количество наблюдений, из которых составлена данная матрица (но не число наблюдений (строк данных) в матричном файле).

  Matrix: Этот параметр определяет тип файла; 1 = *корреляционная матрица*, 2 = *матрица сходства*, 3 = *матрица различия* и 4 = *ковариационная матрица*.

Ниже приведен пример файла данных, содержащего корреляционную матрицу, вычисленную для 100 наблюдений (заметим, что тип матрицы, определяемый в последней строке и равный *1,* означает корреляционную матрицу):


Данные: corr_mtx.STA 10n * 14н

Корреляции (factor.sta)

ЧИСЛОВЫЕ ЗНАЧЕНИЯ	1 WORK_1	2 WORK_2	3 WORK_3	4 HOBBY_1	5 HOBBY_2	6 HOME_1	7 HOME_2	8 HOME_3
WORK_1	1.000000	.64740	.65262	.59812	.52110	.14282	.14514	.13779
WORK_2	.647401	1.00000	.73189	.68855	.69778	.14337	.18185	.23598
WORK_3	.652615	.73189	1.00000	.63695	.63003	.16363	.23830	.25465
HOBBY_1	.598122	.68855	.63695	1.00000	.80469	.53638	.63431	.58284
HOBBY_2	.521100	.69778	.63003	.80469	1.00000	.50591	.49591	.48235
HOME_1	.142818	.14337	.16363	.53638	.50591	1.00000	.65774	.59003
HOME_2	.145137	.18185	.23830	.63431	.49591	.65774	1.00000	.73065
HOME_3	.137792	.23598	.25465	.58284	.48235	.59003	.73065	1.00000
MISCEL_1	.611346	.70863	.69789	.90447	.81101	.49842	.64364	.58587
MISCEL_2	.548943	.68479	.67063	.84320	.75577	.42471	.59340	.51775
Среднее	97.029580	68.16886	98.94298	98.02522	100.10810	99.50832	101.60290	101.37170
Ст. откл.	15.520890	11.26746	12.48538	15.93673	19.94109	11.97813	11.06155	12.73475
Число набл	100.00000							
Матрица	1.000000							

Типы матриц

Корреляционная матрица [значение последнего наблюдения (*Тип матрицы - Matrix*): 1]. Квадратные корреляционные матрицы являются симметричными и содержат коэффициенты корреляции для всех пар переменных. Их можно создавать, сохраняя коэффициенты корреляции в соответствующем диалоговом окне, например, с помощью команды *Сохранить корреляционную матрицу* в диалоговом окне *Множественная регрессия – Описательные статистики* (см. главу12). Корреляционную матрицу можно также создать "вручную", вводя коэффициенты корреляции в обычную таблицу исходных данных и включив в эту таблицу четыре последних наблюдения, как описано выше.

Матрица сходства [значение последнего наблюдения (*Тип матрицы - Matrix*): 2]. В этой матрице хранятся коэффициенты сходства объектов (например, переменных). Матрицу сходства можно также создать "вручную", вводя коэффициенты в обычную

StatSoft

Copyright © StatSoft, 1995

таблицу исходных данных и включив туда же четыре последних строки, как описано выше. Матрицы сходства могут использоваться в модуле *Многомерное шкалирование* (см. том III).

Матрица различия [значение последнего наблюдения (*Тип матрицы - Matrix*): 3]. В этой матрице содержатся расстояния между объектами (например, переменными). Такую матрицу можно создавать "вручную" или с помощью команды *Сохранить матрицу расстояний* в диалоговом окне *Кластерный анализ – Объединение результатов* (см. том III). Такие матрицы могут использоваться также в модуле *Многомерное шкалирование* (см. том III).

Ковариационная матрица [значение последнего наблюдения (*Тип матрицы - Matrix*): 4]. Квадратные ковариационные матрицы содержат коэффициенты ковариации для всех пар переменных, а на главной диагонали - дисперсии для соответствующих переменных. Ковариационные матрицы можно создавать путем сохранения их в модуле *Моделирование структурными уравнениями* (см. том III). Эти матрицы можно также создавать "вручную", вводя ковариации в обычную таблицу исходных данных и включив в эту таблицу четыре последних строки, как описано выше.

Сохранение матричного файла

После того как матричные значения введены в соответствии с приведенным выше описанием, данные могут быть сохранены в виде стандартного файла данных *STATISTICA* с расширением **.sta* с помощью команды *Сохранить как* выпадающего меню *Файл*.

Файлы, содержащие множественные корреляционные матрицы

Некоторые модули, например, *Моделирование структурными уравнениями* (см. том III), поддерживают множественные корреляционные матрицы (например, матрицы "групповых" корреляций) в качестве входных данных. В таких матрицах каждая отдельная корреляционная матрица представляет отдельное подмножество наблюдений (групп). Создание таких матриц можно запрашивать с помощью команды *Дополнительно...* из меню *Быстрые основные статистики* (см. главу 10). При нажатии кнопки *Сохранить* открывается диалоговое окно *Сохранение матрицы*, в котором можно задать имя, под которым будет сохранена матрица. Если предварительно была выбрана группирующая переменная, формируются множественные корреляционные матрицы (т.е. матрицы "групповых корреляций"), которые сохраняются в отдельном файле.

Данные: corr_mtx.STA 4п * 16н

ЧИСЛОВЫЕ ЗНАЧЕНИЯ	1 MEASUR1	2 MEASUR2	3 MEASUR3	4 MEASUR4
MEASUR1	1.00000	-.17992	-.23335	.11451
MEASUR2	-.17992	1.00000	-.14362	.06526
MEASUR3	-.23335	-.14362	1.00000	-.14452
MEASUR4	.11451	.06526	-.14452	1.00000
Средние	6.28571	4.64286	4.32143	5.46429
Ст.откл.	2.08801	2.97165	2.93199	2.98741
Число набл.	28.00000			
Матрица	1.00000			
MEASUR1	1.00000	.20463	.00592	.33463
MEASUR2	.20463	1.00000	.06622	-.09973
MEASUR3	.00592	.06622	1.00000	.02721
MEASUR4	.33463	-.09973	.02721	1.00000
Средние	5.40909	4.40909	3.90909	5.59091
Ст.откл.	2.64861	2.83950	2.48633	2.23945
Число набл.	22.00000			
Матрица	1.00000			

Файл с множественной корреляционной матрицей, приведенный выше, получен из файла данных *Adstudy.sta* с использованием переменной *Gender (пол)* в качестве группирующей переменной. Более подробная информация приведена в главе 10.

Copyright © StatSoft, 1995

StatSoft®

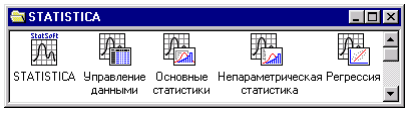
ПРИМЕР

Обзор

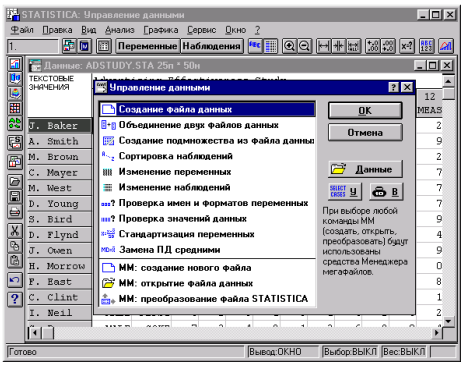
В данном разделе проиллюстрированы некоторые несложные операции управления базами данных. Для понимания этого раздела необходимо предварительно прочесть разделы *Вводный обзор* и *Формат файлов данных* данной главы, в которых вводятся элементарные представления об управлении базами данных и объясняются основные особенности файлов данных системы *STATISTICA*.

Запуск модуля *Управление данными*

Для запуска программного модуля щелкните мышью на значке *Управление данными STATISTICA* или щелчком мыши на значке *STATISTICA* вызовите *Переключатель модулей* и выберите пункт *Управление данными*.



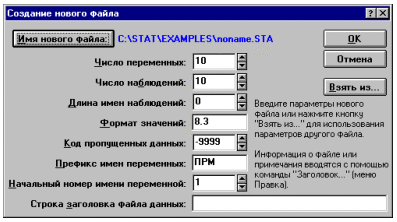
После этого запустится программа, и на экране появятся последний использовавшийся файл данных и стартовая панель модуля *Управление данными*.



Заметим, что если вы не используете *версию Quick STATISTICA*, при запуске модуля откроется также последний использовавшийся файл данных *Менеджера мегафайлов* (см. том III).

Создание нового файла данных

Выберите на стартовой панели модуля *Управление данными* (см. выше) пункт *Создание файла данных*, после чего откроется диалоговое окно *Создание нового файла*.



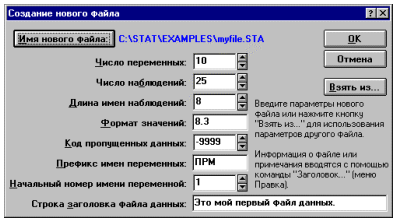
Сначала введите имя нового файла. Для этого нажмите кнопку *Имя нового файла* и введите имя файла в появившемся диалоговом окне *Создание данных: имя файла* (например, пусть это будет имя *Myfile*, возможно, с указанием нового пути). По умолчанию имени файла дается расширение **.sta*. Для создания файла данных, содержащего 25 наблюдений, 10 переменных и 8-символьные имена наблюдений, выполните последовательность шагов, приведенную ниже.

Шаг 1. Измените *число наблюдений* с *10* (устанавливаемого по умолчанию) на *25*.

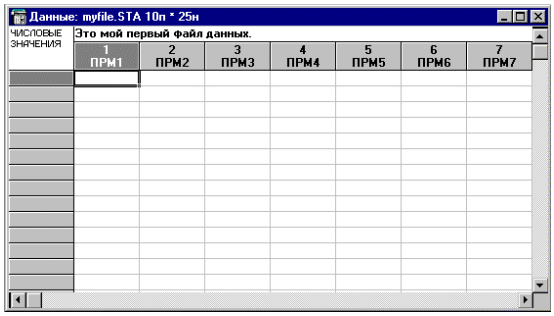
Шаг 2. Измените *длину имен наблюдений* с *0* (устанавливаемого по умолчанию) на *8*.

Шаг3. Введите в поле редактирования *Строка заголовка файла данных* однострочный (необязательный) комментарий, например, *Это мой первый файл данных*.

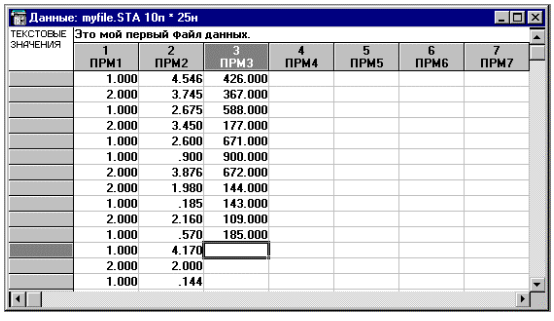
StatSoft
Copyright © StatSoft, 1995



Примите все остальные параметры, предлагаемые в диалоговом окне по умолчанию, например, *Формат значений*, *Код пропущенных данных* и т.п., и нажмите кнопку *OK*. В результате создастся новый файл данных, а на экране появится таблица исходных данных.

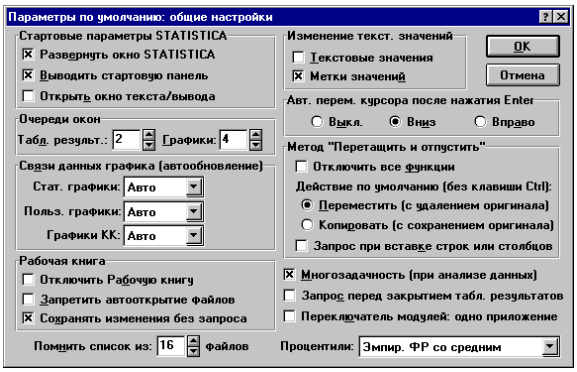


Ввод данных. Созданный вами файл является пустым, то есть он заполнен кодами пропущенных данных. Чтобы начать ввод данных, щелкните мышью на первой ячейке таблицы, находящейся на пересечении первой строки и первого столбца, имеющего имя *ПРМ_1*. Эта ячейка станет выделенной. Это означает, что ее можно редактировать.



При вводе данных в эту ячейку выделение пропадет, и ячейка будет обведена сплошной рамкой, внутри которой будет находиться курсор в виде вертикальной линии. Для записи данных в эту ячейку наберите какое-либо значение (например, 1) и нажмите клавишу *Ввод (Enter)*. При этом активной станет ячейка в следующей строке (в следующем наблюдении). Наберите и в этой ячейке какое-либо значение и нажмите *Ввод (Enter)*. Когда вы дойдете до последней строки в столбце, введете в ячейку значение и нажмете клавишу *Ввод (Enter)*, активной станет первая ячейка первой строки следующего столбца.

Автоматическое перемещение курсора. Если вы хотите вводить данные *по наблюдениям*, после ввода каждого значения вам придется нажимать клавишу *Enter* и затем клавиши управления курсором. Ввод данных можно ускорить (упростить) путем настройки перемещения курсора. Для этого выберите пункт *Общие настройки* выпадающего меню *Сервис*, после чего откроется диалоговое окно *Параметры по умолчанию: общие настройки*.

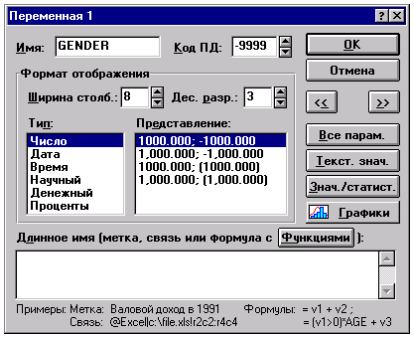


Если вы предпочитаете вводить данные по наблюдениям, а не по переменным (*вниз*), установите переключатель в диалоговом окне *Параметры по умолчанию: общие настройки* в положение *Вправо* (*Авт. перем. курсора после нажатия Enter*). Теперь при нажатии клавиши *Enter* выделение ячейки переместится к

Copyright © StatSoft, 1995
StatSoft®

следующей переменной (направо), а не к следующему наблюдению.

Имена переменных. При создании нового файла переменным по умолчанию присваиваются имена *ПРМ_1*, *ПРМ_2* и т.д. Имена переменных можно отредактировать в диалоговом окне *текущих спецификаций таблицы исходных данных* (см. ниже).

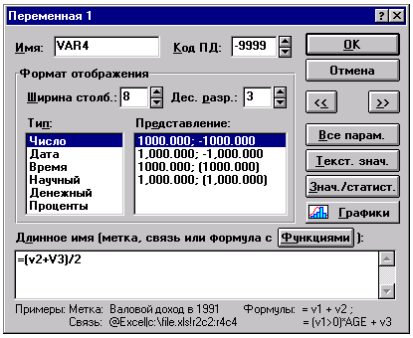


Чтобы изменить имя первой переменной на *Gender (пол)*, дважды щелкните мышью на ячейке, в которой отображается имя переменной. При этом откроется диалоговое окно *текущих спецификаций таблицы исходных данных* (см. выше). Введите в поле *Имя* этого окна имя переменной *Gender*.

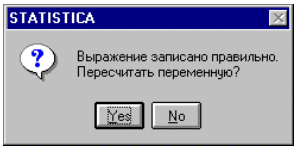
В этом же диалоговом окне можно изменить формат отображения переменной, а также ввести длинное имя или комментарий для переменной. Для этого щелкните мышью в поле *Длинное имя* и введите длинное имя переменной, например, *Это пол респондента*. Это же поле можно использовать для ввода формулы, по которой вычисляются значения переменной (см. ниже и стр. 1349), или связи динамического обмена данными (*DDE*) (см. стр. 1352).

Формулы таблицы исходных данных. Вместо описательной длинной метки в поле *Длинное имя* диалогового окна *текущих спецификаций таблицы исходных данных* (см. выше и стр. 1349) можно ввести формулу, по которой вычисляются значения данной переменной. Как и в Excel, такая формула должна начинаться со знака равенства, что проиллюстрировано далее.

При вычислении в системе *STATISTICA* значений переменной, для которой введена приведенная в примере формула, все ее значения для каждого наблюдения будут заменены на среднее арифметическое значений переменных номер 2 и номер 3. В общем случае, если вы вводите метку, которая начинается со знака равенства, программа полагает, что вместо описательной метки введена формула.



Когда вы нажмете кнопку *OK*, чтобы закрыть диалоговое окно *текущих спецификаций таблицы исходных данных*, система *STATISTICA* проверит синтаксис уравнения на формальную правильность и в случае положительного результата спросит, хотите ли вы пересчитать значения переменной в соответствии с новой формулой.

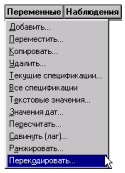


Вы можете нажать *Да* для пересчета значений переменной или *Нет* для того, чтобы отложить пересчет. В последнем случае при необходимости можно пересчитать значения переменной или всех переменных в таблице исходных данных, нажав кнопку .

StatSoft®
Copyright © StatSoft, 1995

(*Пересчитать*) на панели инструментов. Полное описание формул таблицы исходных данных приведено на стр. 1349.

Перекодировка данных. Значения переменных можно перекодировать с помощью средств перекодирования, вызвав из любого модуля *STATISTICA* выпадающее меню *Правка – Переменные* или нажав кнопку **Переменные** на панели инструментов.

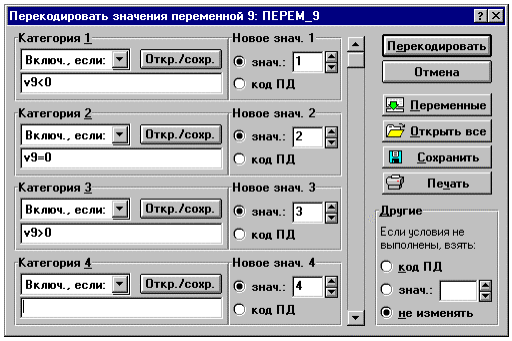


Например, для того чтобы перекодировать все отрицательные значения переменной с номером *9* в единицы (*1),* все нули – в *2*, а все положительные значения – в *3*, введите в диалоговом окне *Перекодировать значения переменной* следующие условия:

Категория 1 Включ.,если: **`v9<0`** *Новое знач. 1:* **`1`**

Категория 2 Включ.,если: **`v9=0`** *Новое знач. 2:* **`2`**

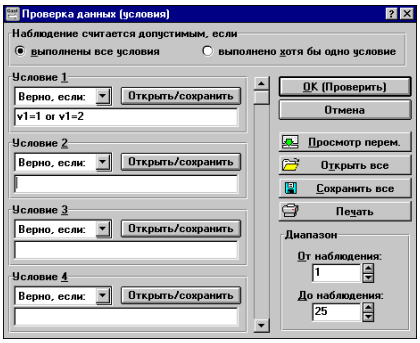
Категория 3 Включ.,если: **`v9=0`** *Новое знач. 3:* **`2`**



Когда вы нажмете кнопку *OK* в этом диалоговом окне, программа спросит вас, когда вы хотите перекодировать переменную – сейчас или позже, а также, если вы предварительно не сохранили условия перекодирования, хотите ли вы сохранить эти условия. Заметим, что пользователь может ввести в этом диалоговом окне до 16 условий перекодировки. Значительно более сложные условия перекодировки данных можно формулировать на языке программирования *STATISTICA BASIC* (см. стр. 1367 этого раздела), который содержит специальные функции перекодирования и допускает запись условий практически неограниченной сложности.

Проверка данных. В модуле *Управление данными* можно проводить интерактивную проверку и очистку данных. Эти средства используются для проверки целостности (логической непротиворечивости) и полноты данных.

Условие 1 Верно, если: **`v1=1 or v1=2`**

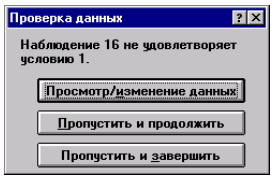


Предположим, что в приведенном выше примере переменная 1 принимает значения только *1* или *2*. Для проверки целостности данных откройте диалоговое окно *Проверка данных*, выбрав пункт *Проверить значения данных* выпадающего меню *Анализ* или стартовой панели модуля *Управление данными*. Затем введите условия проверки, как показано на рисунке.

Определив условия проверки, нажмите *OK (Проверить)*, после чего программа последовательно (по наблюдениям) проверит набор данных или выбранный диапазон данных на непротиворечивость на основании введенных вами условий проверки. В случае обнаружения наблюдения, не удовлетворяющего условиям,

Copyright © StatSoft, 1995
StatSoft®

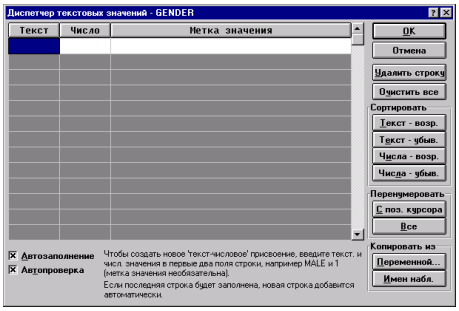
появится выделенная соответствующая строка данных и откроется диалоговое окно *Проверка данных*, в котором будет предложено либо проигнорировать противоречивость данных (и продолжить или остановить проверку), либо отредактировать (скорректировать) выделенное наблюдение.



Заметим, что пользователь может ввести до 16 условий проверки. Более полная информация по этому вопросу приведена на стр. 1380.

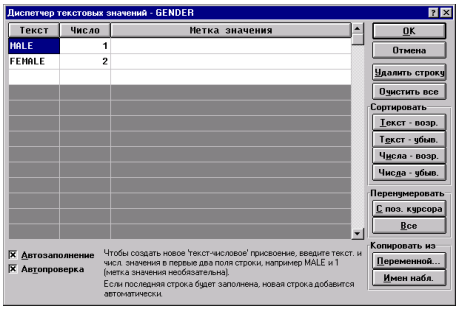
Текстовые значения. Пусть первая переменная, названная *Gender (пол),* содержит только значения 1 (обозначающие мужской пол) и 2 (обозначающие женский пол). Для облегчения считывания данных вы можете ввести для этих двух чисел текстовые (алфавитно-цифровые) эквиваленты. Самый простой способ сделать это – ввести в таблице исходных данных текстовые эквиваленты поверх числовых значений. Например, можно напечатать поверх числового значения *1* значение '*MALE*'*(Мужской)*. После того, как вы нажмете клавишу *Enter*, на месте всех значений, равных *1*, автоматически будет выведен их текстовый эквивалент '*MALE*'.

Текстовые значения и текстово-числовые соответствия можно вводить, а также управлять ими (перегруппировывать, сортировать и т.п.), используя окно *Диспетчер текстовых значений*.



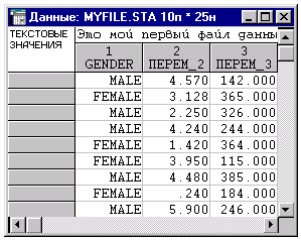
Для ввода новых значений откройте это диалоговое окно, нажав кнопку *Текстовые значения* в диалоговом окне *текущих спецификаций таблицы исходных данных*. Это окно можно вызвать и кнопкой ABC 123 панели инструментов таблицы исходных данных или с помощью команды *Переменные – Текстовые значения* выпадающего меню *Правка*. Поскольку первоначально переменной не присвоено никаких текстовых значений, столбцы будут пустыми.

В столбец с названием *Текст* введите значение *male (мужской)* (при наборе можно использовать как прописные, так и строчные буквы), а в столбце с названием *Число*, наберите значение *1*. Во второй строке наберите соответственно *female (женский)* и *2*.



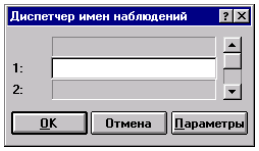
Затем нажмите кнопку *OK* или клавишу *Enter* для ввода установок и возврата в таблицу исходных данных.

StatSoft
Copyright © StatSoft, 1995



Теперь в столбце, соответствующем переменной *Gender (пол),* содержимое всех ячеек с числовым значением *1* автоматически заменилось на текстовое значение *MALE (Мужской)* (набранное прописными буквами независимо от того, какими буквами вводилось значение), а содержимое ячеек, в которых было значение *2*, - на текстовое значение *FEMALE (Женский)*.

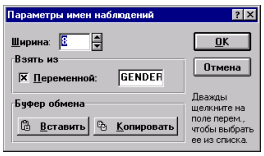
Управление именами наблюдений. При создании файла данных был предложен широкий столбец для 8-символьных имен наблюдений. Если вы создали файл, не содержащий имен наблюдений, просто передвиньте границу первого столбца направо, чтобы сделать его шире. Первоначально столбец пуст. Вы можете ввести в него имена наблюдений, используя диалоговое окно *Диспетчер имен наблюдений*, которое открывается двойным щелчком мыши в поле *Имя наблюдения* таблицы исходных данных или при выборе пункта *Наблюдения – Имена* выпадающего меню *Правка*.



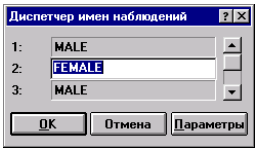
В этом диалоговом окне можно непосредственно ввести имена наблюдений в соответствующие поля или использовать в качестве имен наблюдений значения другой переменной (нажмите кнопку *Параметры*, чтобы открыть диалоговое окно *Параметры имен наблюдений*, см. ниже).

Обычно столбец с именами наблюдений содержит текстовую информацию, которая характеризует набор данных, введенных в таблицу пользователем или импортированных из другого приложения. Однако часто бывает удобно поместить в этот столбец текстовые значения переменной, чтобы эти значения присутствовали в результатах аналитических процедур, куда по умолчанию входят имена наблюдений (например, в графиках, которые строятся при анализе временных рядов).

В нашем примере скопируем значение переменной *1* (*Gender*) в имена наблюдений, установив пометку в поле *Взять из – Переменной* и введя имя переменной – *Gender (пол)* в соответствующее поле ввода.



Для редактирования имен наблюдений дважды щелкните мышью на нужном имени, после чего снова откроется диалоговое окно *Диспетчер имен наблюдений*, в поле редактирования которого будет выведено требуемое имя.



Использование языка программирования *STATISTICA BASIC* для преобразования данных

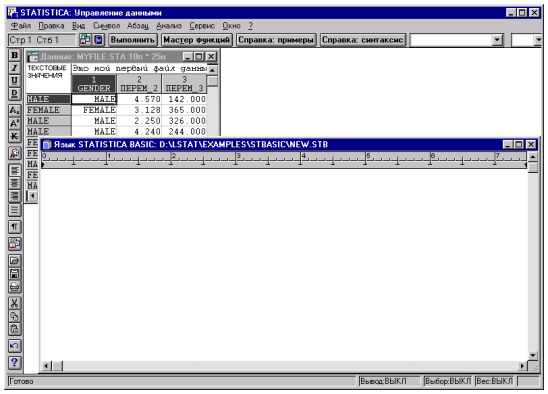
Помимо непосредственного ввода значений с помощью клавиатуры, для определения значений переменной можно использовать формулы в таблице исходных данных или встроенные средства перекодировки (см. выше).

Copyright © StatSoft, 1995

StatSoft®

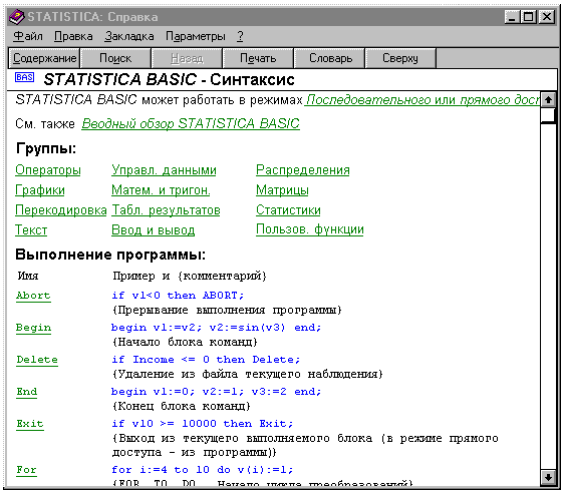
Таким способом можно выполнить подавляющее большинство стандартных преобразований и перекодировок данных в системе *STATISTICA*. Однако иногда необходимо произвести более сложные преобразования данных, например, повторить в цикле последовательность сложных формул для нескольких переменных, использовать сложные условные операторы или выполнить большое количество заданных пользователем правил перекодировки. Для этого служит язык программирования *STATISTICA BASIC*, сходный с языком *BASIC*, который может вызываться из любого модуля *STATISTICA*.

Окно *STATISTICA BASIC* проще всего открыть из выпадающего меню *Анализ* таблицы исходных данных. При выборе этого пункта меню откроется новое окно *STATISTICA BASIC*.



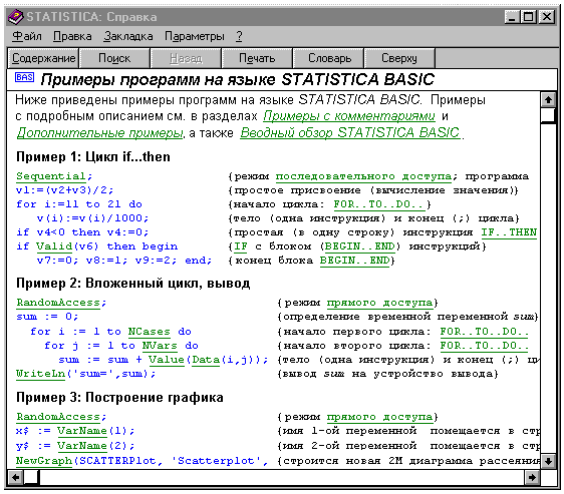
Для выбора и ввода необходимых функций используйте кнопку **Мастер функций** (см. конец настоящего раздела, стр. 1369).

На панели инструментов окна *STATISTICA BASIC* имеются также две кнопки вызова справки: **Справка: примеры** и **Справка: синтаксис**. При нажатии кнопки **Справка: синтаксис** появится справочное окно, содержащее сводную информацию о функциях и синтаксисе языка *STATISTICA BASIC*.



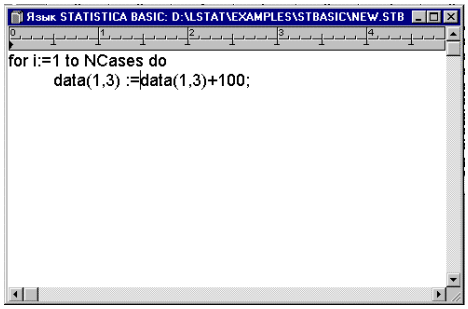
При нажатии кнопки **Справка: примеры** на панели инструментов откроется дополнительное окно *Примеры программ на языке STATISTICA BASIC* (см. ниже), в котором рассмотрены примеры небольших программ.

Заметим, что для сохранения этих справочных окон открытыми (чтобы к ним можно было быстро обратиться) при использовании программ *STATISTICA BASIC*, следует щелкнуть мышью по кнопке *Сверху* на панели инструментов окна *Справка*.



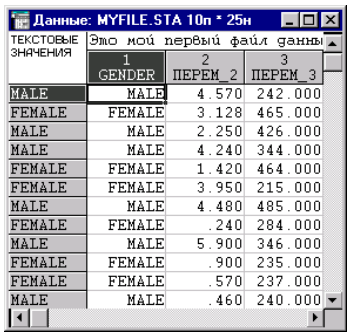
StatSoft®
Copyright © StatSoft, 1995

Введем в редактор программ *STATISTICA BASIC* несложную формулу. Прибавим, например, число 100 к каждому значению переменной номер 3, для чего организуем цикл по всем наблюдениям. Заметим, что NCases является зарезервированной переменной для количества наблюдений в текущем файле данных, а все данные заполняют массив *DATA(наблюдение, переменная)*.



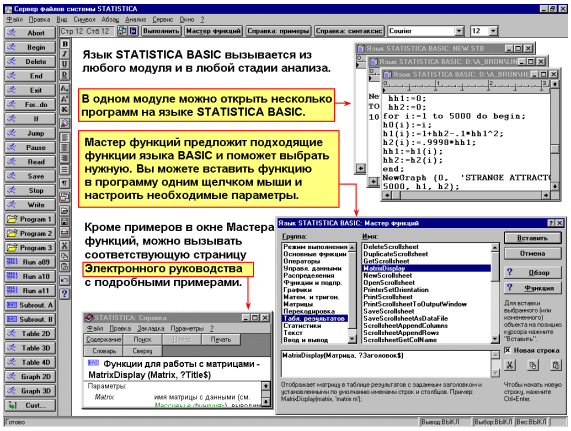
Заметим, что во всех операторах присвоения в *STATISTICA BASIC* вместо знака“ = ” используется знак “ := ”, а обычный знак равенства применяется только в логических выражениях.

Теперь запустим программу, нажав кнопку **Выполнить**. По окончании выполнения программы содержимое таблицы исходных данных автоматически изменится.



Для сохранения программы выберите команду *Сохранить как* выпадающего меню *Файл*. По умолчанию именам файлов, содержащих программы на языке *STATISTICA BASIC*, дается расширение *.stb. Файлы с программами *STATISTICA BASIC* хранятся в обычном текстовом формате.

STATISTICA BASIC - это мощный инструмент разработки программ, с помощью которого можно создавать сложные приложения с использованием графиков, таблиц результатов, пользовательских интерфейсов, операций над матрицами и т.п. В *STATISTICA BASIC* можно организовать программу в виде функций и подпрограмм. В ней можно также вызывать внешние функции, имеющиеся в библиотеках *DLL (Dynamic Link Libraries)*.



Настраиваемый мастер функций (см. выше) *STATISTICA BASIC* позволяет вызвать доступные в данный момент функции и выбрать из них нужную. Выбранную функцию можно *вставить* в программу одним щелчком мыши и затем установить те значения параметров, которые вам нужны.

Наиболее подробное описание всех возможностей языка *STATISTICA BASIC* приведено в *Электронном руководстве*.

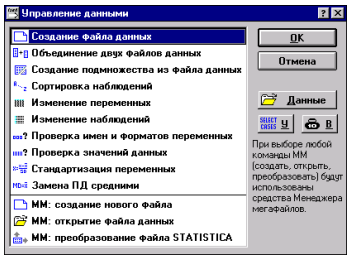
Copyright © StatSoft, 1995

StatSoft®

ДИАЛОГОВЫЕ ОКНА И МЕНЮ

Стартовая панель

Ниже показана стартовая панель модуля *Управление данными*.



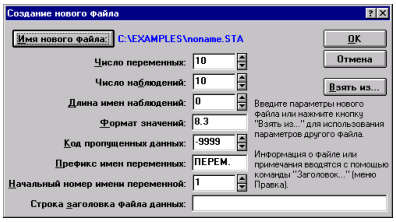
Далее приведено подробное описание этих команд.

Создание нового файла данных

В каждом модуле программы имеется меню *Файл*, с помощью которого можно легко создать новый файл данных. Параметры, которые задаются по умолчанию при открытии файла, можно впоследствии изменить. При создании файла данных в модуле *Управление данными* предлагаются дополнительные возможности, например, автоматическое создание имен переменных в соответствии с критериями, задаваемыми пользователем. Например (см. также раздел *Пример* данной главы):

Шаг 1. Выберите команду *Создать данные* выпадающего меню *Файл*, после чего откроется диалоговое окно *Создание нового файла*.

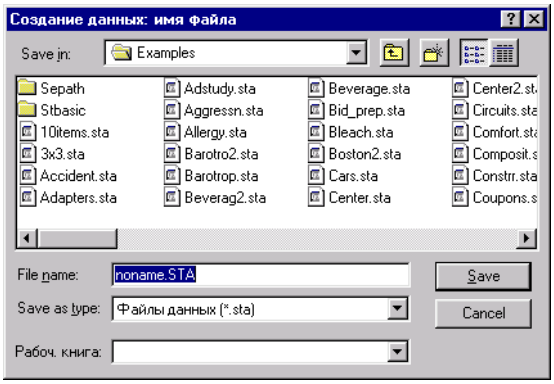
Шаг 2. Нажмите кнопку *Имя нового файла*, задайте имя, полный путь и размеры файла (количество наблюдений и переменных), а также формат значений. Кроме того, можно задать и другую информацию, необходимую для организации файла (каждый из соответствующих вариантов рассмотрен ниже). Заметим, что все спецификации файла, определенные в этом диалоговом окне, можно изменить и после создания файла.



Шаг 3. Нажмите кнопку *OK*. При этом на экране появится таблица исходных данных, в которой можно вводить данные во вновь созданный файл.

Имя нового файла

Нажмите кнопку *Имя нового файла*, после чего откроется диалоговое окно *Создание данных: имя файла*, где можно ввести название нового файла данных.



StatSoft®
Copyright © StatSoft, 1995

Число переменных

Задайте число переменных, которые необходимо ввести в новый файл данных (в одном файле данных *STATISTICA* может содержаться до 4092 переменных). Для создания файлов данных с большим количеством переменных (до 32000) воспользуйтесь средствами, предоставляемыми *Менеджером мегафайлов STATISTICA* (см. том III).

Число наблюдений

Введите число наблюдений, которые должны содержаться в новом файле данных. Число наблюдений в файлах данных *STATISTICA* практически не ограничено, т.е. в одном файле может содержаться до 2 140 000 000 наблюдений.

Длина имен наблюдений

Наблюдения в файлах данных системы *STATISTICA* могут иметь имена длиной до 20 символов. Введите здесь длину имен наблюдений в файле (если длина имени не является одинаковой для всех наблюдений, задайте наибольшую длину из используемых в данном файле).

Формат значений

По умолчанию длина отображаемого числа составляет не более 30 символов. Формат отображения задается в виде *X.Y*, где *X* – общая длина числа, а *Y* – количество знаков после десятичной точки. По умолчанию устанавливается формат *8.3*, т.е. длина отображаемого числа составляет 8 символов, а после десятичной точки отображаются 3 цифры, например, *1234.567*. Формат определяет только способ отображения значений в таблице исходных данных и никак не влияет на внутреннюю точность хранения чисел. Формат отображения можно затем изменить так, чтобы различные переменные могли иметь значения, отображаемые в экспоненциальном, текстовом, денежном формате, формате отображения дат или времени и т.п.

Код пропущенных данных

Установите значение (по умолчанию *-9999*), которое будет использоваться для представления “пропущенных данных ” во всех переменных в файле. Это значение для некоторых переменных может быть изменено в диалоговом окне *текущих спецификаций* (см. стр. 1345). Заметим, что ячейки, соответствующие пропущенным данным, в таблице исходных данных отображаются как пустые.

Префикс имен переменных

Задайте префикс, который будет входить во все имена переменных в файле данных. К концу префикса программа будет приписывать последовательные номера (например, *ПРМ1*, *ПРМ2*, *ПРМ3* и т.п.). При необходимости последние символы префикса могут заменяться цифрами для того, чтобы длина имени переменной не превышала 8 символов.

Начальный номер имени переменной

Введите начальный номер, который будет приписан к префиксу первой переменной (см. выше). Этот номер может отличаться от 1, если вы хотите объединить ваш файл с другим файлом *STATISTICA*.

Строка заголовка файла данных

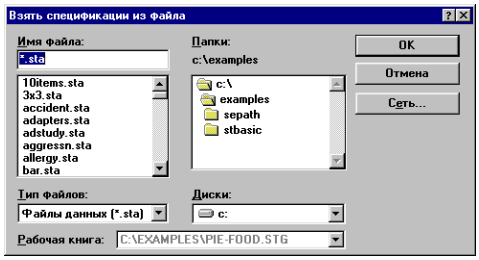
Введите строку длиной не более 77 символов, представляющую собой комментарий к файлу.

Copyright © StatSoft, 1995

Эта *строка заголовка файла* хранится как часть файла и выводится на экран или принтер вместе с таблицей исходных данных в случае выбора пункта *Инф. и примечания* в диалоговом окне *Печать данных* или вместе с результатами анализа, если в диалоговом окне *Параметры страницы/вывода* выбран *краткий, средний* или *полный* тип отчета. Для редактирования заголовка файла и добавления в него комментария (примечаний) откройте диалоговое окно *Файл данных, примечания и Рабочая книга* (см. стр. 1343), дважды щелкнув мышью в области заголовка таблицы исходных данных, или нажав кнопку *Рабочая книга* на панели инструментов.

Взять из…

Нажмите кнопку *Взять из*, после чего откроется диалоговое окно *Взять спецификации из файла*, в котором можно указать файл, содержащий полный набор спецификаций переменных (количество переменных и наблюдений, имена переменных, форматы отображения и код пропущенных данных) уже созданного файла данных *STATISTICA*, который предполагается использовать для вновь создаваемого файла.



Заметим, что спецификации переменных, введенные из файла, будут иметь приоритет над любыми спецификациями, введенными в диалоговом окне.

Объединение двух файлов данных

Часто возникает необходимость объединения файлов данных. Например, требуется добавлять данные о дополнительных участниках эксперимента к концу ("низу") существующего файла, т.е. объединять этот файл с другим файлом, содержащим те же самые переменные (см. *Объединение наблюдений*, стр. 1375). Можно также провести дополнительные эксперименты (т.е. создать дополнительные переменные) для наблюдений, содержащихся в существующем файле данных. Файл, содержащий такие новые переменные, требуется добавить к концу ("правой стороне") существующего файла данных, т.е. объединить его с другим файлом, содержащим те же наблюдения (см. ниже *Объединение переменных*). Можно также объединить соответствия числовых и текстовых значений в двух файлах (см. *Объединение текстовых значений*, стр. 1375).

Пример. Для объединения двух файлов данных *STATISTICA* выполните следующие шаги:

Шаг 1. Выберите команду *Объединить* стартовой панели модуля или выпадающего меню *Анализ*.

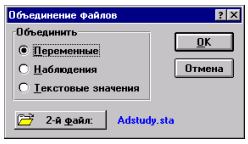
Шаг 2. Выберите конкретный способ объединения (например, *Объединить переменные*) и нажмите кнопку *2-й файл* для выбора файла с переменными, которые нужно добавить после переменных, записанных в текущем файле. Нажмите кнопку *OK* для выхода из этого диалогового окна и открытия окна *Объединить переменные* или *Объединить текстовые значения* (для режима *Объединить наблюдения* промежуточного диалогового окна не возникает).

StatSoft®

Copyright © StatSoft, 1995

Шаг 3. Выберите *режим* объединения и в случае *объединения переменных*, действие, производимое программой *при несовместимости наблюдений*. Нажмите кнопку *OK*, после чего произойдет *объединение* файлов.

При выборе команды *Объединение двух файлов данных* откроется диалоговое окно *Объединение файлов*.



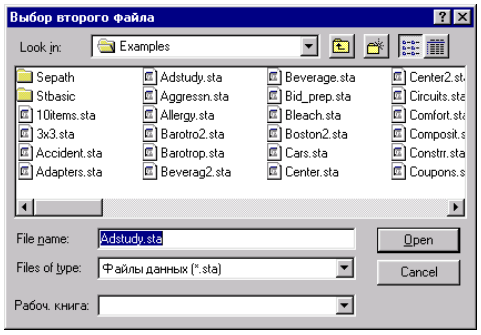
В этом окне необходимо задать тип объединения и имя файла, который требуется объединить с текущим файлом.

Объединить

Выберите один из трех методов объединения. Детальное рассмотрение каждого из них приведено ниже.

Второй файл

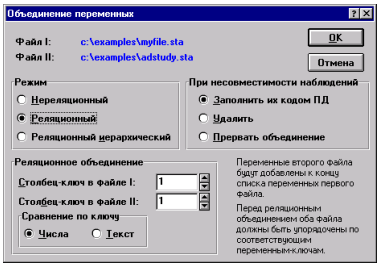
Нажмите кнопку *2-й файл*, после чего откроется диалоговое окно *Выбор второго файла*. В этом окне нужно выбрать файл, который требуется объединить с текущим файлом.



Необходимо отметить, что объединить файл с самим собой невозможно. Другими словами, нельзя выбирать в качестве второго файла текущий файл. Если же вам нужно “удвоить” файл, воспользуйтесь командой *Сохранить как* выпадающего меню *Файл* и создайте копию текущего файла.

Объединение переменных

При выборе варианта *Переменные* в диалоговом окне *Объединение файлов* система *STATISTICA* запишет переменные из второго файла после переменных первого файла (см. *Вводный обзор*, стр. 1335). Для совмещения (выравнивания) наблюдений второго файла относительно наблюдений первого файла используются следующие параметры диалогового окна *Объединение переменных*.



Режим

Если количество и порядок наблюдений в двух файлах данных совпадают не полностью, т.е. если во втором файле данных пропущено хотя бы одно наблюдение или в нем содержится наблюдение, не имеющее эквивалента в первом файле, то соответствие следующих за этим наблюдением данных будет нарушено. Чтобы избежать этой ситуации, необходимо задать режим объединения файлов.

Для реляционного объединения наблюдений в двух файлах требуется в каждом из наборов данных определить *переменную - ключ* (идентификатор наблюдения). Теперь программа

Copyright © StatSoft, 1995

будет для каждого наблюдения проверять значение этой переменной в обоих файлах данных и проводить объединение наблюдений только тогда, когда эти значения совпадают.

Нереляционный режим. При выборе этого режима переменные из второго файла будут просто добавлены к переменным первого файла. Этот режим устанавливается по умолчанию.

Реляционный режим. При выборе этого режима наблюдения из второго файла будут совмещены с соответствующими наблюдениями первого файла в соответствии со значениями *переменной – ключа* (см. ниже).

Реляционный иерархический режим. Этот режим отличается от простого реляционного объединения (см. выше) способом обработки группы записей в первом или втором файле, имеющих одинаковые значения переменной - ключа.

При простом реляционном объединении (см. выше) объединяются последовательные записи с одинаковыми значениями переменной - ключа. Если в файлах имеется неравное количество записей с одинаковыми значениями переменной - ключа, к файлу с меньшим количеством записей в нужных местах добавляются коды пропущенных данных.

При выборе *реляционного иерархического* режима объединения вместо недостающих записей вставляются значения из последней записи, имеющей то же значение переменной - ключа. Приведем пример.

Простое *реляционное объединение*, несовместимые наблюдения заполнены кодом пропущенных данных:

Файл 1		Файл 2		Объединенный файл			
Key1	Var1	Key2	Var2	Key1	Var1	Key2	Var2
1	1	1	1	1	1	1	1
2	4	1	2	Miss	Miss	1	2
2	5	1	3	Miss	Miss	1	3
3	2	2	4	2	4	2	4
3	3	2	5	2	5	2	5
				3	2	Miss	Miss
				3	3	Miss	Miss

Реляционное иерархическое объединение, несовместимые наблюдения заполнены кодом пропущенных данных:

Объединенный файл			
Key1	Var1	Key2	Var2
1	1	1	1
1	1	1	2
1	1	1	3
2	4	2	4
2	5	2	5
3	2	Miss	Miss
3	3	Miss	Miss

Заметим, что последовательность наблюдений при объединении не изменяется, следовательно, перед объединением оба файла необходимо отсортировать в соответствии со значениями переменных - ключей (см. *Сортировка данных*, стр. 1376).

Параметры реляционного объединения

При выборе режимов *Реляционное объединение* или *Реляционное иерархическое объединение* необходимо определить для каждого из объединяемых файлов переменную - ключ. В системе *STATISTICA* предполагается, что перед проведением реляционного объединения в обоих

StatSoft®

Copyright © StatSoft, 1995

файлах была предварительно произведена сортировка (см. *Сортировка данных*, стр. 1376) в соответствии со значениями переменных - ключей.

Столбец-ключ в файле 1. Введите номер переменной, которая будет использоваться в качестве переменной - ключа в первом файле.

Столбец-ключ в файле 2. Введите номер переменной, которая будет использоваться в качестве переменной – ключа во втором файле.

Сравнение по ключу (по числовым или текстовым значениям). Выберите, на основании каких значений будет проводиться сравнение переменных – ключей (*Числа* или *Текст)*.

При несовместимости наблюдений

Выберите один из трех возможных способов обработки несовместимых наблюдений при объединении файлов. Несовместимость наблюдений может возникнуть из-за неодинакового количества наблюдений в объединяемых файлах, либо если некоторые наблюдения не удовлетворяют критериям *реляционного объединения* (см. выше).

Заполнить их кодом ПД. При выборе этого варианта недостающие значения переменных для несовместимых наблюдений заполняются кодом пропущенных данных.

Прервать объединение. При наличии несовместимых наблюдений в файлах на экран выдается сообщение об ошибке и выполнение процедуры объединения прерывается.

Удалить. При выборе этого варианта несовместимые наблюдения из результирующего файла будут удалены.

Объединение наблюдений

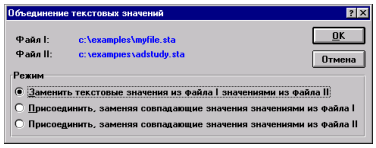
Для того чтобы наблюдения второго файла записались после наблюдений первого файла, нужно выбрать вариант *Наблюдения* в диалоговом окне *Объединение файлов* (см. *Вводный обзор*, стр. 1335). При этом выполняются следующие два правила:

- Объединение выполняется для файлов с равным количеством переменных.
- Списки текстовых значений и метки значений в файлах будут присоединены (объединены последовательно). Если переменная во втором файле имеет числовые значения, которым соответствуют текстовые значения или метки значений, отличные от значений или меток, соответствующим значениям этой же переменной в первом файле, использоваться будут текстовые значения и метки, содержащиеся в первом файле.

Объединение текстовых значений

Для объединения текстовых значений и меток значений из второго файла данных с текстовыми значениями и метками текущего файла данных выберите вариант *Текстовые значения* в диалоговом окне *Объединение файлов*. Заметим, что в этом случае объединению будут подлежать только текстовые значения переменных, имеющих в обоих файлах одинаковые имена. Если две переменные из разных файлов имеют одно и то же имя, но вы не хотите объединять их текстовые значения, временно измените имя одной из переменных, объедините файлы, а затем переименуйте переменную еще раз. При выборе этого варианта открывается диалоговое окно *Объединение текстовых значений*, в котором можно выбрать нужный режим объединения.

Copyright © StatSoft, 1995



Режим

Выберите один из приведенных вариантов объединения текстовых значений.

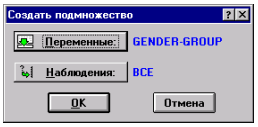
Заменить текстовые значения из файла I значениями из файла II. Все наборы текстовых значений в первом файле заменяются соответствующими наборами значений из второго файла.

Присоединить, заменяя совпадающие значения значениями из файла I. Списки текстовых значений из второго файла объединяются с соответствующими списками значений из первого файла; при этом в случае совпадающих текстовых значений используются текстовые значения из первого файла.

Присоединить, заменяя совпадающие значения значениями из файла II. Списки текстовых значений из второго файла объединяются с соответствующими списками значений из первого файла; при этом в случае совпадающих текстовых значений используются текстовые значения из второго файла.

Подмножество

В отличие от функции *Условия выбора наблюдений* (см. главу 3), позволяющей временно ограничить набор данных, подлежащих анализу, выделив из них подмножество данных, команда *Подмножество* модуля *Управление данными* позволяет выделить часть текущего файла данных, т.е. подмножество данных, содержащихся в этом файле, и сохранить его в другом файле. При выборе этой команды из выпадающего меню *Анализ* или на стартовой панели модуля *Управление данными* откроется диалоговое окно *Создать подмножество*.



После создания файла с подмножеством данных сохраните его под другим именем, используя команду *Сохранить как* выпадающего меню *Файл*.

Переменные

Нажмите кнопку *Переменные*, после чего откроется диалоговое окно *Выбор переменных*, и выберите переменные, которые нужно включить в файл с подмножеством данных.

Наблюдения

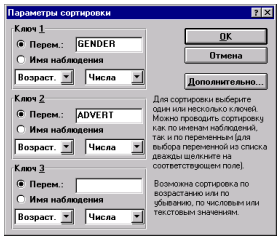
Нажмите кнопку *Наблюдения*, после чего откроется диалоговое окно *Условия выбора наблюдений*, и введите условия выбора наблюдений, которые нужно включить в файл с подмножеством данных (см. главу 1).

Сортировка

В системе *STATISTICA* имеются средства для проведения иерархической сортировки в соответствии с определяемыми пользователем переменными-ключами. Другими словами, если два наблюдения имеют одинаковые значения переменной-ключа, то для упорядочивания будет использоваться следующее значение переменной-ключа. При выборе команды *Сортировка* выпадающего меню *Анализ* или пункта *Сортировка наблюдений* стартовой панели модуля *Управление данными* открывается диалоговое окно *Параметры сортировки*.

StatSoft®

Copyright © StatSoft, 1995



Пример. В этом примере проведем сортировку файла данных *Adstudy.sta* с помощью двух переменных-ключей - *Gender* и *Advert*.

Шаг 1. Открыв диалоговое окно *Параметры сортировки* (см. выше), введите в качестве первой переменной- ключа для сортировки (*Ключ 1*) имя переменной *Gender,* а в качестве второй переменной-ключа (*Ключ 2*) – *Advert.*

Шаг 2. Для каждой переменной-ключа задайте порядок сортировки – сортировать в порядке возрастания или убывания – и объект сортировки - числовые или текстовые значения. В этом примере примем установки по умолчанию – *Возраст.* и *Числа*.

Шаг 3. Нажмите кнопку *OK* для запуска процедуры сортировки.

Пользователь может выбрать в этом диалоговом окне до семи переменных-ключей, используемых для сортировки данных. Для всех переменных-ключей имеется одинаковый набор параметров, детально рассмотренный ниже.

Переменная

Наберите в этом поле имя переменной-ключа или дважды щелкните по нему мышью, после чего откроется диалоговое окно *Выбор переменной* со списком переменных, из которого можно выбрать нужную.

Имя наблюдения

Этот параметр предназначен для сортировки данных по именам наблюдений. Заметим, что при этом для параметра нужно *Режим сортировки* нужно выбрать значение *Текст* (см. ниже).

Порядок сортировки

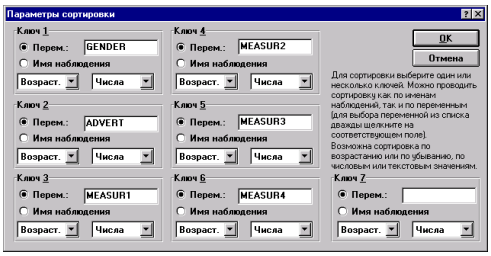
Выберите порядок сортировки – *Возраст.* или *Убыван.*

Режим сортировки

Выберите, какие значения переменной-ключа надо сортировать – *числа* или *текст.* При сортировке по текстовым значениям в случае, если некоторые числовые коды не имеют текстовых эквивалентов, *STATISTICA* полагает, что текстовое значение меньше, чем числовое. Другими словами, если сортировка производится в порядке возрастания значений, все текстовые значения окажутся в начале файла, а числовые значения, не имеющие текстовых эквивалентов, - после них.

Дополнительно

При первом вызове этого диалогового окна на экране отображаются только три ключа. Если вам нужно ввести большее количество ключей, нажмите кнопку *Дополнительно*, после чего количество отображаемых на экране ключей увеличится.



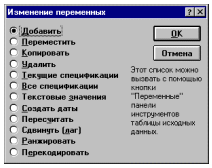
Теперь вы можете ввести до семи ключей сортировки.

Copyright © StatSoft, 1995

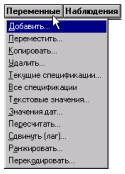
StatSoft®

Изменение переменных

Некоторые наиболее часто используемые операции управления данными можно проводить в диалоговом окне *Изменение переменных*, открывающемся из стартовой панели модуля *Управление данными*.



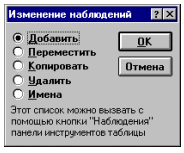
Эти же операции можно произвести, выбрав команду *Изменение переменных* выпадающего меню *Анализ*, которое вызывается при нажатии кнопки **Переменные** панели инструментов таблицы исходных данных.



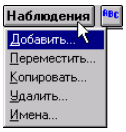
Заметим, что доступ к этому контекстному меню имеется в окне таблицы исходных данных любого модуля. Это меню вызывается также при нажатии правой кнопки мыши. Каждая из команд меню подробно рассмотрена в главе *Окно таблицы исходных данных* (глава 3).

Изменение наблюдений

Некоторые наиболее часто используемые операции управления данными можно проводить в диалоговом окне *Изменение наблюдений*, открывающемся из стартовой панели модуля *Управление данными*.



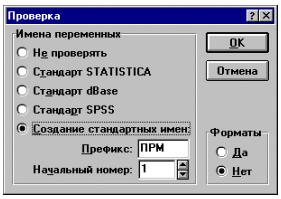
Эти же операции можно произвести, выбрав команду *Изменение наблюдений* выпадающего меню *Анализ*, которое вызывается при нажатии кнопки **Наблюдения** панели инструментов таблицы исходных данных.



Заметим, что доступ к этому контекстному меню имеется в окне таблицы исходных данных любого модуля. Это меню вызывается также при нажатии правой кнопки мыши. Каждая из команд меню подробно рассмотрена в главе *Окно таблицы исходных данных* (глава 3).

Проверка имен и формата переменных

Проверка используется для установления соответствия данных столбцу *Форматы* (см. ниже), для нахождения повторяющихся имен столбцов, а также для проверки того, удовлетворяют ли данные выбранным стандартам представления, например, *STATISTICA*, dBASE, SPSS. При выборе этой команды в выпадающем меню *Анализ* или в стартовой панели модуля *Управление данными* открывается диалоговое окно *Проверка*.



Имена переменных

Этот режим позволяет проверить соответствие имен переменных стандартам *STATISTICA*, dBASE или SPSS, а также создать имена, удовлетворяющие этим стандартам. Для этого

StatSoft®
Copyright © StatSoft, 1995

нужно выбрать один из перечисленных ниже вариантов. В случае обнаружения недопустимого или повторяющегося имени откроется диалоговое окно *Проверка имени* (см. стр. 1379), в котором можно ввести корректное имя, или, в случае повторяющегося имени, отредактировать любое из двух имен (см. ниже).

Не проверять. Выберите этот вариант в случае, если вам нужно проверить только форматы переменных.

Стандарт *STATISTICA*. Выберите этот вариант в случае, если нужно проверить все имена переменных на повторяемость, либо для проверки соответствия стандартам *STATISTICA BASIC*, либо перед экспортом файла.

Стандарт dBASE. Выберите этот вариант в случае, если надо проверить данные на соответствие стандартам dBASE. В этом случае в имена переменных могут входить латинские буквы *A-Z*, *a-z*, цифры *0-9*, а также символ подчеркивания (_). Имя переменной должно начинаться с буквы, причем имена переменных не должны повторяться. Последнее требование можно проверить, используя предыдущий вариант.

Стандарт SPSS. Выберите этот вариант в случае, если надо проверить данные на соответствие стандартам SPSS. В этом случае в имена переменных могут входить латинские буквы *A-Z*, *a-z*, цифры *0-9*, символ подчеркивания (_), точка (.), символы @, # и *$*.. Имя переменной должно начинаться с буквы и символов *$*, @, а также не должно совпадать ни с одним из зарезервированных ключевых слов SPSS (*ALL*, *AND*, *BY*, *EQ*, *GE*, *GT*, *LE*, *LT*, *NE*, *NOT*, *OR*, *TO* и *WITH*).

Создание стандартных имен

При выборе этого варианта для всех переменных создаются “стандартные” имена, состоящие из префикса и порядкового номера.

Префикс. Задайте состоящий из пяти букв префикс для стандартных имен переменных.

Начальный номер. Выберите начальный номер для числовой части имен переменных. Этот режим особенно полезен, если текущий файл нужно объединить с другим файлом *STATISTICA*.

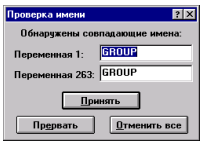
Форматы

Если вы выберете вариант *Да*, *STATISTICA* проверит форматы всех заданных переменных на соответствие значениям данных. Например, значение *1234.5* не соответствует формату *2.1* столбца, в котором оно находится.

Если же формат "длиннее", чем значения в столбце (например, значение равно 12.34 и задан формат *15.7*), длина формата будет автоматически изменена в соответствии с самым длинным значением. При этом, однако, количество десятичных разрядов после десятичной точки останется неизменным.

Повторяющиеся имена

Если во время проверки имен переменных (см. выше диалоговое окно *Проверка*) программа обнаруживает повторяющиеся имена, открывается диалоговое окно *Проверка имени*, в котором эти имена можно отредактировать.



Copyright © StatSoft, 1995

StatSoft®

Переменная 2 – Переменная 1

Повторяющиеся имена выводятся в поле редактирования сразу после номеров переменных. Пользователь может отредактировать одно или оба имени.

Принять

Для ввода отредактированных имен переменных и продолжения процедуры проверки нажмите кнопку *Принять*.

Прервать

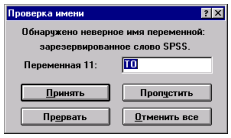
Если вы хотите прервать процедуру проверки и вернуться в модуль *Управление данными*, нажмите кнопку *Прервать*. Все изменения, сделанные до этого момента, будут сохранены.

Отменить все

Если вы хотите остановить процедуру проверки и отменить все изменения, сделанные во время этой процедуры, нажмите кнопку *Отменить все*.

Недопустимое имя переменной

Если во время проверки на корректность имен переменных (см. выше диалоговое окно *Проверка*) программа обнаруживает недопустимые имена, открывается диалоговое окно *Проверка имени*, в котором эти имена можно отредактировать.



Переменная

Недопустимое имя выводится в поле редактирования сразу после номера соответствующей переменной. Пользователь может отредактировать здесь это имя.

Пропустить

Если вы хотите оставить имя переменной неизменным, нажмите кнопку *Пропустить*.

Принять

Для ввода отредактированного имени переменной, продолжения процедуры проверки и перехода к следующему найденному недопустимому имени нажмите кнопку *Принять*.

Прервать

Если вы хотите прервать процедуру проверки и вернуться в модуль *Управление данными*, нажмите кнопку *Прервать*. Все изменения, сделанные до этого момента, будут сохранены.

Отменить все

Если вы хотите остановить процедуру проверки и отменить все изменения, сделанные во время этой процедуры, нажмите кнопку *Отменить все*.

Проверка данных (условия)

После того как данные введены или импортированы в файл, может возникнуть необходимость проверить их на целостность (логическую непротиворечивость) и полноту. Хотя некоторые типы ошибок обнаружить невозможно, существует множество способов проведения проверки данных на основании общих критериев, которым эти данные должны удовлетворять.

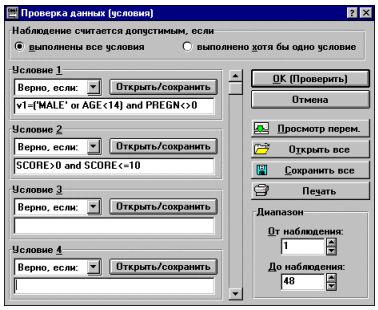
Например, во время проверки можно проверять, являются ли значения переменной допустимыми (в приводимом выше примере переменная *Gender (пол)* может принимать только значения *1*

StatSoft®
Copyright © StatSoft, 1995

и *2*) или находятся ли они в пределах допустимого диапазона значений (так, переменная *Age (Возраст)* должна быть больше *0*, но меньше *200*). Критерии проверки могут также быть очень сложными и содержать множественные логические условия, которым должны удовлетворять соотношения между значениями переменной. Вот относительно простой пример условной проверки: если объект исследования является мужчиной или лицом, не достигшим заданного возраста, количество беременностей для него не должно быть больше нуля. Запишем эти условия:

Верно, если: **(v1='MALE' or AGE<14) and PREGN<>0**

Для проверки данных на основании введенных условий выберите команду *Проверить значения данных* в выпадающем меню *Анализ*. При этом откроется диалоговое окно *Проверка данных (условия)*.



После ввода условий проверки (в соответствии с приведенными ниже правилами) нажмите *OK*, и начнется процедура последовательной проверки набора данных или заданного диапазона данных. При этом все наблюдения последовательно проверяются на полноту и соответствие заданным условиям. При обнаружении наблюдения, не удовлетворяющего условиям, программа выделяет в таблице исходных данных соответствующую строку данных и открывает диалоговое окно *Проверка данных* (см. ниже), в котором пользователь может пропустить указанное наблюдение или отредактировать (скорректировать) его, а затем продолжить или остановить проверку. После окончания редактирования для продолжения процедуры проверки выделенного наблюдения нажмите в левом нижнем углу экрана кнопку *Cont* (продолжить).

Наблюдение считается допустимым, если:

Выберите один из двух приведенных ниже вариантов для определения условия, при выполнении которого наблюдение признается допустимым.

Выполнены все условия. Если вы выбрали этот вариант, наблюдение будет признано допустимым только в случае соответствия всем условиям, приведенным ниже.

Выполнено хотя бы одно условие. При выборе этого варианта наблюдение будет признано допустимым, если оно удовлетворяет, по меньшей мере, одному из заданных условий (см. ниже).

Условие

Задайте не более 16 различных *условий проверки*, служащих критериями корректности.

Синтаксис условий проверки. Более полная информация и примеры условий выбора (проверки) приведены в разделе *Условия выбора наблюдений* главы 3.

- Для ссылок на переменные можно использовать их порядковые номера (например, *v1 = 1*) или их имена (например, *Gender = 1*). Заметим, что прописные и строчные буквы здесь эквивалентны (т.е. *GENDER* эквивалентно *gender*).
- Текстовые значения переменной заключаются в одинарные кавычки (например, *v1 = 'MALE'*). Заметим, что

Copyright © StatSoft, 1995

интерпретатор игнорирует регистр, и вы может набирать текстовые значения как прописными, так и строчными буквами (т.е. *'Yes'* эквивалентно *'YES'*).

- При задании сложных логических условий используйте скобки для изменения установленного по умолчанию порядка выполнения операторов.
- Логические операторы:

=	- равно
# , <>, ><	- не равно
<	- меньше
>	- больше
<=	- меньше или равно
>=	- больше или равно
NOT, ~	- логическое НЕ
AND, &	- логическое И
OR, !	- логическое ИЛИ

Верно/Неверно, если

Выберите, считать ли значение *верным* или *неверным*, *если* оно удовлетворяет условиям проверки.

Открыть/сохранить

При нажатии кнопки *Открыть/сохранить* откроется диалоговое окно *Условия выбора наблюдений* (см. главу 3). В отличие от команд *Открыть все* и *Сохранить все* (см. ниже), которые можно применить только для файлов, содержащих полные наборы условий проверки, эта команда либо открывает предыдущее (одно) условие выбора наблюдений (файл с расширением **.sel*), которое будет использоваться в качестве элемента текущего набора условий, либо сохраняет текущую компоненту условий проверки. Для открытия и сохранения целого набора условий используйте команды *Открыть все* и *Сохранить все* (см. ниже).

Поле ввода

Введите в поле ввода условия выбора наблюдений, которые будут использоваться в качестве критериев проверки (например, *v1 = 1 и v2 = 'MALE' или v3 = 2 и v4 = 'YES'*). Более полная информация об условиях выбора наблюдений приведена в разделе о диалоговом окне *Условия выбора наблюдений* (см. главу 3).

Просмотр переменных

При нажатии кнопки *Просмотр перем.* откроется диалоговое окно *Переменные*, в котором можно просмотреть имена и номера переменных.

Открыть все

С помощью кнопки *Открыть все* вызывается диалоговое окно *Открытие файла с условиями проверки*, в котором можно выбрать файл, содержащий множественные условия выбора наблюдений (в том числе все параметры, определенные в этом диалоговом окне) и открыть их для использования в качестве условий проверки. Условия, задаваемые здесь, совместимы с условиями, используемыми для *перекодировки* переменных (см. стр. 1333), категоризации *частот* и *графиков* (см. главу 9 *Основные статистики и таблицы* и том II настоящего руководства соответственно), а также со всеми средствами системы *STATISTICA*, позволяющими задавать множественные условия выбора наблюдений. Эту команду можно использовать для открытия сохраненного ранее файла с условиями выбора.

Сохранить все

При нажатии на кнопку *Сохранить все* откроется диалоговое окно *Сохранение условий проверки*, в котором можно задать имя файла для сохранения

StatSoft®

Copyright © StatSoft, 1995

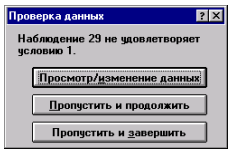
текущих условий. Условия, задаваемые здесь, совместимы с условиями, используемыми для *перекодировки* переменных (см. стр. 1333), категоризации *частот* и *графиков* (см. главу 9 *Основные статистики и таблицы* и том II настоящего руководства соответственно), а также со всеми средствами системы *STATISTICA*, позволяющими задавать множественные условия выбора наблюдений. Условия перекодирования, сохраняемые с помощью этой команды, можно затем использовать в других модулях системы *STATISTICA* в качестве условий выбора.

Диапазон

Определите здесь диапазон значений, подлежащих проверке, задав границы *От наблюдения* и *До наблюдения*.

Проверка данных

Если в процессе проверки данных (см. выше диалоговое окно *Проверка данных*) будет обнаружено наблюдение, не удовлетворяющее заданным условиям, программа выделит в таблице исходных данных соответствующую строку и откроет диалоговое окно *Проверка данных*, в котором предлагаются описанные ниже варианты.



Просмотр/изменение данных

Нажав кнопку *Просмотр/изменение данных*, вы получаете возможность изменить данные, не удовлетворяющие условиям проверки, а затем продолжить процедуру проверки, щелкнув по всплывающей кнопке *Далее*,

находящейся в нижней части окна *STATISTICA*.

Пропустить и продолжить

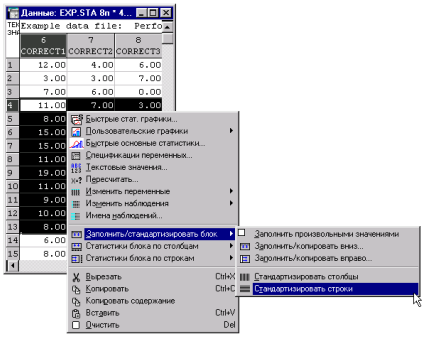
Если нажать кнопку *Пропустить и продолжить*, программа пропустит текущее наблюдение, не удовлетворяющее условиям проверки, и продолжит проверку со следующего наблюдения.

Пропустить и завершить

При нажатии кнопки *Пропустить и завершить* программа пропустит текущее наблюдение, не удовлетворяющее условиям проверки, остановит проверку и вернется в диалоговое окно *Проверка данных* (см. выше).

Стандартизация значений

Команды для стандартизации значений из выбранного (выделенного) блока содержатся в контекстном меню таблицы исходных данных, вызываемом нажатием правой кнопки мыши.



С помощью этих команд можно проводить стандартизацию произвольно выбранных подмножеств значений, а также стандартизацию

Copyright © StatSoft, 1995

StatSoft®

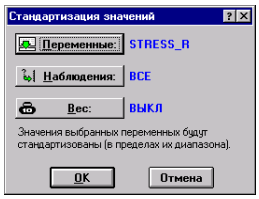
не только в пределах переменных (по столбцам), но и по строкам.

Процедура стандартизации модуля *Управление данными* отличается тем, что она распространяется не только на выделенный блок данных таблицы исходных данных, а на все значения выбранной переменной (переменных). Кроме того, данные из выбранных переменных обрабатываются здесь с учетом заданных условий выбора наблюдений и весовых коэффициентов, зависящих от наблюдений.

В процессе стандартизации все значения выбранных переменных заменяются на стандартизованные значения, вычисляемые следующим образом:

```
Ст. знач. = (Исх.знач.-Среднее)/Ст.откл.
```

Стандартное отклонение (*Ст. откл.*) в приведенной формуле вычисляется как сумма квадратов отклонений, деленная на *n-1*. При выборе этой команды в выпадающем меню *Анализ* или на стартовой панели модуля *Управление данными* открывается диалоговое окно *Стандартизация значений*, в котором пользователь может выбрать переменную для стандартизации.



Если выбранная переменная содержит текстовые значения, то соответствующие им числовые значения будут соответствующим образом изменены. В противном случае после стандартизации многие значения “потеряют” свои алфавитно-цифровые составляющие. В качестве числовых эквивалентов текстового значения используются первые шесть значащих цифр числового значения.

Переменные

По умолчанию все переменные в файле подвергаются стандартизации. Пользователь, по желанию, может задать режим стандартизации только выбранных переменных. При щелчке по клавише *Переменные* открывается диалоговое окно *Выбор переменных*, в котором можно выбрать переменные для стандартизации.

Наблюдения

По умолчанию при вычислении средних значений и стандартных отклонений для каждой переменной используются все наблюдения. Пользователь может задать условия выбора наблюдений, с использованием которых будут вычисляться средние значения и стандартные отклонения и проводиться стандартизация переменных. Таким образом, пользователь может выбрать подмножество данных для стандартизации. Так, в приведенном выше примере можно проводить стандартизацию сначала только для “*Males*”*(Мужчин),* а затем для “*Females*”*(Женщин)*, в результате чего получатся стандартизованные значения, соответствующие средним значениям и стандартным отклонениям для отдельных групп. При нажатии кнопки *Наблюдения* открывается диалоговое окно *Условия выбора наблюдений* (см. главу 3), в котором можно задать новые условия выбора наблюдений или открыть уже существующие условия.

Вес

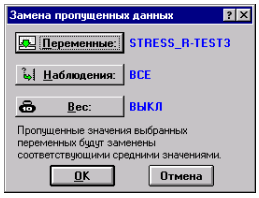
По умолчанию все (выбранные) наблюдения вносят одинаковый вклад в вычисляемые средние значения и стандартные отклонения. Нажав кнопку *Вес*, в диалоговом окне *Задание веса* можно выбрать "весовую" переменную. При этом вклад каждого наблюдения в вычисляемые значения будет прямо пропорционален значению *весовой* переменной для этого наблюдения. Хотя в действительности

StatSoft®
Copyright © StatSoft, 1995

в системе *STATISTICA* используется более быстрый алгоритм, можно представить себе, что каждое наблюдение учитывается при расчетах несколько раз, причем количество повторений задается соответствующим значением весовой переменной.

Замена пропущенных данных

При выборе этой команды пропущенные данные в выбранных переменных будут автоматически заменены средними значениями этих переменных. Эта команда также применяется во многих модулях системы *STATISTICA* в качестве одного из вариантов оперирования с пропущенными данными, однако, в таких случаях замена производится лишь на время текущего анализа. При выборе этого пункта открывается диалоговое окно *Замена пропущенных данных*.



Переменные

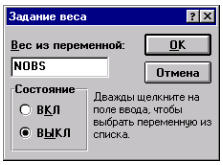
По умолчанию во всех переменных, входящих в текущий файл, пропущенные данные замещаются средними значениями. Пользователь может дать указание программе заменить пропущенные данные только в выбранных переменных. Для этого нужно нажать кнопку *Переменные* и выделить переменные в открывшемся диалоговом окне *Выбор переменных*.

Наблюдения

По умолчанию при вычислении средних значений для каждой переменной используются все наблюдения. Пользователь может изменить условия выбора наблюдений, по которым будут вычисляться средние значения. При нажатии кнопки *Наблюдения* открывается диалоговое окно *Условия выбора наблюдений* (см. главу 3), в котором можно выбрать новые или открыть уже существующие условия выбора наблюдений.

Вес

По умолчанию все (выбранные) наблюдения вносят одинаковый вклад в вычисляемые средние значения. Нажав кнопку *Вес*, в диалоговом окне *Задание веса* можно выбрать "весовую" переменную.



После выбора такой переменной вклад каждого наблюдения в вычисляемые значения будет прямо пропорционален значению *весовой* переменной для этого наблюдения. Хотя в действительности в *STATISTICA* используется более быстрый алгоритм, можно представить себе, что каждое наблюдение учитывается при расчетах несколько раз, а количество повторений задается соответствующим значением весовой переменной.

Copyright © StatSoft, 1995

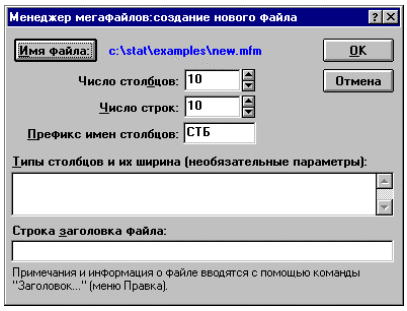
СРЕДСТВА *МЕНЕДЖЕРА МЕГАФАЙЛОВ* МОДУЛЯ УПРАВЛЕНИЕ ДАННЫМИ

Рассмотренные ниже средства *Менеджера мегафайлов* вызываются из стартовой панели модуля *Управление данными* (см. стр. 1370), выпадающего меню *Файл* и выпадающего меню *Анализ*. Заметим, что в модуле Quick *STATISTICA* для Windows нет доступа к *Менеджеру мегафайлов*.

Обзор возможностей *Менеджера мегафайлов* приведен на стр. 1338 данной главы. Наиболее полное описание *Менеджера мегафайлов* и программного языка *ММ* можно найти в томе III.

Создание новых файлов данных ММ

Для создания новых файлов в *Менеджере мегафайлов* откройте диалоговое окно *Создание нового файла*, выбрав пункт *Создать данные ММ* в выпадающем меню *Файл* либо пункт *ММ: Создание нового файла* в выпадающем меню *Анализ* или в стартовой панели модуля *Управление данными*.

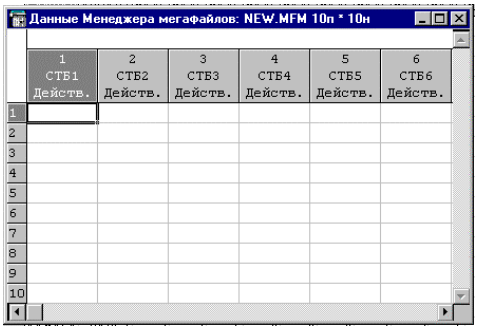


Пример. Для создания нового файла данных *Менеджера мегафайлов* выполните следующую последовательность шагов.

Шаг 1. Выберите пункт *Создать данные ММ* выпадающего меню *Файл*, после чего откроется диалоговое окно *Менеджер мегафайлов: создание нового файла* (см. выше).

Шаг 2. В открывшемся диалоговом окне задайте имя файла и путь к нему, нажав кнопку *Имя файла*, и размеры файла (количество строк и столбцов). Можно также задать формат, т.е. типы и ширину столбцов. Если вы не зададите формат, будут созданы столбцы с типом *Действительный*. Можно задать и другую информацию, необходимую для создания файла данных. Каждый из соответствующих параметров рассмотрен ниже. Заметим, что все спецификации файла, введенные в этом диалоговом окне, можно изменять после создания файла (см. том III).

Шаг 3. Нажмите кнопку *OK*. На экране появится таблица исходных данных *Менеджера мегафайлов*, в которой вы можете вводить данные в новый файл.



Имя файла

Нажмите кнопку *Имя файла* и введите в открывшемся диалоговом окне *Создать данные: имя файла* имя нового файла *Менеджера мегафайлов*.

StatSoft®
Copyright © StatSoft, 1995

Число столбцов

Задайте число столбцов в новом файле данных. Файл *Менеджера мегафайлов* может содержать до 32000 столбцов.

Число строк

Задайте число строк в новом файле данных. Файл *STATISTICA* может содержать практически неограниченное количество строк (до 2.14 миллионов).

Префикс имен столбцов

Задайте префикс имен столбцов, с которого будут начинаться имена всех столбцов в новом файле данных. Имена столбцов будут образованы путем присоединения к префиксу последовательных чисел, например, *СТБ1*, *СТБ2*, *СТБ3* и т.д.

Типы и ширина столбцов

Эти спецификации представляют собой набор инструкций для *Менеджера мегафайлов* по формированию структуры нового файла. После создания файла, дважды щелкнув мышью на заголовке любого столбца таблицы исходных данных, пользователь может изменить его формат в открывшемся диалоговом окне *Спецификации столбца* (см. том III).

Этот параметр определяет в явном виде содержание файла *Менеджера мегафайлов* для модулей *STATISTICA*. Спецификации формата вводятся в виде *nXm*, где *n* – целое число, указывающее, сколько раз формат необходимо повторить (по умолчанию полагается равным 1); *X* – тип столбца (см. следующий параграф); *m* – ширина столбца. Например, формат *10A8* означает 10 текстовых (алфавитно-цифровых) полей, длина каждого из которых равна 8 символам.

Спецификации типов столбцов

В *Менеджере мегафайлов* используются следующие спецификации типов столбцов (более полная информация по этому вопросу приведена в томе III):

A - Текстовый (алфавитно-цифровой)

F - С плавающей точкой (*R* - действительный, эквивалентный формату *Действ. 4 байта* в диалоговом окне *Спецификации столбцов*)

D - С двойной точностью (*DR* – двойной действительный, эквивалентный формату *Действ. 8 байт* в диалоговом окне *Спецификации столбцов*)

I - Целочисленный (*NI* – обычный целочисленный, эквивалентный формату *Целый 2 байта* в диалоговом окне *Спецификации столбцов*)

S - Короткий целочисленный (*SI*, эквивалентный формату *Целый 1 байт* в диалоговом окне *Спецификации столбцов*)

LI - Длинный целочисленный (*J*, эквивалентный формату *Целый 4 байта* в диалоговом окне *Спецификации столбцов*)

L - Логический

Например, формат с параметрами *2a15 i3 f5* задает четыре столбца, из которых первые два содержат текст (символ *a*), отображаемый на экране в виде 15 символов, третий столбец содержит *целочисленные* значения (символ *i*), отображаемая на экране длина которых равна 3; а четвертый столбец имеет формат *с плавающей точкой* (символ *f*) и содержит значения, отображаемая длина которых равна 5. Более

Copyright © StatSoft, 1995

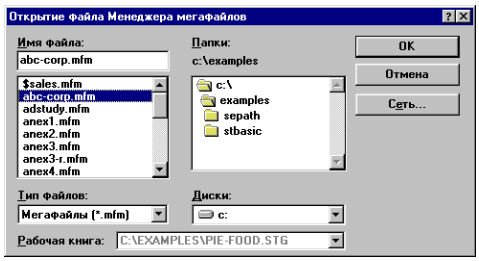
полная информация о типах столбцов и их ширине в файлах *Менеджера мегафайлов* приведена в томе III.

Строка заголовка файла

Введите строку комментария о файле длиной не более 77 символов. Более длинные комментарии и примечания можно ввести в область *Инф. и примечания* диалогового окна *Заголовок файла данных, примечания и Рабочая книга* (см. главу 3 *Окно таблицы исходных данных*).

Открытие данных *Менеджера мегафайлов*

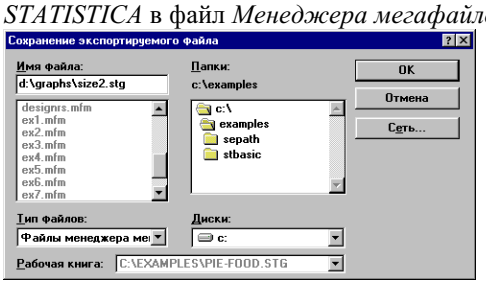
При выборе этой команды открывается диалоговое окно *Открытие файла Менеджера мегафайлов*, в котором можно выбрать файл *Менеджера мегафайлов* (с расширением **.mfm*).



Файлы Менеджера мегафайлов содержат значения данных, метки, форматы, заголовки, примечания, а также другую информацию. Чтобы открыть файл, созданный не в формате *Менеджера мегафайлов*, например, файл Excel, Lotus, Quattro, dBASE, Paradox, ASCII или иной файл, воспользуйтесь командой *Импорт* выпадающего меню *Файл* (см. том III).

Преобразование файлов *STATISTICA* в формат ММ

Данная команда содержится в выпадающем меню *Файл* (выберите пункт *Создать данные ММ*). В выпадающем меню *Анализ* также имеется пункт *ММ: Преобразовать файл STATISTICA*. Аналогичная команда имеется также в стартовой панели модуля *Управление данными*. Любым из перечисленных выше способов можно открыть диалоговое окно *Сохранение экспортируемого файла* и экспортировать или преобразовать файл данных *STATISTICA* в файл *Менеджера мегафайлов*.



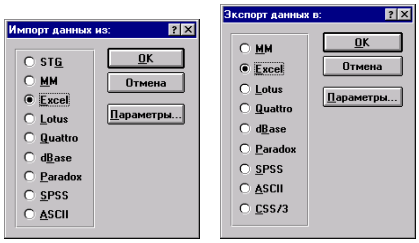
Более полная информация об экспорте файлов данных *STATISTICA* в *Менеджер мегафайлов* приведена в параграфе *Экспорт в Менеджер мегафайлов* на стр. 1408.

StatSoft®
Copyright © StatSoft, 1995

ИМПОРТ И ЭКСПОРТ ДАННЫХ

Импорт данных

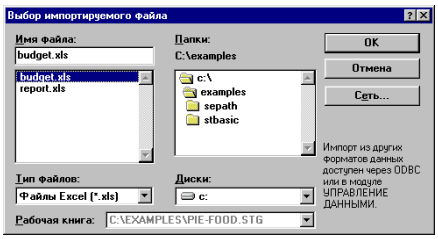
В каждом модуле системы *STATISTICA* можно использовать средства для импорта или экспорта файлов, вызываемые из выпадающего меню *Файл*. С помощью этого же меню в каждом модуле можно осуществлять доступ к различным файлам других приложений, используя интерфейс *ODBC*, в том числе файлам СУБД больших компьютеров. Наиболее полная информация по этому вопросу приведена в главе 3 и в *Электронном руководстве*. В модуле *Управление данными* имеются дополнительные возможности импорта и экспорта файлов. Для доступа к ним следует выбрать пункт *Дополнительно…* в подменю *Импорт данных* или *Экспорт данных*,



после чего программа предложит набор средств обмена данными с конкретными приложениями. Наиболее важной является возможность обмена данными, имеющими в системе *STATISTICA* двойное (текстовое и числовое) представление и длинные метки значений, с другими приложениями, не поддерживающими метки или двойное представление данных. В том числе предусмотрены возможности "маркировки" и разделения числовых и текстовых значений и их меток при экспорте, а также объединения значений и меток при импорте данных. В системе *STATISTICA* имеются также средства для доступа к файлам, записанным в сравнительно редко используемых форматах, не удовлетворяющих стандартам *ODBC*, таких как файлы в формате SPSS *PFF* (*Portable File Format*), а также к графическим файлам STG.

Пример. Импорт данных представляет собой несложную процедуру, для проведения которой выберите в выпадающем меню *Файл* модуля *Управление данными* команду *Импорт данных-Дополнительно,* и выполните перечисленные ниже действия.

Шаг 1. Выберите в диалоговом окне *Выбор импортируемого файла* имя импортируемого файла. Если необходимо, измените для этого *Тип файлов*.



Шаг 2. В диалоговом окне *Импорт данных из* подтвердите или измените тип импортируемого файла.

Шаг 3. Нажмите в указанном выше диалоговом окне на кнопку *OK*, после чего данные будут импортированы и сохранены в формате файла данных (таблицы исходных данных) системы *STATISTICA*.

Каждый из вариантов импорта поддерживает набор форматов (например, *Lotus* позволяет

Copyright © StatSoft, 1995

StatSoft®

импортировать файлы из Lotus (версии 1A, 2, 3), Symphony, Lotus 1-2-3 для Windows и т.п.). При импорте система *STATISTICA* использует имена столбцов/полей/переменных файла – источника. Если длина импортируемого имени превышает 8 символов, то в качестве имени переменной в *STATISTICA* используются первые 8 символов. Более подробная информация обо всех форматах импортируемых данных приведена ниже.

Чтобы открыть диалоговое окно с параметрами импорта, соответствующими выбранному формату, нажмите кнопку *Параметры*.

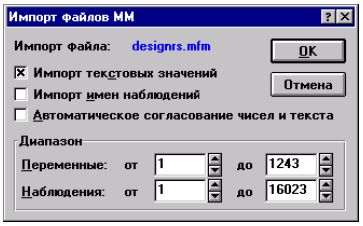
Импорт графических файлов *STATISTICA* (STG)

STG–файлы - это графические файлы системы *STATISTICA*. Если в диалоговом окне *Выбор импортируемого файла* указать тип *Графический файл STG*, то откроется диалоговое окно *Импорт данных из*, где по умолчанию будет выбран графический файл системы *STATISTICA*. Для импорта выбранного STG–файла нажмите кнопку *OK*.

Каждый график, содержащийся в STG-файле, будет импортирован в виде пары переменных, содержащих значения *X* и *Y*, или в виде тройки переменных, содержащих значения *X*, *Y* и *Z*. Если в исходном файле оси графиков не имеют названий, то импортированным переменным будут по умолчанию присвоены имена *Граф1X*, *Граф1Y*, *Граф2X, Граф2Y* и т.п. в случае двумерных графиков и *Граф1X*, *Граф1Y*, *Граф1Z, Граф2X, Граф2Y*, *Граф2Z* и т.п. в случае трехмерных графиков. Если в STG–файле на графиках имелись надписи, при импорте они будут использованы для создания имен переменных, соответствующих графикам. Если количество значений переменных, соответствующих различным графикам одного файла, не совпадает, то более "короткие" графики будут дополнены кодами пропущенных данных. Первая строка заголовка графического файла будет преобразована в строку заголовка файла данных *STATISTICA*.

Импорт файлов, созданных *Менеджером мегафайлов*

В систему *STATISTICA* можно импортировать различных типы данных, созданные *Менеджером мегафайлов* (см. том III), в том числе числовые, текстовые и логические. Если в диалоговом окне *Выбор импортируемого файла* указать тип *Файл Менеджера мегафайлов*, то откроется диалоговое окно *Импорт данных из* (по умолчанию будет выбран MFM-файл). Чтобы открыть диалоговое окно *Импорт файлов ММ* нажмите кнопку *Параметры*.



Импорт текстовых значений

При выборе этого варианта непарные столбцы текстовых значений, т.е. столбцы, не имеющие эквивалентов в виде столбцов числовых значений (см. ниже), будут импортированы в систему *STATISTICA* в “двойные переменные”, при этом всем текстовым значениям будут поставлены в соответствие числовые коды, представляющие собой последовательные целые числа. Если ширина текстового столбца превышает 8 символов, лишняя часть текстовых значений будет обрезана.

Если указанный вариант не выбран, *STATISTICA* не будет импортировать непарные текстовые столбцы (столбцы, не имеющие числовых

StatSoft®
Copyright © StatSoft, 1995

столбцов-эквивалентов), что может привести к уменьшению количества переменных.

Импорт имен наблюдений

При выборе этого варианта имена наблюдений будут импортированы из первого столбца файла, созданного *Менеджером мегафайлов*. Если же эта переменная содержит числовые, а не текстовые значения, в качестве имен наблюдений будут выступать созданные "текстовые отображения" этих значений.

Автоматическое согласование чисел и текста

Этот режим позволяет автоматически объединять столбцы файла *Менеджера мегафайлов* для образования "двойных переменных", имеющих числовые и текстовые значения (см. стр. 1325) и, возможно, метки значений.

При выборе этого режима система *STATISTICA* комбинирует столбцы для получения "двойных переменных", сопоставляя столбцу с числовыми значениями один или два текстовых столбца с соответствующими именами по следующим правилам:

- Числовому столбцу могут быть поставлены в соответствие один или два следующих непосредственно за ним текстовых столбца, которые преобразуются в текстовые значения и метки значений переменных *STATISTICA*.
- Первый текстовый столбец, который может иметь длину до 8 символов, преобразуется в текстовые значения переменной *STATISTICA*. Имя этого столбца должно совпадать с именем числового столбца, причем, если имя этого столбца короче 8 символов, к концу имени должен быть добавлен символ *$*, а если длина имени равна 8 символам, второй символ этого имени должен быть заменен на символ *$*. Заметим, что если ширина столбца превышает 8 символов, этот столбец игнорируется, и соответствующие текстовые значения не импортируются.
- Второй текстовый столбец преобразуется в метки значений. Имя этого столбца должно совпадать с именем числового столбца, причем, если имя этого столбца короче 8 символов, к концу имени должен быть добавлен символ @, а если длина имени равна 8 символам, второй символ этого имени должен быть заменен на символ @.

Между числовыми и текстовыми значениями в таких столбцах должно существовать взаимно однозначное соответствие: каждому числовому значению может соответствовать только одна текстовая запись в текстовом столбце (столбцах). В процессе создания файла данных *STATISTICA* программа проверяет наличие такого соответствия. В случае если одному числовому значению соответствует более одного текстового значения, на экран выводится предупреждение. Система *STATISTICA* импортирует также длинные имена столбцов файлов *Менеджера мегафайлов* (в виде длинных имен переменных *STATISTICA*), строки заголовка файлов данных и примечания к файлу *Менеджера мегафайлов.* При экспорте файла данных *STATISTICA* в *Менеджер мегафайлов* пары или тройки столбцов файла *Менеджера мегафайлов*, удовлетворяющие указанным выше условиям, формируются автоматически.

Диапазон импортируемых значений

По умолчанию из файла *Менеджера мегафайлов* импортируются все наблюдения и переменные. С помощью параметра *Диапазон* можно выбрать требуемый диапазон импортируемых наблюдений и/или переменных. Заметим, что

Copyright © StatSoft, 1995

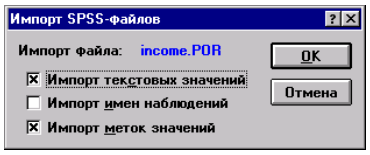
если пользователь импортирует имена наблюдений (см. выше), а в заданный диапазон переменных входит первый столбец, содержащий имена наблюдений, реально импортироваться будут столбцы, начиная со второго.

Переменные. Задайте диапазон импортируемых переменных (*от* переменной и *до* переменной включительно).

Наблюдения. Задайте диапазон импортируемых наблюдений (*от* наблюдения и *до* наблюдения включительно).

Импорт SPSS-файлов

Система *STATISTICA* может импортировать переносимые SPSS-файлы. При этом имена переменных в SPSS-файле становятся именами переменных файла системы *STATISTICA*, а длинные имена переменных SPSS-файла (длиной до 128 символов) становятся длинными именами переменных в *STATISTICA*. При нажатии кнопки *Параметры* в диалоговом окне *Импорт данных из* открывается диалоговое окно *Импорт SPSS-файлов*.



Импорт текстовых значений

С помощью этого параметра можно управлять импортом текстовых переменных из SPSS-файлов. При выборе этого режима текстовые переменные из SPSS–файлов преобразуются в "двойные переменные" системы *STATISTICA*, при этом каждому текстовому значению ставится в соответствие целое число из ряда последовательных целых чисел. Текстовые значения, длина которых превышает 8 символов, обрезаются. При этом в качестве текстовых значений будут использоваться только первые восемь символов. Если указанный режим не выбран пользователем, система *STATISTICA* не будет импортировать текстовые переменные SPSS-файлов.

Импорт имен наблюдений

Имена наблюдений могут быть импортированы из первой переменной SPSS–файла. Если эта переменная содержит числовые, а не текстовые значения, в качестве имен наблюдений будут выступать созданные "текстовые отображения" этих значений.

Импорт меток значений

При выборе этого варианта система *STATISTICA* импортирует метки значений SPSS–файла длиной до 40 символов. В случае если числовая SPSS–переменная имеет метки значений, программа создает текстовые эквиваленты (см. стр. 1325 и 1366) для числовых значений из первых восьми символов этих меток.

Пропущенные значения

Заметим, что пропущенные значения в SPSS–переменных могут определяться различными способами:

- Если в исходной переменной имеется одно пропущенное значение, то в соответствующей переменной *STATISTICA* этому значению будет соответствовать код пропущенных данных.
- Если в переменной SPSS–файла целый диапазон значений определен как пропущенные данные, то *STATISTICA* импортирует текущие значения данных (для получения максимального количества

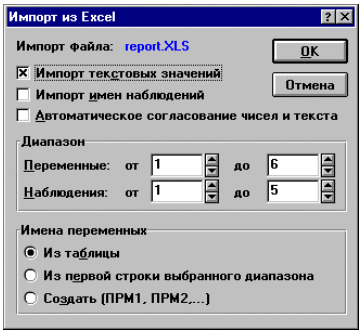
StatSoft®
Copyright © StatSoft, 1995

информации из файла-источника) и в дальнейшем рассматривает первое значение из списка пропущенных значений как код пропущенных данных.

Импорт файлов Excel

Система *STATISTICA* может импортировать из Excel числовые и текстовые значения, а также даты. Исходные текстовые значения автоматически конвертируются в текстовые значения “двойных переменных” системы *STATISTICA*. Если длина текстового значения в таблице Excel превышает 8 символов, то система *STATISTICA* в качестве текстового значения будет использовать только первые 8 символов. Импортируемые из Excel значения-даты переводятся в юлианский формат, т.е. количество дней, прошедших с 1 января 1900 года. Так, например, дата 1/21/1968 будет преобразована в число 24858. Для того чтобы даты отображались на экране и в таблицах результатов не в юлианском, а в общепринятом формате, измените формат отображения переменных в диалоговом окне *текущих спецификаций таблицы исходных данных* со значения *Число* на значение *Дата*. Более подробная информация об использовании дат в системе *STATISTICA* приведена в разделах *Даты* и *Операции над датами* главы 3.

Заметьте, что при импорте система *STATISTICA* автоматически преобразует таблицу исходных данных Excel в “базу данных”, поэтому нет необходимости перед импортом сохранять файл Excel как “базу данных”. В приведенном ниже диалоговом окне *Импорт из Excel* предлагается набор параметров для импорта данных.



Импорт текстовых значений

При выборе этого режима непарные столбцы текстовых значений, т.е. столбцы, не имеющие эквивалентов в виде столбцов числовых значений (см. ниже), будут импортированы в *STATISTICA* в “двойные переменные”, при этом текстовым значениям будут поставлены в соответствие числовые значения, представляющие собой последовательные целые числа. Если указанный режим не выбран, *STATISTICA* не будет импортировать непарные текстовые столбцы.

Импорт имен наблюдений

Имена наблюдений могут быть импортированы из первого столбца заданного пользователем диапазона данных в файле Excel. Если эта переменная содержит числовые, а не текстовые значения, в качестве имен наблюдений будут выступать созданные “текстовые отображения” этих значений.

Автоматическое согласование чисел и текста

Этот режим позволяет автоматически объединять столбцы файла Excel для образования “двойных переменных ” *STATISTICA*, имеющих числовые и текстовые значения и, возможно, метки значений. При

Copyright © StatSoft, 1995

выборе этого режима *STATISTICA* комбинирует столбцы файла Excel для получения "двойных переменных", сопоставляя столбцу с числовыми значениями один или два текстовых столбца с соответствующими именами по следующим правилам (заметим, что этот способ именования столбцов по умолчанию используется при экспорте в программу Excel файлов данных из модуля *Управление данными*):

- Числовому столбцу могут быть поставлены в соответствие один или два следующих непосредственно за ним текстовых столбца, которые преобразуются в текстовые значения и метки значений переменных системы *STATISTICA*.
- Первый текстовый столбец, который может иметь длину до 8 символов, преобразуется в текстовые значения переменной системы *STATISTICA*. Имя этого столбца должно совпадать с именем числового столбца, причем, если имя этого столбца короче 8 символов, к концу имени должен быть добавлен символ *$*, а если длина имени равна 8 символам, второй символ этого имени должен быть заменен на символ *$*. Заметим, что если ширина столбца превышает 8 символов, этот столбец игнорируется, и соответствующие текстовые значения не импортируются.
- Второй текстовый столбец преобразуется в метки значений. Имя этого столбца должно совпадать с именем числового столбца, причем если имя этого столбца короче 8 символов, к концу имени должен быть добавлен символ @, а если длина имени равна 8 символам, то второй символ этого имени должен быть заменен на символ @.

Между числовыми и текстовыми значениями в таких столбцах должно существовать взаимно однозначное соответствие: каждому числовому значению может соответствовать только одна текстовая запись в текстовом столбце (столбцах). В процессе создания файла данных *STATISTICA* программа проверяет наличие такого соответствия. В случае если одному числовому значению соответствует более одного текстового значения, на экран выводится предупреждение.

Диапазон импортируемых значений

По умолчанию импортируется вся активная область таблицы исходных данных. Параметр *Диапазон* позволяет выбрать требуемый диапазон импортируемых наблюдений и/или переменных.

Переменные. Задайте диапазон импортируемых переменных (*от* переменной и *до* переменной включительно).

Наблюдения. Задайте диапазон импортируемых наблюдений (*от* наблюдения и *до* наблюдения включительно).

Имена переменных

С помощью этого параметра можно выбрать, каким образом из данных файла Excel будут сформированы имена переменных в файле *STATISTICA*.

Из таблицы. При выборе этого варианта в качестве имен переменных в *STATISTICA* будут использоваться указатели столбцов Excel (A, B, C и т. д.).

Из первой строки выбранного диапазона. При выборе этого варианта в качестве имен переменных используются метки из первой строки выбранного диапазона (см. выше). Метки длиной более 8 символов обрезаются; для создания имен переменных используются только 8 первых символов.

Создать (*ПРМ1*, *ПРМ2*, . . .). При использовании этого варианта в импортированном файле создаются имена

StatSoft®

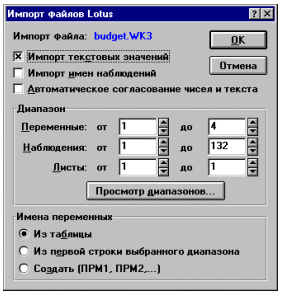
Copyright © StatSoft, 1995

переменных, состоящие из префикса *ПРМ*, к которому добавляются последовательные номера (например, *ПРМ1*, *ПРМ2* и т. д.).

Заметим, что в системе *STATISTICA* можно получить доступ к данным таблицы Excel не только при применении операции импорта файлов, но также при использовании динамических связей *DDE* (см. раздел *Установить связь* главы 1) или буфера обмена. Система *STATISTICA* распознает формат буфера обмена Excel и правильным образом интерпретирует “форматированные” значения, такие как *$100* или *2,000*.

Импорт файлов Lotus

В систему *STATISTICA* можно импортировать рабочие листы из Lotus 1-2-3 версии 1A (*Lotus/1*), версии 2.01/2.2 (*Lotus/2*), версии 3 (*Lotus/3*) и Lotus для Windows (*123w*), а также из Symphony (*Symph*). Поскольку файлы из 1-2-3/3 и Lotus для Windows имеют трехмерную структуру, при импорте в *STATISTICA* их последовательным страницам соответствуют последовательные группы переменных в файле *STATISTICA*. Импортируемые из Lotus значения-даты переводятся в юлианский формат, т.е. в числа, соответствующие количеству дней, прошедших с 1 января 1900 года. Например, дата 1/21/1968 будет преобразована в число 24858. Для того чтобы даты отображались на экране и в таблицах результатов не в юлианском, а в общепринятом формате, измените формат отображения переменных в диалоговом окне *текущих спецификаций таблицы исходных данных* с значения *Число* на значение *Дата* (см. главу 3). Если в диалоговом окне *Выбор импортируемого файла* указать файл Lotus, то откроется диалоговое окно *Импорт данных из* (по умолчанию будет выбран файл Lotus). Для открытия диалогового окна *Импорт файлов Lotus* нажмите кнопку *Параметры*. Это диалоговое окно используется для всех версий Lotus.



Импорт текстовых значений

При выборе этого варианта непарные столбцы текстовых значений, меткам которых не поставлены в соответствие числовые значения (см. ниже), будут импортированы в систему *STATISTICA* в “двойные переменные”, при этом текстовым значениям будут поставлены в соответствие числовые значения, представляющие собой последовательные целые числа. Если указанный режим не выбран, система *STATISTICA* не будет импортировать непарные метки.

Импорт имен наблюдений

Имена наблюдений могут быть импортированы из первого столбца заданного пользователем диапазона данных в файле Lotus. Если эта переменная содержит числовые, а не текстовые значения, в качестве имен наблюдений будут выступать созданные “текстовые отображения” этих значений.

Автоматическое согласование чисел и текста

Этот режим позволяет автоматически объединять метки Lotus 1-2-3 для образования “двойных переменных” системы *STATISTICA*,

Copyright © StatSoft, 1995

StatSoft®

имеющих числовые и текстовые значения и, возможно, метки значений. При выборе этого режима система *STATISTICA* комбинирует столбцы Lotus для получения "двойных переменных", сопоставляя столбцу с числовыми значениями один или два текстовых столбца с соответствующими именами по следующим правилам (заметим, что этот способ именования столбцов используется при экспорте файлов данных из системы *STATISTICA* в Lotus по умолчанию):

- Числовому столбцу могут быть поставлены в соответствие один или два следующих непосредственно за ним текстовых столбца, которые преобразуются в текстовые значения и метки значений переменных системы *STATISTICA*.
- Первый текстовый столбец, который может иметь длину до 8 символов, преобразуется в текстовые значения переменной системы *STATISTICA*. Имя этого столбца должно совпадать с именем числового столбца, причем, если имя этого столбца короче 8 символов, к концу имени должен быть добавлен символ *$*, а если длина имени равна 8 символам, второй символ этого имени должен быть заменен на символ *$*. Заметим, что если ширина столбца превышает 8 символов, этот столбец игнорируется, и соответствующие текстовые значения не импортируются.
- Второй текстовый столбец преобразуется в метки значений. Имя этого столбца должно совпадать с именем числового столбца, причем если имя этого столбца короче 8 символов, к концу имени должен быть добавлен символ @, а если длина имени равна 8 символам, то второй символ этого имени должен быть заменен на символ @.

Между числовыми и текстовыми значениями в таких столбцах должно существовать взаимно однозначное соответствие: каждому числовому значению может соответствовать только одна текстовая запись в текстовом столбце (столбцах). В процессе создания файла данных системы *STATISTICA* программа проверяет наличие такого соответствия. В случае если одному числовому значению соответствует более одного текстового значения, на экран выводится предупреждение.

Диапазон импортируемых значений

По умолчанию импортируется вся активная область рабочего листа. Пользователь может выбрать для импорта часть рабочего листа, задав диапазон импортируемых данных (см. ниже).

Переменные. Задайте диапазон импортируемых переменных (*от* переменной и *до* переменной включительно).

Наблюдения. Задайте диапазон импортируемых наблюдений (*от* наблюдения и *до* наблюдения включительно).

Листы. Этот режим можно выбрать только в том случае, если вы импортируете файл из Lotus 1-2-3 Версия 3 или Lotus для Windows. В любом другом случае он недоступен. По умолчанию импортируются все листы, однако, используя данный параметр, пользователь может выбрать нужные листы (*от* листа и *до* листа включительно).

Просмотр диапазонов. Нажав кнопку *Просмотр диапазонов*, откройте окно *Просмотр диапазонов*, в котором можно выбрать именованный диапазон из списка всех именованных диапазонов выбранного файла (см. ниже).

StatSoft®

Copyright © StatSoft, 1995

Имена переменных

С помощью этого параметра можно выбрать, каким образом будут сформированы имена переменных в файле *STATISTICA*.

Из таблицы. При выборе этого варианта в качестве имен переменных в *STATISTICA* будут использоваться указатели столбцов Lotus 1-2-3 (A, B, C и т. д.). Lotus 1-2-3 версии 3 использует имена *A:A*, *A:B*, *B:A*, *B:B* и т.д., которые преобразуются в *STATISTICA* в *A_A*, *A_B*, *B_A*, *B_B*. Этот режим особенно полезен при работе с файлами версии 3, которые впоследствии предстоит экспортировать (использование указателей столбцов Lotus позволяет сохранить информацию о происхождении переменных, принадлежащих различным листам).

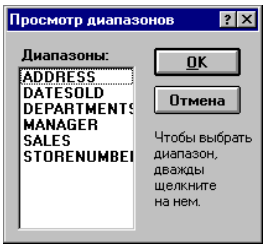
Из первой строки выбранного диапазона. При выборе этого варианта в качестве имен переменных используются метки из первой строки выбранного диапазона. Метки длиной более 8 символов обрезаются; для создания имен переменных используются только 8 первых символов.

Создать (*ПРМ1*, *ПРМ2*, . . .). При использовании этого варианта в импортированном файле создаются имена переменных, состоящие из префикса *ПРМ*, к которому добавляются последовательные номера (например, *ПРМ1*, *ПРМ2* и т. д.).

Заметим, что данные таблиц Lotus доступны в системе *STATISTICA* не только при применении операции импорта файлов, но также при использовании динамических связей *DDE* (см. раздел *Установить связь* главы 1), или буфера обмена. Система *STATISTICA* распознает формат буфера обмена Lotus и правильным образом интерпретирует “форматированные” значения, такие как *$100* или *2,000*.

Импорт данных Lotus – Просмотр диапазонов

В базе данных Lotus можно задавать “именованные диапазоны” данных и определять тем самым требуемые части рабочего листа. При нажатии кнопки *Просмотр диапазонов* в диалоговом окне *Импорт из Lotus - STATISTICA* появляется окно со списком диапазонов данных рабочего листа Lotus, выбранного пользователем для импорта.

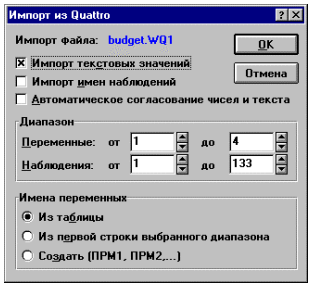


Выбрав в этом окне именованный диапазон данных рабочего листа, пользователь может импортировать не весь рабочий лист, который, возможно, содержит заголовки столбцов и другую текстовую информацию, а только подмножество содержащихся в этом рабочем листе данных. Для этого нужно выделить имя диапазона и нажать кнопку *OK*.

Импорт файлов Quattro

В систему *STATISTICA* можно импортировать рабочие листы из Quattro Pro версий для DOS и Windows (см. замечания об импорте данных из Quattro для Windows в конце данного раздела). Нажав в диалоговом окне *Импорт данных из* кнопку *Параметры*, откройте диалоговое окно *Импорт из Quattro*.

Copyright © StatSoft, 1995



Импорт текстовых значений

При выборе этого варианта непарные столбцы текстовых значений, меткам которых не поставлены в соответствие числовые значения (см. ниже), будут импортированы в систему *STATISTICA* в “двойные переменные”, при этом текстовым значениям будут поставлены в соответствие числовые значения, представляющие собой последовательные целые числа. Если указанный режим не выбран, система *STATISTICA* не будет импортировать непарные метки.

Импорт имен наблюдений

Имена наблюдений могут быть импортированы из первого столбца заданного пользователем диапазона данных в файле Quattro. Если эта переменная содержит числовые, а не текстовые значения, в качестве имен наблюдений будут выступать созданные “текстовые отображения” этих значений.

Автоматическое согласование чисел и текста

Этот режим позволяет автоматически объединять метки Quattro для образования “двойных переменных” *STATISTICA*, имеющих числовые и текстовые значения и, возможно, метки значений. При его выборе система *STATISTICA* комбинирует столбцы Quattro для получения “двойных переменных”, сопоставляя столбцу с числовыми значениями один или два текстовых столбца с соответствующими именами по следующим правилам (заметим, что этот способ именования столбцов используется по умолчанию при экспорте файлов данных из модуля *Управление данными* в Quattro):

- Числовому столбцу могут быть поставлены в соответствие один или два следующих непосредственно за ним текстовых столбца, которые преобразуются в текстовые значения и метки значений переменных системы *STATISTICA*.
- Первый текстовый столбец, который может иметь длину до 8 символов, преобразуется в текстовые значения переменной системы *STATISTICA*. Имя этого столбца должно совпадать с именем числового столбца, причем, если имя этого столбца короче 8 символов, к концу имени должен быть добавлен символ *$*, а если длина имени равна 8 символам, второй символ этого имени должен быть заменен на символ *$*. Заметим, что если ширина столбца превышает 8 символов, этот столбец игнорируется, и соответствующие текстовые значения не импортируются.
- Второй текстовый столбец преобразуется в метки значений. Имя этого столбца должно совпадать с именем числового столбца, причем если имя этого столбца короче 8 символов, к концу имени должен быть добавлен символ @, а если длина имени равна 8 символам, то второй символ этого имени должен быть заменен на символ @.

Между числовыми и текстовыми значениями в таких столбцах должно существовать взаимно однозначное соответствие: каждому числовому значению может соответствовать только одна текстовая запись в текстовом столбце (столбцах). В процессе создания файла данных система *STATISTICA* программа проверяет наличие

StatSoft®
Copyright © StatSoft, 1995

такого соответствия. В случае если одному числовому значению соответствует более одного текстового значения, на экран выводится предупреждение.

Диапазон импортируемых значений

По умолчанию импортируется вся активная область рабочего листа. Пользователь может выбрать для импорта часть рабочего листа, задав диапазон импортируемых данных (см. ниже).

Переменные. Задайте диапазон импортируемых переменных (*от* переменной и *до* переменной включительно).

Наблюдения. Задайте диапазон импортируемых наблюдений (*от* наблюдения и *до* наблюдения включительно).

Имена переменных

С помощью этого параметра можно выбрать, каким образом будут сформированы имена переменных в файле *STATISTICA*.

Из таблицы. При выборе этого варианта в качестве имен переменных в *STATISTICA* будут использоваться указатели столбцов Quattro (A, B, C и т. д.).

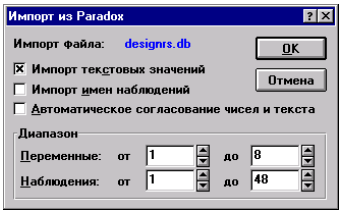
Из первой строки выбранного диапазона. При выборе этого варианта в качестве имен переменных используются метки из первой строки выбранного диапазона. Метки длиной более 8 символов обрезаются; для создания имен переменных используются только 8 первых символов.

Создать (*ПРМ1*, *ПРМ2*, . . .). При использовании этого варианта в импортированном файле создаются имена переменных, состоящие из префикса *ПРМ*, к которому добавляются последовательные номера (например, *ПРМ1*, *ПРМ2* и т.д.).

Заметим, что данные таблиц Quattro для Windows доступны в системе *STATISTICA* не только при применении операции импорта файлов, но также при использовании динамических связей *DDE* (см. раздел *Установить связь* главы 1) или буфера обмена. *STATISTICA* распознает формат буфера обмена Quattro для Windows и правильным образом интерпретирует “форматированные” значения, такие как *$100* или *2,000*.

Импорт файлов Paradox

В систему *STATISTICA* можно импортировать файлы данных из базы Paradox. Нажав в диалоговом окне *Импорт данных из* кнопку *Параметры*, откройте диалоговое окно *Импорт из Paradox*.



Импорт текстовых значений

При выборе этого варианта непарные текстовые значения, меткам которых не поставлены в соответствие числовые значения (см. ниже), будут импортированы в “двойные переменные”, при этом текстовым значениям будут поставлены в соответствие числовые значения, представляющие собой последовательные целые числа. Если указанный режим не выбран, система *STATISTICA* не будет импортировать непарные метки.

Copyright © StatSoft, 1995

Импорт имен наблюдений

Имена наблюдений могут быть импортированы из первого столбца заданного пользователем диапазона данных в файле Paradox. Если эта переменная содержит числовые, а не текстовые значения, в качестве имен наблюдений будут выступать созданные "текстовые отображения" этих значений.

Автоматическое согласование чисел и текста

Этот режим позволяет автоматически объединять метки Paradox для образования "двойных переменных" *STATISTICA*, имеющих числовые и текстовые значения и, возможно, метки значений. При его выборе *STATISTICA* комбинирует столбцы Paradox для получения "двойных переменных", сопоставляя столбцу с числовыми значениями один или два текстовых столбца с соответствующими именами по следующим правилам (заметим, что этот способ именования столбцов используется при экспорте файлов данных из модуля *Управление данными* в Paradox по умолчанию):

- Числовому столбцу могут быть поставлены в соответствие один или два следующих непосредственно за ним текстовых столбца, которые преобразуются в текстовые значения и метки значений переменных системы *STATISTICA*.
- Первый текстовый столбец, который может иметь длину до 8 символов, преобразуется в текстовые значения переменной системы *STATISTICA*. Имя этого столбца должно совпадать с именем числового столбца, причем, если имя этого столбца короче 8 символов, к концу имени должен быть добавлен символ *$*, а если длина имени равна 8 символам, второй символ этого имени должен быть заменен на символ *$*. Заметим, что если ширина столбца превышает 8 символов, этот столбец игнорируется, и соответствующие текстовые значения не импортируются.
- Второй текстовый столбец преобразуется в метки значений. Имя этого столбца должно совпадать с именем числового столбца, причем если имя этого столбца короче 8 символов, к концу имени должен быть добавлен символ @, а если длина имени равна 8 символам, то второй символ этого имени должен быть заменен на символ @.

Между числовыми и текстовыми значениями в таких столбцах должно существовать взаимно однозначное соответствие: каждому числовому значению может соответствовать только одна текстовая запись в текстовом столбце (столбцах). В процессе создания файла данных система *STATISTICA* проверяет наличие такого соответствия. В случае если одному числовому значению соответствует более одного текстового значения, на экран выводится предупреждение.

Диапазон импортируемых значений

По умолчанию импортируется вся активная область рабочего листа. Пользователь может выбрать для импорта часть рабочего листа, задав диапазон импортируемых данных (см. ниже).

Переменные. Задайте диапазон импортируемых переменных (*от* переменной и *до* переменной включительно).

Наблюдения. Задайте диапазон импортируемых наблюдений (*от* наблюдения и *до* наблюдения включительно).

Импорт файлов dBASE

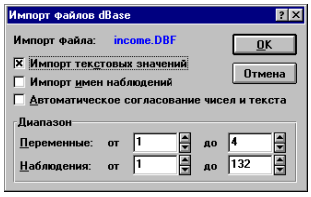
В системе *STATISTICA* можно импортировать файлы данных из dBASE. В каждой записи

StatSoft®

Copyright © StatSoft, 1995

файла dBASE содержится флаг, указывающий, нужно ли считывать запись; система *STATISTICA* игнорирует все записи, помеченные как удаленные.

Если в диалоговом окне *Выбор импортируемого файла* указать файл dBASE, откроется диалоговое окно *Импорт данных из* (по умолчанию будет выбран файл dBASE). Чтобы открыть диалоговое окно *Импорт из dBASE*, нажмите кнопку *Параметры*. Это диалоговое окно используется для всех версий dBASE.



Импорт текстовых значений

При выборе этого варианта непарные текстовые поля, т.е. те, которым не поставлены в соответствие числовые поля, будут импортированы в систему *STATISTICA* в "двойные переменные", при этом текстовым значениям будут поставлены в соответствие числовые значения, представляющие собой последовательные целые числа. Если указанный режим не выбран, система *STATISTICA* не будет импортировать непарные текстовые поля.

Импорт имен наблюдений

При выборе этого варианта имена наблюдений могут быть импортированы из первого поля файла dBASE. Если это поле содержит числовые, а не текстовые значения, в качестве имен наблюдений будут выступать созданные "текстовые отображения" этих значений.

Автоматическое согласование чисел и текста

Этот режим позволяет автоматически объединять метки dBASE для образования "двойных переменных" системы *STATISTICA*, имеющих числовые и текстовые значения и, возможно, метки значений. При его выборе система *STATISTICA* комбинирует поля файла dBASE для получения "двойных переменных", сопоставляя поле с числовыми значениями одному или двум текстовым полям с соответствующими именами по следующим правилам (заметим, что этот способ именования столбцов используется по умолчанию при экспорте файлов данных из модуля *Управление данными* в dBASE):

- Числовому полю могут быть поставлены в соответствие одно или два следующих непосредственно за ним текстовых поля, которые преобразуются в текстовые значения и метки значений переменных системы *STATISTICA*.
- Первое текстовое поле, которое может иметь длину до 8 символов, преобразуется в текстовые значения переменной системы *STATISTICA*. Имя этого поля должно совпадать с именем числового поля, причем, если имя этого поля короче 8 символов, к концу имени должен быть добавлен символ *$*, а если длина имени равна 8 символам, второй символ этого имени должен быть заменен на символ *$*. Заметим, что если ширина поля превышает 8 символов, это поле игнорируется, и соответствующие текстовые значения не импортируются.
- Второе текстовое поле преобразуется в метки значений. Имя этого поля должно совпадать с именем числового поля, причем если имя этого поля короче 8 символов, к концу имени должен быть добавлен символ @, а если длина имени равна 8 символам, то

Copyright © StatSoft, 1995

второй символ этого имени должен быть заменен на символ @.

Между числовыми и текстовыми значениями в таких полях должно существовать взаимно однозначное соответствие: каждому числовому значению может соответствовать только одна текстовая запись в текстовом поле (полях). В процессе создания файла данных системы *STATISTICA* программа проверяет наличие такого соответствия. В случае если одному числовому значению соответствует более одного текстового значения, на экран выводится предупреждение.

Диапазон импортируемых значений

По умолчанию импортируется вся активная область файла данных. Пользователь может выбрать для импорта часть файла данных, задав диапазон импортируемых данных (см. ниже).

Переменные. Задайте диапазон импортируемых переменных (*от* переменной и *до* переменной включительно).

Наблюдения. Задайте диапазон импортируемых наблюдений (*от* наблюдения и *до* наблюдения включительно).

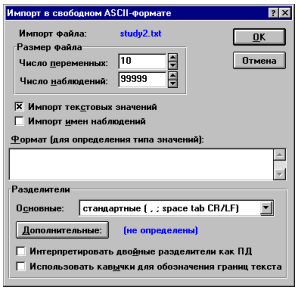
Импорт ASCII-файлов

В системе *STATISTICA* можно импортировать файлы как в свободном (с разделителями), так и в фиксированном (без разделителей) ASCII-формате. Для каждого типа ASCII-файлов предусмотрен свой набор параметров. Подробная информация о них приведена ниже.

Импорт файлов в свободном ASCII-формате

В отличие от импорта файлов в фиксированном (без разделителей) ASCII-формате (см. стр. 1406), импорт файлов в свободном (с разделителями) формате не требует, чтобы данные во всех переменных ASCII-файла находились в заданных (фиксированных) столбцах. В ASCII-файлах, записанных в свободном формате, для разделения значений (переменных) в наблюдениях используются разделители или ограничители, например, запятые или пробелы. Тип разделителя, используемого в конкретных файлах, задается пользователем в диалоговом окне *Разделители для свободного ASCII-формата* (см. стр. 1405).

При нажатии кнопки *Параметры* в диалоговом окне *Импорт данных из – Свободный ASCII-формат* открывается диалоговое окно *Импорт в свободном ASCII-формате*.



Размер файла

Как и в случае импорта файлов в фиксированном ASCII-формате, при импорте файлов в свободном ASCII-формате необходимо точно задать количество переменных.

Число переменных. Задайте число переменных в ASCII-файле.

Число наблюдений. Задайте число наблюдений в ASCII-файле. Если вы не уверены, что знаете абсолютно точно, сколько

StatSoft®
Copyright © StatSoft, 1995

наблюдений содержится в импортируемом файле, задайте количество наблюдений с заведомым избытком. В процессе импорта файла *STATISTICA* определит реальную длину файла. Заметим, что длина строки в импортируемом ASCII-файле может быть не более 4000 символов. Это ограничение действительно только для длин строк и не распространяется на общую длину "наблюдения", поскольку каждое импортируемое наблюдение в файле-источнике может состоять из нескольких строк.

Импорт текстовых значений

При выборе этого варианта текстовые поля преобразуются в "двойные переменные" системы *STATISTICA*, при этом каждому текстовому значению ставится в соответствие целое число из ряда последовательных целых чисел. Текстовые значения, длина которых превышает 8 символов, обрезаются. При этом в качестве текстовых значений будут использоваться только первые восемь символов. Если указанный режим не выбран пользователем, система *STATISTICA* не будет импортировать текстовые переменные.

Импорт имен наблюдений

При выборе этого варианта имена наблюдений могут быть импортированы из первого поля каждой записи ASCII-файла. Если эта переменная содержит более 20 символов, для создания имен наблюдений будут использоваться только первые 20 символов. Если это поле содержит числовые, а не текстовые значения, в качестве имен наблюдений будут выступать созданные "текстовые отображения" этих значений.

Формат

Этот параметр в явном виде определяет для системы *STATISTICA* содержание ASCII-файла - источника. Необходимо заметить, что формат – это набор инструкций для интерпретации структуры файла-источника, а не определение создаваемых переменных *STATISTICA*. Спецификации формата вводятся в виде *nX*, где *n* – целое число, указывающее, сколько раз формат необходимо повторить (по умолчанию полагается равным 1); *X* – тип столбца. Например, формат *40F* означает 40 полей, содержащих числовые значения (здесь - с плавающей точкой).

Спецификации типов столбцов

В ASCII-файлах используются следующие спецификации типов столбцов:

A - Текстовый (алфавитно-цифровой)

F - С плавающей точкой (*R* - Действительный)

D - С двойной точностью (*DR* – Двойной действительный)

I - Целочисленный (*NI* – нормальный целочисленный; диапазон от -32768 до 32767)

S - Короткий целочисленный (*SI*; диапазон от -128 до 127)

LI - Длинный целочисленный (*J*; диапазон от – 2147483648 до 2147483647)

L - Логический

В *логическом* формате должны содержаться текстовые значения *истина* и *ложь*. При импорте данных *STATISTICA* распознает следующие три варианта обозначений:

TRUE или *FALSE*, если длина поля составляет не менее 5 символов.

Copyright © StatSoft, 1995

YES или *NO*, если длина поля равна 3 или 4 символам.

Y, *N* или *T*, *F* соответственно, если поле имеет длину 1 символ.

Логические значения *TRUE* или *YES* импортируются как *1*, а *FALSE* или *NO* импортируются как *0*.

В *текстовых* полях (тип *A*) непосредственно за буквой *A* должно следовать число от 1 до 255, определяющее максимально возможную длину текстового значения в этом поле.

В спецификациях формата может использоваться символ "/" (косая черта), указывающий, что оставшаяся часть текущей строки файла-источника при импорте данных должна быть пропущена, т.е. далее данные будут импортированы со следующей строки файла-источника. Если перед списком форматов, заключенным в скобки, стоит множитель, этот список должен быть повторен количество раз, равное этому множителю.

Например, формат *2(2L a5)* эквивалентен формату *L L A5 L L A5* и определяет две *логические* переменные, за которыми следует текстовая переменная длиной 5 символов, затем снова две логические переменные и одна текстовая переменная длиной 5 символов.

Разделители

С помощью этого параметра можно задать символы, используемые в файле-источнике в качестве разделителей. Окончательный список используемых разделителей будет представлять собой комбинацию набора выбранных *основных* и *дополнительных* разделителей.

Основные. Из четырех заданных наборов разделителей (*CR* – возврат каретки, *LF* – перевод строки и *FF* – перевод страницы) выберите тип разделителя, используемого в файле-источнике.

Стандартные. В этот набор входят: запятая (*,*), точка с запятой (*;*),<*space*> (пробел), <*tab*> (табулятор), <*CR\LF*> (возврат каретки\перевод строки).

Не задано. Эта установка означает, что не будет использоваться основной набор разделителей (см. ниже описание кнопки *Дополнительные*).

Непечат. символы. Набор непечатаемых символов содержит: <*space*> (пробел), <*tab*> (табуляцию), <*FF*> (перевод страницы), <*LF*> (перевод строки) и <*CR\LF*>(возврат каретки\перевод строки).

Не числа. В этот набор входят все символы, кроме *0-9*, десятичной точки (.), минуса (-) и плюса (+).

Дополнительные. Нажав кнопку *Дополнительные*, выберите разделитель или разделители в открывшемся диалоговом окне *Свободный ASCII-формат* (см. стр. 1405).

Интерпретировать двойные разделители как ПД. При выборе этого режима система *STATISTICA* интерпретирует каждую пару расположенных рядом разделителей как пропущенные данные (отсутствующее значение) и помещает в позицию между двумя разделителями код пропущенных данных (-9999). Если указанный режим не выбран, *STATISTICA* будет обрабатывать несколько расположенных рядом разделителей как один разделитель. В этом случае коды пропущенных данных должны быть введены в ASCII-файл в явном виде как заданное значение, например, *-9999*. Этот режим полезен, например, когда значения в файле данных разделены пробелами, причем между значениями может быть различное количество пробелов. Если в качестве разделителей используются *пробелы*, каждая пара пробелов интерпретировалась бы как пропущенное значение, и в результате

StatSoft®

Copyright © StatSoft, 1995

импортированный файл содержал бы большое количество пропущенных значений.

Использовать кавычки для обозначения границ текста. Этот режим следует выбрать, если в качестве маркеров границ текстовых значений используются двойные (") или одинарные (') кавычки и в состав текстовых значений входят символы, используемые в качестве разделителей. Например, в записи *"John Jones*, *Ph.D."*, запятая используется и как часть текстового значения, и как разделитель. В этом случае при импорте данных *STATISTICA* рассматривает кавычки только как маркеры границ текстовых значений и импортирует заключенный между ними текст, включая разделители, как единое текстовое значение. Кавычки при выборе этого режима не входят в состав текстового значения и не импортируются.

Если кавычки входят в состав текстового значения, например, в название книги (*"Моби Дик"*), для их импорта можно воспользоваться одним из двух приведенных ниже способов.

- Выберите этот режим и заключите текстовое значение в кавычки, не совпадающие с используемыми в текстовом значении. Например, значение '*Уильям Шекспир,* "*Король Лир*"*, Акт 2, Сцена 1*' будет импортировано в виде *Уильям Шекспир,* "*Король Лир*"*, Акт 2, Сцена 1.*
- Выберите этот режим и заключите текстовое значение в дополнительные кавычки, например, значение "*Уильям Шекспир* ""*Король Лир*"", ""*Макбет*"" *и* ""*Гамлет*""" будет импортировано как *Уильям Шекспир* "*Король Лир*", "*Макбет*" *и* "*Гамлет*".

Если указанный режим не выбран, соответствующие символы будут интерпретироваться как разделители, а не как часть текстового значения, а кавычки будут импортироваться как часть текста. Например, '*Уильям Шекспир,* "*Король Лир*"*, Акт 2, Сцена 1*' будут импортированы в виде '*Уильям Шекспир,* в одном столбце, "*Король Лир*" – в следующем столбце*, Акт 2* – в третьем столбце и *Сцена 1*' – в четвертом.

Разделители в свободном ASCII-формате

Чтобы при импорте файла в свободном ASCII-формате выбрать в качестве разделителя не заданный ранее символ, откройте диалоговое окно *Свободный ASCII-формат*. Для этого нажмите кнопку *Дополнительные* в диалоговом окне *Импорт в свободном ASCII-формате*.

В этом диалоговом окне представлен список символов, из которых можно выбрать следующие символы – разделители:

<space> <tab> <CR/LF>

*! " # $ % & ') * (, [\ + - .*

0 1 2 3 4 5 6 7 8 9

] ^ _ / : ; < = > ? @ ` { | } ~

Прописные латинские буквы (т.е. *A B C . . . Z*)

Строчные латинские буквы (т.е. *a b c . . . z*)

Заметьте, что если вы хотите импортировать числовые значения, нельзя выбирать в качестве разделителей цифры (от *0* до *9*) или знаки "+", "-", ".".

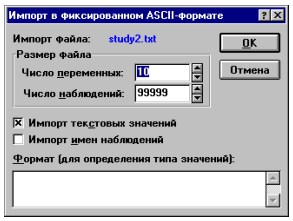
Copyright © StatSoft, 1995

StatSoft®

Импорт файлов в фиксированном ASCII-формате

В отличие от импорта файлов в свободном (с разделителями) ASCII-формате (см. стр. 1402), при импорте файлов в фиксированном (без разделителей) формате предполагается, что данные во всех переменных ASCII-файла находятся в заданных (фиксированных) столбцах, а положение каждой переменной во всех наблюдениях одинаково. Например, значения переменной *Gender (Пол)* могут находиться в столбцах 3 и 4 для всех наблюдений в файле данных.

Если в диалоговом окне *Выбор импортируемого файла* указан файл в фиксированном ASCII-формате, то в диалоговом окне *Импорт данных из* этот формат будет выбран по умолчанию. При расширении **.fix* автоматически будет принято, что файл имеет фиксированный ASCII-формат. Если программа не может распознать по расширению, записан ли файл в фиксированном ASCII-формате, будет выбрана версия *Свобод.* ASCII-формат, которую нужно будет изменить на *Фиксир.*. При нажатии в диалоговом окне *Выбор импортируемого файла* кнопки *Параметры* откроется диалоговое окно *Импорт в фиксированном ASCII-формате*.



Размер файла

Как и в случае импорта файлов в свободном ASCII-формате (см. стр. 1402), для правильной интерпретации системой *STATISTICA* данных, хранящихся в файле, необходимо точно задать количество переменных и наблюдений в этом файле.

Число переменных. Задайте число переменных в ASCII-файле.

Число наблюдений. Задайте число наблюдений в ASCII-файле. Если вы не уверены, что знаете абсолютно точно, сколько наблюдений содержится в импортируемом файле, задайте количество наблюдений с заведомым избытком. В процессе импорта файла система *STATISTICA* определит реальную длину файла. Заметим, что длина строки в импортируемом ASCII-файле не должна превышать 4000 символов. Это ограничение действительно только для длин строк и не распространяется на общую длину "наблюдения", поскольку каждое импортируемое наблюдение в файле-источнике может состоять из нескольких строк.

Импорт текстовых значений

При выборе этого варианта текстовые поля конвертируются в "двойные переменные" системы *STATISTICA*, при этом каждому текстовому значению ставится в соответствие целое число из ряда последовательных целых чисел. Текстовые значения, длина которых превышает 8 символов, обрезаются. При этом в качестве текстовых значений будут использоваться только первые восемь символов. Если указанный режим не выбран пользователем, система *STATISTICA* не будет импортировать текстовые переменные.

Импорт имен наблюдений

При выборе этого варианта имена наблюдений могут быть импортированы из первого поля каждой записи ASCII-файла. Если это поле содержит более 20 символов, для создания имен наблюдений будут использоваться только первые

StatSoft®
Copyright © StatSoft, 1995

20 символов. Если это поле содержит числовые, а не текстовые значения, в качестве имен наблюдений будут выступать созданные "текстовые отображения" этих значений.

Формат

Этот параметр в явном виде определяет для системы *STATISTICA* содержание ASCII-файла - источника. Необходимо заметить, что формат – это набор инструкций для интерпретации структуры файла-источника, а не определение создаваемых переменных системы *STATISTICA*. Спецификации формата вводятся в виде *nXm*, где *n* – целое число, указывающее, сколько раз формат необходимо повторить (по умолчанию полагается равным 1); *X* – тип столбца, а *m* – ширина столбца. Например, формат *10A8* означает 10 текстовых (алфавитно-цифровых) полей, длина каждого из которых равна 8 символам.

Спецификации типов столбцов

В ASCII-файлах используются следующие спецификации типов столбцов:

A - Текстовый (алфавитно-цифровой)

F - С плавающей точкой (*R* - Действительный)

D - С двойной точностью (*DR* – Двойной действительный)

I - Целочисленный (*NI* – нормальный целочисленный; диапазон от -32768 до 32767)

S - Короткий целочисленный (*SI*; диапазон от -128 до 127)

LI - Длинный целочисленный (*J*; диапазон от – 2147483648 до 2147483647)

L - Логический

T - Спецификатор "Перейти к " (см. ниже)

В *логическом* формате должны содержаться текстовые указатели *истина* и *ложь*. При импорте данных *STATISTICA* распознает следующие три варианта обозначений:

TRUE или *FALSE*, если длина поля составляет не менее 5 символов.

YES или *NO*, если длина поля равна 3 или 4 символам.

Y, *N* или *T*, *F* соответственно, если поле имеет длину 1 символ.

Логические значения *TRUE* или *YES* импортируются как *1*, а *FALSE* или *NO* импортируются как *0*.

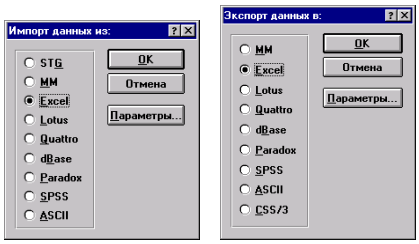
Дополнительное обозначение *Tx* используется для того, чтобы дать инструкцию системе *STATISTICA* перейти к символу с номером *x* в текущей строке исходного файла, т.е. переместиться в эту позицию. Такую инструкцию можно дать для того, чтобы "перескочить" через нежелательные символы или чтобы вернуться к одному из предшествующих символов, если *x* задает позицию, символ в которой уже был считан. В спецификациях формата может использоваться символ "/" (косая черта), указывающий, что оставшаяся часть текущей строки файла-источника при импорте данных должна быть пропущена, т.е. далее данные будут импортированы со следующей строки файла-источника. Например, формат *10I3 A8 T55 3I5 /* задает импорт 14 значений: десять трехзначных целых чисел, за которыми следует текстовое значение длиной 8 символов, затем требуется перескочить на 55-й символ в записи, импортировать три пятизначных целых числа и проигнорировать всю остальную часть записи. Если перед списком форматов, заключенным в скобки, стоит множитель, этот список должен быть повторен количество раз, равное этому множителю (например, *10(2A3,I5)*)..

Copyright © StatSoft, 1995

StatSoft®

Экспорт данных

В каждом модуле системы *STATISTICA* можно использовать средства для импорта и экспорта файлов, которые вызываются в выпадающем меню *Файл*. С помощью этого же меню в каждом модуле можно, используя интерфейс *ODBC*, осуществлять доступ к различным файлам других приложений, в том числе файлам СУБД больших компьютеров. Наиболее полная информация по этому вопросу приведена в главе 3 и в *Электронном руководстве*. В модуле *Управление данными* имеются дополнительные возможности для импорта и экспорта файлов. Для доступа к ним следует выбрать пункт *Дополнительно…* в подменю *Импорт* или *Экспорт*:



,

после чего программа предложит набор возможностей для обмена данными с конкретными приложениями. Наиболее существенной является возможность обмена данными, имеющими в системе *STATISTICA* двойное (текстовое и числовое) представление и длинные метки значений, с другими приложениями, не поддерживающими метки или двойное представление данных. В частности, предусмотрены возможности "маркировки" и разделения числовых и текстовых значений и их меток при экспорте данных, а также объединения значений и меток при импорте данных. В системе *STATISTICA* имеются также средства для доступа к файлам, записанным в сравнительно редко используемых форматах, не удовлетворяющих стандартам *ODBC*, таких как файлы в формате SPSS *PFF* (*Portable File Format*), а также к графическим файлам STG.

Пример. Для экспорта файла система *STATISTICA* выполните действия, приведенные ниже.

Шаг 1. Выберите в выпадающем меню *Файл* пункт *Экспорт данных*, после чего откроется диалоговое окно *Экспорт данных в*.

Шаг 2. Выберите один из вариантов экспорта - тип файла, в который будут экспортированы данные (например, *Lotus*). Можно также нажать кнопку *Параметры* и выбрать параметры экспорта (см. ниже).

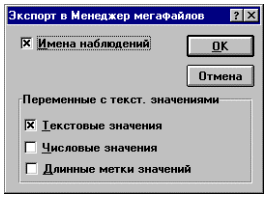
Шаг 3. Нажмите в указанном выше диалоговом окне кнопку *OK*, после чего данные будут экспортированы и сохранены в требуемом формате.

Экспорт данных в файлы *Менеджера мегафайлов*

В системе *STATISTICA* можно экспортировать наборы данных в *Менеджер мегафайлов*. При выборе этого варианта в диалоговом окне *Экспорт данных в* откроется диалоговое окно *Экспорт в Менеджер мегафайлов*.

StatSoft®

Copyright © StatSoft, 1995



При экспорте данных длинные имена переменных *STATISTICA* (если они имеются) преобразуются в длинные имена столбцов файла *Менеджера мегафайлов*, а заголовок файла *STATISTICA* – в строку заголовка файла *Менеджера мегафайлов*.

Имена наблюдений

При выборе этого варианта имена наблюдений экспортируются в первый столбец файла *Менеджера мегафайлов*.

Переменные с текстовыми значениями

В *Менеджере мегафайлов* "двойные переменные" (числовые/текстовые) могут быть по желанию пользователя расширены до трех столбцов, содержащих числовые и текстовые значения и, возможно, метки значений переменных *STATISTICA*. Имена этих столбцов будут соответствовать именам переменных в файле *STATISTICA*. Пользователь может управлять их созданием с помощью параметров, позволяющих при импорте файла в формат *STATISTICA* снова преобразовывать пары и тройки столбцов *Менеджера мегафайлов* (т.е. "сжимать") в отдельные переменные *STATISTICA*. Если не выбран ни один из этих вариантов, экспортируются только те переменные *STATISTICA*, которые не имеют текстовых значений.

Текстовые значения. При выборе этого варианта система *STATISTICA* создаст для "двойных переменных" текстовые столбцы с алфавитно-цифровыми значениями. При создании текстовых столбцов для образования их имен используются имена "двойных переменных", с добавлением к ним символа *$* по следующим правилам:

- Если имя переменной *STATISTICA* короче восьми символов, к концу имени текстового столбца *Менеджера мегафайлов* будет добавлен символ *$*;
- Если длина имени переменной *STATISTICA* составляет восемь символов, второй символ имени текстового столбца *Менеджера мегафайлов* будет заменен на *$*. Это позволяет избежать в дальнейшем конфликтов с ограничениями, налагаемыми на имена *Менеджером мегафайлов* и связанными с повторяющимися именами столбцов.

Числовые значения. При выборе этого варианта создаются столбцы числовых значений, соответствующие столбцам текстовых значений, созданным при выборе предыдущего режима. При экспорте такой переменной в *Менеджер мегафайлов* будет использовано то же имя переменной. Если указанный вариант не выбран, экспортироваться будут только числовые значения переменных, у которых не имеется текстовых эквивалентов.

Длинные метки значений. С помощью этого параметра можно управлять созданием столбцов текстовых меток значений. При выборе этого варианта система *STATISTICA* создает третий столбец, который содержит текстовые значения, представляющие собой *Длинные метки значений* переменной. При создании такого столбца для образования его имени используется имя "двойной переменной" с добавлением к нему символа @ по следующим правилам:

- Если имя переменной *STATISTICA* короче восьми символов, к концу имени текстового столбца *Менеджера мегафайлов* будет добавлен символ @;

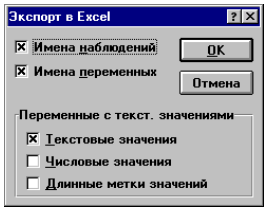
Copyright © StatSoft, 1995

- Если длина имени переменной *STATISTICA* составляет восемь символов, второй символ имени текстового столбца *Менеджера мегафайлов* будет заменен на @.

Заметим, что при экспорте только текстовых значений или меток имя соответствующего столбца в файле *Менеджера мегафайлов* будет совпадать с именем переменной *STATISTICA*, а не строиться по правилам, изложенным выше, с использованием символов *$* и @.

Экспорт файлов в Excel

Файлы Excel могут содержать до 256 столбцов и 16384 строк. Если размер файла *STATISTICA* превышает эти ограничения, пользователь может создать *подмножества* файла и экспортировать их по отдельности (см. стр. 1376). При выборе пункта *Excel* в диалоговом окне *Экспорт данных в* откроется диалоговое окно *Экспорт в Excel*.



Имена наблюдений

При выборе этого варианта имена наблюдений будут экспортироваться в первый столбец файла Excel.

Имена переменных

При выборе этого варианта система *STATISTICA* будет экспортировать имена переменных в первую строку таблицы Excel. Если указанный режим не выбран, имена переменных не будут входить в файл Excel.

Переменные с текстовыми значениями

"Двойные переменные" (числовые/текстовые) могут быть по желанию пользователя расширены до трех столбцов таблицы Excel, содержащих числовые и текстовые значения и, возможно, метки значений переменных системы *STATISTICA*. Имена этих столбцов будут соответствовать именам переменных в файле системы *STATISTICA*. Их созданием пользователь может управлять с помощью функций, позволяющих затем преобразовывать (т.е. "сжимать") пары и тройки столбцов таблицы Excel в отдельные переменные при импорте файла в систему *STATISTICA*. Если не выбран ни один из этих вариантов, экспортируются только те переменные системы *STATISTICA*, которые не имеют текстовых значений.

Текстовые значения. При выборе этого варианта система *STATISTICA* создаст для "двойных переменных" текстовые столбцы с алфавитно-цифровыми значениями. При создании текстовых столбцов система *STATISTICA* использует для образования имен этих столбцов имена "двойных переменных", добавляя к ним символ *$* по следующим правилам:

- Если имя переменной системы *STATISTICA* короче восьми символов, к концу имени текстового столбца Excel будет добавлен символ *$*;
- Если длина имени переменной системы *STATISTICA* составляет восемь символов, второй символ имени текстового столбца Excel будет заменен на *$*. Это упрощает объединение числовых и текстовых данных при импорте файла обратно в систему *STATISTICA*.

Числовые значения. При выборе этого варианта создаются столбцы числовых значений, соответствующие столбцам текстовых значений,

StatSoft®
Copyright © StatSoft, 1995

созданным при выборе предыдущего варианта. Если указанный режим не выбран, экспортироваться будут только числовые значения переменных, у которых не имеется текстовых эквивалентов.

Длинные метки значений. С помощью этого параметра можно управлять созданием столбцов текстовых меток значений. При этом система *STATISTICA* создает третий столбец, который содержит текстовые значения, представляющие собой *длинные метки значений* переменной. Для образования его имени используется имя "двойной переменной" с добавлением к нему символа @ по следующим правилам:

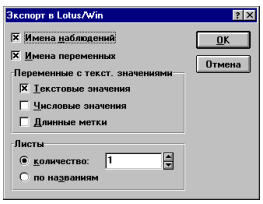
- Если имя переменной системы *STATISTICA* короче восьми символов, к концу имени текстового столбца Excel будет добавлен символ @;
- Если длина имени переменной системы *STATISTICA* составляет восемь символов, второй символ имени текстового столбца Excel будет заменен на @.

Заметим, что при экспорте только текстовых значений или меток имя соответствующего столбца в файле Excel будет совпадать с именем переменной системы *STATISTICA*, а не строиться по правилам, изложенным выше, с использованием символов *$* и @.

Экспорт файлов в Lotus

Используя диалоговое окно *Экспорт в Lotus*, в *STATISTICA* можно экспортировать файлы в Lotus 1-2-3 Версии 1A, 2.01/2.2, Версии3, Lotus для Windows и Symphony. Для открытия этого окна нажмите кнопку *Параметры* в диалоговом окне *Экспорт в Lotus/Win*.

Файлы Lotus 1-2-3 Версии 2.01/2.2 и Версии 3 могут содержать до 256 столбцов и 8192 строк. Если размер файла *STATISTICA* превышает эти ограничения, пользователь может создать *подмножества* файла и экспортировать их по отдельности (см. стр. 1376).



В диалоговом окне *Экспорт в Lotus* предлагаются следующие возможности.

Листы (только для Lotus 1-2-3 Версия 3 и Lotus 1-2-3/w)

Ниже приведены параметры экспорта для Lotus 1-2-3 Версии 1A и 2/2.2, Версии 3, и Lotus для Windows. Эти варианты одинаковы для всех указанных версий, за исключением Lotus 1-2-3 Версии 3 и Lotus для Windows, для которых предусмотрены дополнительные возможности:

Количество. По умолчанию все переменные в файле системы *STATISTICA* экспортируются в один рабочий лист. Пользователь может по желанию разделить переменные системы *STATISTICA* между заданным количеством рабочих листов. С помощью этого параметра можно задать количество рабочих листов.

По названиям. Переменные системы *STATISTICA* можно также разделить по рабочим листам в соответствии с их именами (*по именам)*: переменные с именами от *A_A* до *A_IV* поместить в рабочий лист *A*, переменные с именами от *B_A* до *B_IV* – в рабочий лист *B* и т.д.. Такую сортировку можно организовать, выбрав вариант *Из таблицы* при импорте имен переменных из рабочего листа версии 3 или Lotus для Windows (см. стр. 1395.)

Copyright © StatSoft, 1995

Приведенные ниже параметры являются общими для всех версий Lotus.

Имена наблюдений

При выборе этого варианта имена наблюдений экспортируются в первый столбец файла Lotus.

Имена переменных

При выборе этого варианта система *STATISTICA* будет экспортировать имена переменных в первую строку таблицы исходных данных Lotus. Если указанный режим не выбран, имена переменных не будут входить в файл Lotus.

Переменные с текстовыми значениями

"Двойные переменные" системы *STATISTICA* (числовые/текстовые) могут быть по желанию пользователя расширены в таблице Lotus до трех столбцов, содержащих числовые и текстовые значения и, возможно, метки значений переменных системы *STATISTICA*. Имена этих столбцов будут соответствовать именам переменных в файле *STATISTICA*. Созданием этих столбцов пользователь может управлять с помощью параметров, позволяющих затем при импорте файла обратно в систему *STATISTICA* конвертировать пары и тройки столбцов таблицы Lotus (т.е. "сжимать") в отдельные переменные. Если не выбран ни один из этих вариантов, экспортируются только те переменные, которые не имеют текстовых значений.

Текстовые значения. При выборе этого варианта система *STATISTICA* создаст для "двойных переменных" текстовые столбцы с алфавитно-цифровыми значениями. При создании текстовых столбцов система *STATISTICA* использует для образования имен этих столбцов имена "двойных переменных", добавляя к ним символ *$* по следующим правилам:

- Если имя переменной системы *STATISTICA* короче восьми символов, к концу имени текстового столбца Lotus будет добавлен символ *$*;
- Если длина имени переменной системы *STATISTICA* составляет восемь символов, второй символ имени текстового столбца Lotus будет заменен на *$*. Это упрощает объединение числовых и текстовых данных при импорте файла обратно в систему *STATISTICA*.

Числовые значения. При выборе этого варианта создаются столбцы числовых значений, соответствующие столбцам текстовых значений, созданным при выборе предыдущего режима. Если указанный вариант не выбран, экспортироваться будут только числовые значения переменных, которые не имеют текстовых эквивалентов.

Длинные метки. С помощью этого параметра можно управлять созданием столбцов текстовых меток значений. В этом случае система *STATISTICA* создает третий столбец, который содержит текстовые значения, представляющие собой *длинные метки значений* переменной. При создании такого столбца для образования его имени используется имя "двойной переменной" с добавлением к нему символа @ по следующим правилам:

- Если имя переменной системы *STATISTICA* короче восьми символов, к концу имени текстового столбца Lotus будет добавлен символ @;
- Если длина имени переменной системы *STATISTICA* составляет восемь символов, второй символ имени текстового столбца Lotus будет заменен на @.

Как и при экспорте в *Менеджер мегафайлов* (см. стр. 1408), в случае экспорта только текстовых значений или меток имя соответствующего

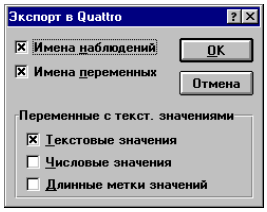
StatSoft®

Copyright © StatSoft, 1995

столбца в файле Lotus будет совпадать с именем переменной системы *STATISTICA*, а не строиться по правилам, изложенным выше, с использованием символов *$* и *@*.

Экспорт файлов в Quattro

Система *STATISTICA* позволяет экспортировать файлы в таблицы Quattro. Для этого в диалоговом окне *Экспорт данных в* нужно выбрать тип *Quattro*, после чего откроется диалоговое окно *Экспорт в Quattro*.



Имена наблюдений

При выборе этого варианта имена наблюдений экспортируются в первый столбец файла Quattro.

Имена переменных

При выборе этого варианта система *STATISTICA* будет экспортировать имена переменных в первую строку таблицы Quattro. Если указанный режим не выбран, имена переменных не будут входить в файл Quattro.

Переменные с текстовыми значениями

"Двойные переменные" системы *STATISTICA* (числовые/текстовые) могут быть по желанию пользователя расширены в Quattro до трех столбцов, содержащих числовые и текстовые значения и, возможно, метки значений переменных системы *STATISTICA*. Имена этих столбцов будут соответствовать именам переменных в файле системы *STATISTICA*.

Созданием этих столбцов пользователь может управлять с помощью параметров, позволяющих затем конвертировать (т.е. "сжимать") пары и тройки столбцов таблицы Quattro в отдельные переменные при импорте файла в систему *STATISTICA*. Если не выбран ни один из этих вариантов, экспортируются только те переменные системы *STATISTICA*, которые не имеют текстовых значений.

Текстовые значения. При выборе этого режима система *STATISTICA* создаст для "двойных переменных" текстовые столбцы с алфавитно-цифровыми значениями. При создании текстовых столбцов для образования их имен используются имена "двойных переменных", с добавлением к ним символа *$* по следующим правилам:

- Если имя переменной системы *STATISTICA* короче восьми символов, к концу имени текстового столбца Quattro будет добавлен символ *$*;
- Если длина имени переменной системы *STATISTICA* составляет восемь символов, второй символ имени текстового столбца Quattro будет заменен на *$*. Это упрощает объединение числовых и текстовых данных при импорте файла обратно в систему *STATISTICA*.

Числовые значения. При выборе этого варианта создаются столбцы числовых значений, соответствующие столбцам текстовых значений, созданным при выборе предыдущего режима. Если указанный вариант не выбран, экспортироваться будут только числовые значения переменных, которые не имеют текстовых эквивалентов.

Длинные метки значений. С помощью этого параметра можно управлять созданием столбцов текстовых меток значений. При этом система *STATISTICA* создает третий столбец, который содержит текстовые значения, представляющие

Copyright © StatSoft, 1995

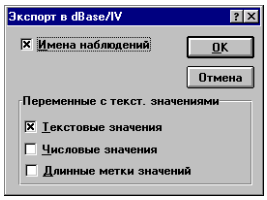
собой *длинные метки значений* переменной. Для образования его имени используется имя "двойной переменной" с добавлением к нему символа @ по следующим правилам:

- Если имя переменной системы *STATISTICA* короче восьми символов, к концу имени текстового столбца Quattro будет добавлен символ @;
- Если длина имени переменной система *STATISTICA* составляет восемь символов, второй символ имени текстового столбца Quattro будет заменен на @.

Заметим, что в случае экспорта только текстовых значений или меток имя соответствующего столбца в файле Quattro будет совпадать с именем переменной системы *STATISTICA*, а не строиться по правилам, изложенным выше, с использованием символов *$* и @.

Экспорт файлов в dBASE

В системе *STATISTICA* можно экспортировать файлы в форматы dBASE. В файлах dBASE III+ может содержаться не более 128 полей, а в файлах dBASE IV – не более 255 полей. Если размер файла системы *STATISTICA* превышает эти ограничения, пользователь может создать *подмножества* файла и экспортировать их по отдельности (см. стр. 1376) При выборе пункта *dBASE* в диалоговом окне *Экспорт данных в* откроется диалоговое окно *Экспорт в dBASE (III+ или IV)*.



Имена наблюдений

При выборе этого варианта имена наблюдений экспортируются в первое поле файла dBASE.

Переменные с текстовыми значениями

"Двойные переменные" системы *STATISTICA* (числовые/текстовые) могут быть по желанию пользователя расширены в таблице dBASE до трех полей, содержащих числовые и текстовые значения и, возможно, метки значений переменных системы *STATISTICA*. Имена этих столбцов будут соответствовать именам переменных в файле системы *STATISTICA*.

Созданием этих столбцов пользователь может управлять с помощью параметров, позволяющих затем конвертировать (т.е. "сжимать") пары и тройки полей таблицы dBASE обратно в отдельные переменные при импорте файла в систему *STATISTICA*. Если не выбран ни один из этих вариантов, экспортируются только те переменные системы *STATISTICA*, которые не имеют текстовых значений.

Текстовые значения. При выборе этого варианта система *STATISTICA* создаст для "двойных переменных" текстовые поля с алфавитно-цифровыми значениями. При создании текстовых полей система *STATISTICA* использует для образования имен этих полей имена "двойных переменных", добавляя к ним символ *$* по следующим правилам:

- Если имя переменной системы *STATISTICA* короче восьми символов, к концу имени текстового поля dBASE будет добавлен символ *$*;
- Если длина имени переменной системы *STATISTICA* составляет восемь символов, второй символ имени текстового поля dBASE будет заменен на *$*. Это упрощает объединение числовых и текстовых данных

StatSoft®
Copyright © StatSoft, 1995

при импорте файла обратно в систему *STATISTICA*.

Числовые значения. При выборе этого варианта создаются поля числовых значений, соответствующие полям текстовых значений, созданным при выборе предыдущего режима. Если указанный вариант не выбран, экспортироваться будут только числовые значения переменных, которые не имеют текстовых эквивалентов.

Длинные метки значений. С помощью этого параметра можно управлять созданием полей текстовых меток значений. При этом система *STATISTICA* создает третье поле, которое содержит текстовые значения, представляющие собой *длинные метки значений* переменной. При создании такого поля система *STATISTICA* использует для образования его имени имя "двойной переменной", добавляя к нему символ @ по следующим правилам:

- Если имя переменной системы *STATISTICA* короче восьми символов, к концу имени текстового поля dBASE будет добавлен символ @;
- Если длина имени переменной системы *STATISTICA* составляет восемь символов, второй символ имени текстового поля dBASE будет заменен на @.

Замечания

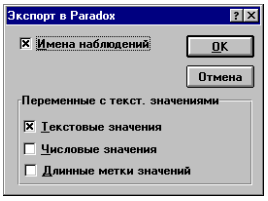
Имена текстовых переменных. В случае экспорта только текстовых значений или меток имя соответствующего поля в файле dBASE будет совпадать с именем переменной *STATISTICA*, а не строиться по правилам, изложенным выше, с использованием символов *$* и *@*.

Формат числовых значений. Хотя внутренняя точность хранения значений переменных в системе *STATISTICA* всегда двойная, при экспорте данных в dBASE (III+ или IV) сохраняется только часть каждого значения, задаваемая соответствующим форматом отображения.

Например, если значение, хранящееся в файле *STATISTICA* в виде *7.3926541009725183*, отображается в таблице исходных данных как *7.39*, то экспортированное значение будет равно *7.39*. Если вы хотите экспортировать значения с более высокой точностью, измените формат отображения переменной в диалоговом окне *текущих спецификаций таблицы исходных данных* (см. главу 3 *Окно таблицы исходных данных*).

Экспорт файлов в Paradox

В системе *STATISTICA* можно экспортировать файлы в форматы Paradox. При выборе пункта *Paradox* в диалоговом окне *Экспорт данных в* откроется диалоговое окно *Экспорт в Paradox*.



Имена наблюдений

При выборе этого варианта имена наблюдений экспортируются в первый столбец файла Paradox.

Переменные с текстовыми значениями

"Двойные переменные" системы *STATISTICA* (числовые/текстовые) могут быть по желанию пользователя расширены в базе данных Paradox до трех столбцов, содержащих числовые и текстовые значения и, возможно, метки значений

Copyright © StatSoft, 1995

StatSoft®

переменных системы *STATISTICA*. Имена этих столбцов будут соответствовать именам переменных в файле системы *STATISTICA*. Созданием этих столбцов пользователь может управлять с помощью параметров, позволяющих при импорте файла обратно в систему *STATISTICA* преобразовывать (т.е. "сжимать") пары и тройки столбцов таблицы Paradox в отдельные переменные *STATISTICA*. Если не выбран ни один из этих вариантов, экспортируются только те переменные системы *STATISTICA*, которые не имеют текстовых значений.

Текстовые значения. При выборе этого режима система *STATISTICA* создаст для "двойных переменных" текстовые столбцы с алфавитно-цифровыми значениями. При создании текстовых столбцов система *STATISTICA* использует для образования имен этих столбцов имена "двойных переменных", добавляя к ним символ *$* по следующим правилам:

- Если имя переменной системы *STATISTICA* короче восьми символов, к концу имени текстового столбца Paradox будет добавлен символ *$*;
- Если длина имени переменной системы *STATISTICA* составляет восемь символов, второй символ имени текстового столбца Paradox будет заменен на *$*. Это упрощает объединение числовых и текстовых данных при импорте файла обратно в систему *STATISTICA*.

Числовые значения. При выборе этого варианта создаются поля числовых значений, соответствующие полям текстовых значений, созданных при выборе предыдущего режима. После экспорта этих переменных в файл Paradox их имена будут совпадать с именами соответствующих переменных в системе *STATISTICA*. Если указанный вариант не выбран, экспортироваться будут только числовые значения переменных, которые не имеют текстовых эквивалентов.

Длинные метки значений. С помощью этого параметра можно управлять созданием столбцов текстовых меток значений. При этом система *STATISTICA* создает третий столбец, который содержит текстовые значения, представляющие собой *длинные метки значений* переменной. При создании такого столбца система *STATISTICA* использует для образования его имени имя "двойной переменной", добавляя к нему символ @ по следующим правилам:

- Если имя переменной системы *STATISTICA* короче восьми символов, к концу имени текстового столбца Paradox будет добавлен символ @;
- Если длина имени переменной системы *STATISTICA* составляет восемь символов, второй символ имени текстового столбца Paradox будет заменен на @.

Заметим, что в случае экспорта только текстовых значений или меток имя соответствующего столбца в файле Paradox будет совпадать с именем переменной системы *STATISTICA*, а не строиться по правилам, изложенным выше, с использованием символов *$* и @.

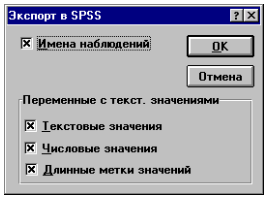
Экспорт файлов в SPSS

Этот вариант позволяет экспортировать файлы в переносимый формат SPSS, совместимый также с форматом SAS. При этом метки значений переменных системы *STATISTICA* используются в качестве меток значений в SPSS. Переменные в SPSS могут иметь имена длиной до 8 символов, которые должны начинаться с прописной латинской буквы (от A до Z) или символов *$* или @ и могут содержать цифры и знаки подчеркивания (_).

StatSoft®

Copyright © StatSoft, 1995

Определенный набор имен зарезервирован для использования в качестве ключевых слов. Имена, входящие в этот набор, а именно *ALL*, *AND*, *BY*, *EQ*, *GE*, *GT*, *LE*, *LT*, *NE*, *NOT*, *OR*, *TO*, и *WITH*, не могут использоваться в качестве имен переменных. Для проверки того, удовлетворяют ли имена переменных требованиям SPSS, можно использовать команду *Проверить имена и форматы* выпадающего меню *Анализ* (см. стр. 1378.) При выборе в диалоговом окне *Экспорт данных в* пункта *SPSS* откроется диалоговое окно *Экспорт в SPSS*.



Имена наблюдений

При выборе этого варианта имена наблюдений экспортируются в первую переменную файла SPSS.

Переменные с текстовыми значениями

В SPSS можно экспортировать различные компоненты “двойных переменных ” системы *STATISTICA*, используя приведенные ниже варианты.

Текстовые значения. Выберите этот вариант в случае, если нужно экспортировать алфавитно-цифровые значения “двойных переменных ” системы *STATISTICA* в текстовые переменные SPSS.

Числовые значения. Выберите этот вариант для экспорта числовых значений “двойных переменных ” системы *STATISTICA* в числовые переменные SPSS.

Длинные метки значений. При выборе этого режима метки значений переменных системы *STATISTICA* экспортируются в метки значений переменных SPSS или текстовые переменные SPSS (см. ниже). При различных комбинациях выбора возможны следующие варианты:

- *Текстовые значения/Числовые значения.* При выборе этой комбинации создаются *числовые* переменные SPSS с *текстовыми* значениями системы *STATISTICA* в качестве меток значений.
- *Текстовые значения/Метки.* При выборе этой комбинации создаются *текстовые* переменные SPSS с *длинными метками значений* системы *STATISTICA* в качестве меток значений.
- *Числовые значения/метки.* При выборе этой комбинации создаются *числовые* переменные SPSS с *длинными метками значений* системы *STATISTICA* в качестве меток значений.
- При выборе всех трех пунктов (*Текстовые значения*/*Числовые значения*/*Метки*) создаются два столбца: *числовой* и *текстовый* с *длинными метками значений* системы *STATISTICA* в качестве меток значений.
- При выборе только одного из вариантов будет создана переменная только одного типа (числового или текстового).
- Если не выбран ни один из вариантов, будут экспортированы только те переменные системы *STATISTICA*, которые не содержат текстовых значений.

Экспорт файлов в свободном ASCII-формате

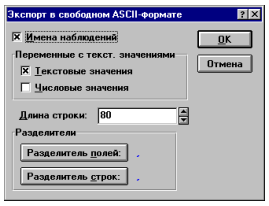
В системе *STATISTICA* можно экспортировать файлы в свободном (с разделителями) ASCII-формате. В отличие от экспорта файлов в

Copyright © StatSoft, 1995

фиксированном (без разделителей) ASCII-формате (см. стр. 1419), экспорт файлов в свободном (с разделителями) формате не требует, чтобы данные во всех переменных ASCII-файла находились в заданных (фиксированных) столбцах. В ASCII-файлах, записанных в свободном формате, для разделения значений (переменных) в наблюдениях используются разделители или ограничители, например, запятые или пробелы (см. ниже раздел *Разделители)*.

При выборе экспорта в файл в фиксированном ASCII-формате (см. стр. 1419), значения, оказавшиеся короче формата, дополняются пробелами таким образом, чтобы абсолютное положение каждого значения в строке результирующего файла было “фиксированным”. В файлах в свободном (с разделителями) ASCII-формате дополнять пробелами значения переменных не требуется, поскольку для разделения расположенных рядом значений в этом формате используются определенные пользователем символы – разделители.

При выборе в диалоговом окне *Экспорт данных в* пункта *ASCII- Свободный* открывается диалоговое окно *Экспорт в свободном ASCII-формате*.



Имена наблюдений

При выборе этого варианта имена наблюдений будут экспортированы в первую переменную ASCII - файла.

Переменные с текстовыми значениями

Выбрав один из этих вариантов, вы можете экспортировать только *текстовые* и/или *числовые* компоненты “двойных переменных” системы *STATISTICA*. При выборе обоих вариантов создаются два столбца: один для текстовых значений “двойной переменной”, а другой – для числовых значений этой же переменной. Если не выбран ни один из этих режимов, экспортироваться будут только те переменные системы *STATISTICA*, которые не содержат текстовых значений.

Текстовые значения. Выберите этот вариант в том случае, если вы хотите экспортировать текстовые значения “двойной переменной” системы *STATISTICA*.

Числовые значения. Выберите этот вариант в том случае, если вы хотите экспортировать числовые значения “двойной переменной” системы *STATISTICA*.

Длина строки

Количество переменных в строке ASCII-файла зависит от задания длины строки. Введите длину строки ASCII-файла (от 80 до 4000). Если введенная длина строки недостаточна для того, чтобы в строке содержались значения всех переменных, то после последнего полностью записанного значения переменной строка прервется (т.е. переменная не будет разделена по разным строкам). Если требуется, чтобы записи были непрерывающимися, т.е. каждая запись занимала только одну строку файла, задайте длину строки, не меньшую, чем сумма форматов всех переменных.

Разделители

Разделитель полей. Нажмите кнопку *Разделитель полей*, после чего откроется диалоговое окно *Свободный ASCII-формат* (см.

StatSoft®

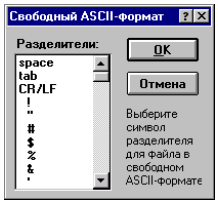
Copyright © StatSoft, 1995

ниже), в котором можно задать символ, используемый в дальнейшем для разделения значений переменной в ASCII-файле. По умолчанию в качестве такого разделителя используется запятая “,”. В качестве разделителя нельзя задавать цифры (от *0* до *9)*, десятичную точку, знаки “+” и “-”.

Разделитель строк. *Разделитель строк* используется для обозначения конца каждой из строк (наблюдений) данных (одна строка данных может представлять собой несколько строк ASCII-файла). Нажмите кнопку *Разделитель строк*, после чего откроется диалоговое окно *Свободный ASCII-формат* (см. ниже), в котором можно задать символ, используемый в дальнейшем для обозначения конца строк ASCII-файле. По умолчанию в качестве такого разделителя используется запятая “,”. В качестве разделителя нельзя задавать цифры (от *0* до *9)*, десятичную точку, знаки “+” и “-”. Заметим, что если длина строки в файле данных системы *STATISTICA* превышает заданную *длину строки* ASCII-файла (см. выше), то данные из одной строки системы *STATISTICA* будут находиться в одной или нескольких строках ASCII-файла.

Экспорт в свободный ASCII-формат - Разделители

Для задания разделителей полей или строк при экспорте в *файлы в свободном ASCII-формате* (см. выше) откройте диалоговое окно *Свободный ASCII-формат*.



В этом диалоговом окне приведен список символов, из которых пользователь может выбрать символ-разделитель. В список символов этого диалогового окна входят следующие символы (*CR/LF* обозначает возврат каретки/перевод строки):

<space> <tab> <CR/LF>

*! “ # $ % & ') * (, [\ + - .*

0 1 2 3 4 5 6 7 8 9

] ^ _ / : ; < = > ? @ ` { | } ~

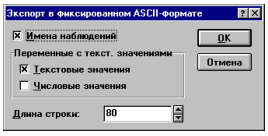
Прописные латинские буквы (т.е. *A B C . . . Z*)

Строчные латинские буквы (т.е. *a b c . . . z*)

Заметьте, что при экспорте числовых значений нельзя задавать в качестве разделителей цифры (*0 - 9*), а также знаки “+”, “-” и “.”.

Экспорт файлов в фиксированном ASCII-формате

В отличие от файлов в свободном ASCII-формате (см. стр. 1417), файлы в фиксированном ASCII-формате (без разделителей) создаются с использованием текущих форматов переменных, при этом значения, оказавшиеся короче формата, дополняются пробелами таким образом, чтобы абсолютное положение каждого значения в строке результирующего файла было “фиксированным”. При выборе в диалоговом окне *Экспорт данных в* пункта *ASCII-Фиксированный* открывается диалоговое окно *Экспорт в фиксированном ASCII-формате*.



Имена наблюдений

При выборе этого варианта имена наблюдений будут экспортированы в первую переменную ASCII-файла.

Copyright © StatSoft, 1995

StatSoft®

Переменные с текстовыми значениями

Выбрав один из этих вариантов, вы можете экспортировать только *текстовые* и/или *числовые* компоненты "двойных переменных" системы *STATISTICA*. При выборе обоих вариантов создаются два столбца: один для текстовых значений "двойной переменной", а другой – для числовых значений этой же переменной. Если не выбран ни один из этих режимов, экспортироваться будут только те переменные системы *STATISTICA*, которые не содержат текстовых значений.

Текстовые значения. Выберите этот вариант в том случае, если вы хотите экспортировать текстовые значения "двойной переменной" системы *STATISTICA*.

Числовые значения. Выберите этот вариант в том случае, если вы хотите экспортировать числовые значения "двойной переменной" системы *STATISTICA*.

Длина строки

Количество переменных в строке ASCII-файла зависит от заданной длины строки. Введите длину строки ASCII-файла (от 80 до 4000). Если введенная длина строки недостаточна для того, чтобы в строке содержались значения всех переменных, то после последнего полностью записанного значения переменной строка прервется (т.е. переменная не будет разделена по разным строкам). Если требуется, чтобы записи были непрерывающимися, т.е. каждая запись занимала только одну строку файла, задайте длину строки, не меньшую, чем сумма форматов всех переменных.

Экспорт файлов в CSS/3

Из системы *STATISTICA* для Windows можно экспортировать данные в CSS/3-формат (т.е. CSS:*STATISTICA* или *STATISTICA*/DOS). Указанные файлы DOS поддерживают большинство спецификаций формата и соглашений системы *STATISTICA* для Windows. В процессе экспорта все эти спецификации переносятся автоматически. При этом необходимо иметь в виду следующее:

- Если длина длинного имени переменной превышает 40 символов, т.е. ограничение формата в DOS-версии, то оно будет обрезано до 40 символов;
- В DOS-версии не поддерживаются форматы отображения (например, дат, времени, экспоненциальное представление, значения с запятыми и т.п.);
- В DOS-версии не поддерживается информация Рабочей книги;
- В DOS-версии не поддерживаются примечания и комментарии к файлу; экспортируется только строка заголовка файла;
- Файл должен содержать не более 300 переменных;
- В DOS-версию не экспортируется информация о цвете таблицы исходных данных, стилях и шрифтах.

StatSoft®

Copyright © StatSoft, 1995

ПРЕДМЕТНЫЙ УКАЗАТЕЛЬ

А

Автоматическое перемещение курсора, 1363
Автоматическое согласование чисел и текста, 1391
Автопересчет при изменении данных, 1352
Алфавитно-цифровые значения, 1325, 1354, 1366

Б

Буфер обмена, 1328

В

Ввод данных, 1328, 1363
Всплывающая кнопка *Далее (Cont)*, 1383

Г

Глобальные операции и операции с использованием буфера обмена
  операции над наблюдениями, 1331
  операции над переменными, 1331
Группирующие переменные, 1323

Д

Даты, 1329
Даты юлианского календаря в таблицах исходных данных, 1326
Двойное представление значений, 1325, 1336, 1354, 1366
Деление на нуль, 1352
Денежный формат отображения, 1348
Динамический обмен данными, связи, 1329, 1349
Диспетчер связей (DDE), 1330
Длинные имена переменных, 1349
Добавление файлов (см. объединение), 1335, 1372

З

Заголовок файла, 1343
Записи, определение, 1322

И

Иерархические связи между базами данных в *Менеджере мегафайлов*, 1341
Изменение наблюдений, 1378
Изменение переменных, 1378
Изменение структуры файлов данных, 1331
Имена наблюдений, 1344, 1367
Имена наблюдений в формулах таблицы исходных данных, 1350
Имена переменных, 1345, 1364
Имена переменных в формулах таблицы исходных данных, 1350
Импорт, 1389
  ASCII-файлы в свободном формате, 1402
  имена наблюдений, 1391, 1392, 1393, 1395, 1398, 1400, 1401, 1403
  просмотр диапазонов данных Lotus, 1397
  разделители в свободном ASCII-формате, 1405
  текстовые значения, 1390, 1392, 1393, 1395, 1398, 1399, 1401, 1403, 1406
  файлы dBASE, 1400
  файлы Excel, 1393
  файлы Lotus, 1395
  файлы Oracle, 1337
  файлы Paradox, 1399
  файлы Quattro, 1397
  файлы SAS, 1416
  файлы SPSS, 1392
  файлы STG, 1390
  файлы Sybase, 1337
  файлы в фиксированном ASCII-формате, 1406
  файлы *Менеджера мегафайлов*, 1390
Импорт данных, 1328, 1329
Импорт диапазона значений пропущенных данных из SPSS-файла, 1392
Импорт файлов Oracle, 1337
Импорт файлов Sybase, 1337
Интерактивные средства перекодировки, 1334
Использование перекодировки для создания новых переменных, 1334

К

Категориальные переменные, 1323
Ковариационная матрица, 1361
Комментарий к файлу, 1343
Константы в формулах таблицы исходных данных, 1351
Корреляционные матрицы, 1360

Л

Листы (экспорт в Lotus 1-2-3), 1411

М

Матрица различия, 1361
Матрица сходства, 1360
Менеджер мегафайлов, 1339, 1386
  длинные текстовые значения, 1341
  иерархические связи между базами данных, 1341
  импорт файлов *Менеджера мегафайлов*, 1390
  обзор, 1339
  обмен данными с файлами системы *STATISTICA*, 1342
  обработка, анализ данных, 1341
  открыть данные MM, 1388
  поддержка больших архивных банков данных, 1341
  предварительная обработка длинных записей исходных данных, 1341
  преобразование файла данных *STATISTICA*, 1388
  создание нового файла данных MM, 1386
  текстовые значения, 1341
  файлы данных, 1358
  язык программирования *MM*, 1339

Н

Наблюдения (определение), 1322, 1343

Copyright © StatSoft, 1995

Научный формат отображения, 1348
Недопустимое имя переменной, 1380
Недопустимые преобразования, 1352
Непрерывные переменные, 1324

О

Обмен данными с файлами системы *STATISTICA*, 1342
Обработка, анализ данных, 1341
Объединение, 1372
  наблюдения, 1335, 1375
  нереляционное, 1374
  переменные, 1335, 1373
  реляционное, 1335, 1374
  реляционное иерархическое, 1374
  текстовые значения, 1375
Операторы в формулах таблицы исходных данных, 1350
Операции метода *Перетащить и отпустить*, 1336
Операции над датами, 1327, 1328
Операции над наблюдениями и переменными с использованием буфера обмена, 1331
Операции над таблицей исходных данных с использованием буфера обмена, 1332
Операции с блоками, 1332
Операция Trunc(x), 1351
Операция *Взять из файла*, 1372

П

Перекодировка данных, 1324, 1333, 1365
Переменные (определение), 1322
Перемещение курсора, 1363
Пересчет по формулам таблицы исходных данных (переменные), 1352
Печать
  комментарии к файлам и заголовки, 1343
  переменные, спецификации, связи DDE и т.п., 1354
Повторяющиеся имена, 1379
Поддержка больших архивных банков данных в *Менеджере мегафайлов*, 1341
Подмножество файла данных, 1376
Поля (определение), 1322
Преобразование данных, 1332, 1367
Присоединить, 1376
Проверка
  данные, 1330, 1365, 1383
  имена, 1378
  синтаксис условий, 1381
  условия, 1380
  формат, 1378
Проверка данных, 1338, 1365
Программа CSS/3
  экспорт файлов CSS/3, 1420
Программа dBASE
  импорт файлов dBASE, 1400
  экспорт файлов в dBASE, 1414
Программа Excel
  импорт файлов Excel, 1393
  экспорт файлов в Excel, 1410
Программа Lotus
  импорт файлов Lotus, 1395
  листы, 1411
  просмотр диапазонов данных, 1397
  экспорт файлов в Lotus, 1411
Программа Paradox
  импорт файлов Paradox, 1399
  экспорт файлов Paradox, 1415
Программа Quattro
  импорт файлов Quattro, 1397
  экспорт файлов в Quattro, 1413
Программа SAS, 1416
Программа SPSS
  экспорт файлов в SPSS, 1416
Пропущенные данные, 1324, 1344, 1352, 1371, 1385

Р

Рабочие книги, 1336, 1343, 1420
Разделение файлов данных, 1335
Распределения и их интегралы, 1351
Редактирование файлов данных, 1331
Реляционное иерархическое объединение, 1374
Реляционное объединение, 1335, 1373

С

Свободный ASCII-формат (с разделителями)
  длина строки, 1403, 1406, 1418, 1420
  импорт файлов из, 1402
  разделители, 1405, 1419
  экспорт файлов в, 1417
Связи DDE, 1329, 1349, 1352
Связи данных, 1349
Связи динамического обмена данными, 1352
Система SPSS
  импорт диапазона значений пропущенных данных, 1392
  импорт файлов SPSS, 1392
Случайные числа, 1351
  из заданного распределения, 1351
Создание файлов данных, 1328, 1362, 1370
Сортировка данных, 1376
Спецификации переменных, 1345
Средства для импорта файлов, 1389
Средства для экспорта файлов, 1408
Средства поддержки ODBC, 1337
Стандартизация данных, 1383
Стандартизация переменных, 1333
Стандартизация переменных или блоков данных, 1383

Т

Таблица исходных данных (управление данными), 1322
  операции с использованием буфера обмена, 1332
Текстовые (алфавитно-цифровые) значения, 1325, 1354, 1366
Точность хранения данных, 1325
Транспонирование файлов данных, 1323

У

Удаление групп столбцов (строк), 1331
Управление данными, 1322
  алфавитно-цифровые значения, 1325, 1354, 1366
  архивные банки данных, 1341
  ввод данных, 1363
  группирующие переменные, 1323
  двойное представление значений, 1325, 1336, 1354, 1366
  деление на нуль, 1352
  длинные записи, 1341

StatSoft®
Copyright © StatSoft, 1995

длинные текстовые значения, 1341
заголовок файла, 1343, 1371
замена пропущенных данных средними значениями, 1385
изменение наблюдений, 1378
изменение переменных, 1378
изменение структуры файлов, 1331
имена наблюдений, 1344, 1367
имена переменных, 1345, 1364, 1370
импорт ASCII-файлов в свободном формате, 1402
импорт SPSS-файлов, 1392
импорт STG-файлов, 1390
импорт файлов dBASE, 1400
импорт файлов Excel, 1393
импорт файлов Lotus, 1395
импорт файлов Paradox, 1399
импорт файлов Quattro, 1397
импорт файлов SAS, 1416
импорт файлов в фиксированном ASCII-формате, 1406
импорт файлов данных, 1328, 1389
импорт файлов *Менеджера мегафайлов*, 1390
категориальные переменные, 1323
Командный язык системы *STATISTICA*, 1339
комментарий к файлу, 1343
Менеджер мегафайлов, 1338, 1339, 1341, 1386
метод *Перетащить и отпустить*, 1332, 1336
наблюдения, 1322, 1343
недопустимое имя переменной, 1380
недопустимые преобразования, 1352
непрерывные переменные, 1324
неразрешенные символы, 1345
обмен данными с *Менеджером мегафайлов*, 1342
объединение наблюдений, 1375
объединение переменных, 1373
объединение текстовых значений, 1375
объединение файлов данных, 1335, 1372
операции языка *STATISTICA BASIC*, 1339
организация файлов данных, 1322, 1343
перекодировка данных, 1324, 1333, 1365
переменные, 1322, 1343
перемещение курсора, 1363
повторяющиеся имена, 1379
подмножество файла данных, 1376
поля, 1322
предварительная обработка длинных записей, 1341
преобразование данных, 1332, 1367
проверка данных, 1330, 1365, 1383
проверка данных (условия), 1380
проверка имен, 1378
проверка формата, 1378
пропущенные данные, 1324, 1344, 1385
пропущенные данные, в формулах таблицы исходных данных, 1352
пропущенные данные, замена, 1385
пропущенные данные, при импорте ASCII-файлов, 1404
пропущенные значения в SPSS-переменных, 1392
Рабочие книги, 1336, 1343, 1420
разделение файлов данных, 1335
разделители в свободном ASCII-формате, 1405, 1419
распределения и их интегралы, 1351
редактирование файлов, 1331
реляционное объединение, 1335, 1374
создание файлов данных, 1328, 1362, 1370
сортировка данных, 1376
спецификации переменных, 1345
средства поддержки ODBC, 1337
стандартизация переменных, 1333
стандартизация переменных или блоков данных, 1383
текстовые значения, 1325, 1341, 1354, 1366
точность хранения данных, 1325
транспонирование файлов данных, 1323
файлы данных, 1322
файлы данных большого объема, 1341
формат матричных файлов, 1359
форматы отображения данных, 1325
формулы в таблице исходных данных, 1349, 1350, 1364
числовые и алфавитно-цифровые значения, 1325, 1354, 1366
экспорт в CSS/3, 1420
экспорт в dBASE, 1414
экспорт в Excel, 1410
экспорт в Lotus, 1411
экспорт в Paradox, 1415
экспорт в Quattro, 1413
экспорт в SAS, 1416
экспорт в SPSS, 1416
экспорт в *Менеджер мегафайлов*, 1408
экспорт в свободном ASCII-формате, 1417
экспорт в фиксированном ASCII-формате, 1419
экспорт файлов данных, 1329, 1408
Условные преобразования, 1350
Установить связь, 1330

Ф

Файлы данных, 1322, 1336
Фиксированный ASCII-формат
импорт файлов из, 1406
спецификации типов столбцов, 1403, 1407
формат, 1403
экспорт файлов в, 1419
Формат времени в таблицах исходных данных, 1327
Формат дат в таблицах исходных данных, 1326
Формат матричных файлов, 1359
Формат отображения времени, 1347

Copyright © StatSoft, 1995

Формат отображения дат, 1346
Формат отображения переменных
  время, 1347
  дата, 1346
  денежный, 1348
  научный, 1348
  проценты, 1349
  число, 1345
Формат отображения процентов, 1349
Формат файлов данных, 1343
Форматы отображения, 1345
Форматы отображения данных, 1325
Форматы числовых данных, 1325
Формулы в таблице исходных данных, 1333, 1349, 1350, 1351, 1352, 1364
  автопересчет при изменении данных, 1352
  распределения и их интегралы, 1351
  случайные числа, 1351
Функции в формулах таблицы исходных данных, 1351
Функция Abs(x), 1351
Функция Cos(x), 1351
Функция Exp(x), 1351
Функция Log(x), 1351
Функция Log2(x), 1351
Функция Rnd(x), 1351
Функция Sign(x), 1351
Функция Sin(x), 1351
Функция Sqrt(x), 1351

Ч

Число Эйлера (число *e*), 1351
Числовой формат отображения, 1345
Числовой эквивалент, 1325

Э

Экспорт, 1329, 1408
  в *Менеджер мегафайлов*, 1408
  разделители в свободном ASCII-формате, 1419
  файлы CSS/3, 1420
  файлы dBASE, 1414
  файлы Excel, 1410
  файлы Lotus, 1411
  файлы Paradox, 1415
  файлы Quattro, 1413
  файлы SAS, 1416
  файлы SPSS, 1416
  файлы в свободном ASCII-формате, 1417
  файлы в фиксированном ASCII-формате, 1419

Я

Язык *STATISTICA BASIC*, 1338, 1367
Язык программирования *MM*, 1339

StatSoft®
Copyright © StatSoft, 1995

Глава 8:

ЭЛЕМЕНТАРНЫЕ ПОНЯТИЯ СТАТИСТИКИ

Оглавление

ОБЗОР ЭЛЕМЕНТАРНЫХ ПОНЯТИЙ СТАТИСТИКИ 1426
Что такое переменные? 1426
Исследование зависимостей и с экспериментальные исследования 1426
Зависимые и независимые переменные 1426
Шкалы измерений 1427
Зависимости между переменными 1428
Почему зависимости между переменными являются важными 1428
Две основные черты всякой зависимости между переменными 1428
Что такое Статистическая значимость (*p-уровень*)? 1429
Как определить, является ли результат действительно значимым 1429
Статистическая значимость и число испытаний 1430
Величина зависимости между переменными в сравнении с надежностью зависимости 1430
Почему более сильные зависимости между переменными являются более значимыми 1430
Почему объем выборки влияет на значимость зависимости 1430
Пример: “отношение числа мальчиков к числу девочек” 1431
Почему слабые зависимости могут быть значимо доказаны только на больших выборках 1431
Можно ли отсутствие связей рассматривать как значимый результат? 1432
Как измерить величину зависимости между переменными 1432
Общая конструкция большинства статистических тестов 1433
Как вычисляется статистическая значимость 1433
Почему важно “Нормальное распределение” 1434
Иллюстрация того, как нормальное распределение используется в статистических рассуждениях 1434
Все ли статистики критериев нормально распределены? 1434
Как узнать последствия нарушений предположений нормальности? 1435

ПРЕДМЕТНЫЙ УКАЗАТЕЛЬ 1436

Copyright © StatSoft, 1995

Глава 8:

ЭЛЕМЕНТАРНЫЕ ПОНЯТИЯ СТАТИСТИКИ

ОБЗОР ЭЛЕМЕНТАРНЫХ ПОНЯТИЙ СТАТИСТИКИ

Это введение представляет собой краткое обсуждение элементарных статистических понятий, лежащих в основе конкретных процедур в любой области статистического анализа данных. Мы выбрали темы, которые иллюстрируют основные допущения, принимаемые в большинстве статистических методов для понимания “численной природы” действительности (см. работу Nisbett, and all., 1987).

Ввиду ограниченности объема мы вынуждены сосредоточить внимание на основных "функциональных" аспектах обсуждаемых понятий, прекрасно понимая, что предлагаемое описание будет кратким. Дальнейшую информацию о каждом из этих понятий можно найти в разделах *Вводный обзор* и *Примеры* этого руководства, а также в учебниках по статистике. Мы рекомендуем следующие учебники: Kachigan (1986) и Runyon and Haber (1976); для углубленного обсуждения элементарной теории и основных понятий статистики см. классические книги Hays (1988), и Kendall and Stuart (1979).

Что такое переменные?

Переменные – это то, что можно измерить, контролировать или чем можно манипулировать в исследованиях. Понятно, что переменные отличаются многими аспектами, особенно той ролью, которую они играют в исследованиях, типом измерений и т.д.

Исследование зависимостей в сравнении с экспериментальными исследованиями

Большинство эмпирических исследований данных можно отнести к одному из названных в заголовке типов . В исследовании корреляций вы не влияете (или, по крайней мере, пытаетесь не влиять) на какие-либо переменные, а только измеряете их и хотите найти зависимости (корреляции) между некоторыми из измеренных переменных, например, между кровяным давлением и уровнем холестерина. В экспериментальных исследованиях, напротив, вы варьируете некоторые переменные и измеряете воздействия этих изменений на другие переменные. Например, исследователь может искусственно увеличить кровяное давление, а затем измерить уровень холестерина. Анализ данных в экспериментальном исследовании также приходит к вычислению "корреляций" между переменными, именно, между переменными, на которые воздействуют, и теми переменными, на которые влияет это воздействие. Тем не менее, экспериментальные данные потенциально снабжают нас более качественной информацией. Только экспериментальные данные могут убедительно продемонстрировать наличие причинных связей между переменными. Например, если вы обнаружили, что всякий раз, когда изменяется переменная *A,* изменяется и переменная *B*, то можете сделать вывод – “переменная *A* оказывает влияние на *B*.” Результаты *корреляционного* исследования могут быть проинтерпретированы в каузальных (причинных) терминах на основе некоторой теории, но сами по себе не могут отчетливо доказать причинность.

Зависимые и независимые переменные

Независимыми переменными называются переменные, которые варьируются исследователем, тогда как *зависимые* переменные – это переменные, которые измеряются или регистрируются. Может показаться, что

StatSoft®

Copyright © StatSoft, 1995

проведение этого различия создает путаницу в терминологии, поскольку как говорят некоторые студенты “все переменные зависят от чего-нибудь”. Тем не менее, однажды отчетливо проведя это различие, вы поймете его необходимость. Термины *зависимая* и *независимая* переменная применяются в основном в экспериментальном исследовании, где экспериментатор манипулирует некоторыми переменными, и в этом смысле они “независимы” от реакций, свойств, намерений и т.д. присущих объектам исследования. Некоторые другие переменные, как предполагается, должны "зависеть" от действий экспериментатора или от экспериментальных условий. Иными словами, зависимость проявляется в ответной реакции исследуемого объекта. Отчасти в противоречии с данным разграничением понятий находится использование их в исследованиях, где вы не варьируете независимые переменные, а только приписываете объекты к “экспериментальным группам”, основываясь на некоторых их предэкспериментальных свойствах. Например, если в эксперименте мужчины сравниваются с женщинами относительно числа лейкоцитов (WCC), то *Пол* можно назвать *независимой* переменной, а *WCC зависимой* переменной.

Шкалы измерений

Переменные различаются также тем "насколько хорошо" они могут быть измерены, или другими словами, как много измеряемой информации обеспечивает шкала их измерений. Очевидно, в каждом измерении присутствует некоторая ошибка, определяющая границы "количества информации", какую можно получить в данном измерении. Другим фактором, определяющим количество информации, содержащейся в переменной, является тип шкалы, в которой проведено измерение. Различают следующие типы шкал:(a) *номинальная*, (b) *порядковая (ординальная)*, (c) *интервальная* (d) *относительная* (*шкала отношения*).

Соответственно, имеются четыре типа переменных: (a) *номинальная*, (b) *порядковая (ординальная)*, (c) *интервальная* и (d) *относительная*.

(a) *Номинальные* переменные используются только для *качественной классификации*. Это означает, что данные переменные могут быть измерены только в терминах принадлежности к некоторым существенно различным классам, при этом вы не сможете определить количество или упорядочить эти классы. Например, вы можете сказать только, что 2 индивидуума различимы в терминах данной переменной A (например, принадлежат к разным расам), но не сможете сказать, у кого из них данное качество выражено в большой степени. Типичными примерами *номинальных* переменных являются пол, национальность, цвет глаз, город и т.д. Иногда номинальные переменные называют категориальными.

(b) *Порядковые* переменные позволяют ранжировать (упорядочить) объекты, указав какие из них в большей или меньшей степени обладают качеством, выраженным данной переменной. Однако они не позволяют сказать “на сколько больше” или “на сколько меньше”. Порядковые переменные иногда также называют ординальными. Типичный пример – социоэкономический статус семьи. Мы понимаем, что *верхний средний уровень* выше *среднего уровня*, однако сказать, что разница между ними равна, скажем, 18% мы не сможем. Само расположение шкал в следующем порядке: *номинальная*, *порядковая*, *интервальная* является хорошим примером порядковой переменной. Например, можно сказать, что измерения в *номинальной* шкале доставляют меньше информации, чем в *порядковой* шкале, а в *порядковой* меньше, чем в *интервальной*. Однако вы не сможете придать термину "меньше" точный количественный смысл или сравнить между собой эти различия.

Copyright © StatSoft, 1995

(c) *Интервальные* переменные позволяют не только упорядочивать объекты измерения, но и численно выражать и сравнивать различия между ними. Например, температура, измеренная в градусах Фаренгейта или Цельсия, образует интервальную шкалу. Вы можете не только сказать, что температура 40 градусов выше, чем температура 30 градусов, но и что увеличение температуры с 20 до 40 градусов вдвое больше увеличения температуры от 30 до 40 градусов.

(d) *Относительные* переменные очень похожи на интервальные переменные. В дополнение ко всем свойствам переменных, измеренных в интервальной шкале, их характерной чертой является наличие определенной точки абсолютного нуля, таким образом, для этих переменных являются обоснованными предложения типа: *x* в два раза больше, чем *y*. Типичными примерами шкал отношений являются измерения времени или пространства. Например, температура по Кельвину образует шкалу отношения, и вы можете не только утверждать, что температура 200 градусов выше, чем 100 градусов, но и что она вдвое выше. Интервальные шкалы (например, шкала Цельсия) не обладают данным свойством шкалы отношения. Однако в большинстве статистических процедур не делается различия между свойствами интервальных шкал и шкал отношения.

Зависимости между переменными

Независимо от типа, две или более переменных связаны (зависимы) между собой, если наблюдаемые значения этих переменных распределены согласованным образом. Другими словами, мы говорим, что переменные зависимы, если их значения систематическим образом согласованы друг с другом в имеющихся у нас наблюдениях. Например, переменные *Пол* и *WCC* (число лейкоцитов) могли бы рассматриваться как зависимые, если бы большинство мужчин имело высокий уровень *WCC,* а большинство женщин – низкий *WCC*, или наоборот. *Рост* связан с *Весом,* потому что обычно высокие индивиды тяжелее низких; *IQ* (коэффициент интеллекта) связан с *Количеством ошибок* в тесте, т.к. люди высоким значением IQ делают меньше ошибок.

Почему зависимости между переменными являются важными

Вообще говоря, конечная цель всякого исследования или научного анализа – нахождение связей (зависимостей) между переменными. Философия науки учит, что не существует иного способа представления знания, кроме как в терминах зависимостей между количествами или качествами.

Таким образом, развитие науки всегда заключается в нахождении новых зависимостей между переменными. Исследование корреляций по существу состоит в измерении таких зависимостей непосредственным образом. Тем не менее, экспериментальное исследование не является в этом смысле чем-то отличным. Например, отмеченное выше экспериментальное сравнение *WCC* у мужчин и женщин может быть описано как поиск связи между двумя переменными: *Пол* и *WCC*. Назначение статистики состоит в том, чтобы помочь вам оценить зависимости между переменными. Действительно, все сотни описанных в данном руководстве процедур могут быть проинтерпретированы в терминах оценки различных типов взаимозависимостей между переменными.

Две основные черты всякой зависимости между переменными

Можно отметить два самых простых свойства зависимости между переменными: (a) величина зависимости и (b) надежность зависимости.

StatSoft®

Copyright © StatSoft, 1995

(a) *Величина.* Величина зависимости легче понять и измерить, чем *надежность*. Например, если любой мужчина в вашей выборке имел значение *WCC* выше чем любая женщина, то вы можете сказать, что величина зависимости между двумя переменными (*Пол* и *WCC*) очень высокая. Другими словами, вы могли бы предсказать значения одной переменной по значениям другой.

(b) *Надежность ("истинность"). Надежность* взаимозависимости – менее наглядное понятие, чем величина зависимости, однако чрезвычайно важное. Оно непосредственно связано с репрезентативностью той определенной выборки, на основе которой строятся выводы. Другими словами, надежность говорит, насколько вероятно, что зависимость, подобная найденной вами, будет вновь обнаружена (иными словами, подтвердится) на данных другой выборки, извлеченной из той же самой популяции. Следует помнить, что конечной целью почти никогда не является изучение данной конкретной выборки; выборка представляет интерес лишь постольку, поскольку она дает информацию обо всей популяции. Если ваше исследование удовлетворяет некоторым специальным критериям (об этом будет сказано позже), то надежность найденных зависимостей между переменными вашей выборки можно количественно оценить и представить с помощью стандартной статистической меры (называемой *p-уровень* или статистический *уровень значимости*, смотри следующий раздел).

Что такое Статистическая значимость (*p-уровень*)?

Статистическая значимость результата представляет собой оцененную меру уверенности в его "истинности" (в смысле "репрезентативности выборки"). Выражаясь более технически, *p-уровень* это показатель, находящийся в убывающей зависимости от надежности результата. Более высокий *p-уровень* соответствует более низкому уровню доверия к найденной в выборке зависимости между переменными. Именно, *p-уровень* представляет собой вероятность ошибки, связанной с распространением наблюдаемого результата на всю популяцию. Например, *p-уровень = .05* (т.е. $^1/_{20}$) показывает, что имеется 5% вероятность того, что найденная в выборке зависимость между переменными является лишь случайной особенностью данной выборки. Иными словами, если данная зависимость в популяции отсутствует, а вы многократно проводили бы подобные эксперименты, то примерно в одном из двадцати повторений эксперимента можно было бы ожидать такой же или более сильной зависимости между изучаемыми переменными. Во многих исследованиях *p-уровень* .05 рассматривается как "приемлемая граница" уровня ошибки.

Как определить, является ли результат действительно значимым

Не существует никакого способа избежать произвола при принятии решения о том, какой уровень значимости следует действительно считать "значимым". Выбор определенного уровня значимости, выше которого результаты отвергаются как ложные, является достаточно произвольным. На практике окончательное решение обычно зависит от того, был ли результат предсказан априори (т.е. до проведения опыта) или обнаружен *апостериорно* в результате многих анализов и сравнений, выполненных с множеством данных, а также на традиции, имеющейся в данной области исследований.

Обычно во многих областях результат $p \leq .05$ является приемлемой границей статистической значимости, однако следует помнить, что этот уровень все еще включает довольно большую вероятность ошибки (5%). Результаты, значимые на уровне $p \leq .01$ обычно рассматриваются как

Copyright © StatSoft, 1995

StatSoft®

статистически значимые, а результаты с уровнем $p \leq .005$ или $p \leq .001$ как высоко значимые. Однако следует понимать, что данная классификация уровней значимости достаточно произвольна и является всего лишь неформальным соглашением, принятым на основе практического опыта.

Статистическая значимость и число испытаний

Понятно, что чем большее число анализов вы провели над совокупностью данных, тем большее число результатов среди них имеют шанс удовлетворить выбранному уровню значимости. Например, если вычисляют корреляции между 10 переменными (т.е. имеется 45 различных коэффициентов корреляции), то можно ожидать, что примерно два коэффициента корреляции (один на каждые 20) случайно окажутся значимыми на уровне $p \leq .05$, даже если переменные совершенно случайны и некоррелированы в популяции.

Некоторые статистические методы, включающие много сравнений, и, таким образом, имеющие хороший шанс повторить такого рода ошибки, используют специальную корректировку или поправку на общее число сравнений. Тем не менее, многие статистические методы (особенно простые методы разведочного анализа данных) не предлагают какого-либо способа решения этой проблемы. Поэтому исследователь должен с осторожностью оценивать надежность неожиданных находок. Многие примеры, обсуждаемые в данном руководстве, предлагают специальные советы как это сделать.

Величина зависимости между переменными в сравнении с надежностью зависимости

Как было уже сказано, величина зависимости и надежность представляют собой две различные характеристики зависимостей между переменными. Тем не менее, нельзя сказать, что они совершенно независимы. Говоря общим языком, чем больше величина зависимости (связи) между переменными в выборке обычного объема, тем более она надежна (см. следующий раздел).

Почему более сильные зависимости между переменными являются более значимыми

Если предполагать отсутствие зависимости между соответствующими переменными в популяции, то наиболее вероятно ожидать, что в исследуемой выборке связь между этими переменными также будет отсутствовать. Таким образом, чем более сильная зависимость обнаружена в выборке, тем менее вероятно, что этой зависимости нет в популяции, из которой она извлечена. Как вы видите, величина зависимости и значимость тесно связаны между собой, и можно было бы попытаться вывести значимость из величины зависимости и наоборот. Однако указанная связь между зависимостью и значимостью имеет место только при фиксированном объеме выборки, поскольку при различных объемах выборки одна и та же зависимость может оказаться как высоко значимой, так и незначимой вовсе (см. следующий раздел).

Почему объем выборки влияет на значимость зависимости

Если наблюдений мало, то соответственно имеется мало возможных комбинаций значений этих переменных и таким образом, вероятность случайно обнаружить комбинацию значений, показывающую сильную зависимость, относительно высока. Рассмотрим следующий пример. Если вы исследуете зависимость двух переменных (*Пол*: мужчина/женщина и *WCC*: высокий/низкий) и имеете только 4 субъекта в выборке (2 мужчины и 2 женщины), то вероятность того, что чисто случайно вы найдете 100% зависимость между двумя переменными

StatSoft®

Copyright © StatSoft, 1995

равна 1/8. А именно, вероятность того, что оба мужчины имеют высокий *WCC,* а обе женщины – низкий *WCC*, или наоборот, – равна 1/8. Теперь рассмотрим вероятность подобного совпадения для 100 субъектов; легко видеть, что эта вероятность равна практически нулю.

Рассмотрим более общий пример. Представим популяцию, в которой среднее значение *WCC* для мужчин и женщин одно и тоже. Если теперь вы начнете повторять эксперимент, состоящий в извлечении пары случайных выборок (одна – мужчины, другая – женщины), и вычислении разности выборочных средних *WCC* для каждой пары, то в большинстве экспериментов результат будет близок к 0. Однако, время от времени, будут встречаться пары выборок, в которых различие между мужчинами и женщинами будут существенно отличаться от 0. Как часто будет это происходить? Чем меньше объем выборки в каждом эксперименте, тем более вероятно появление таких ложных результатов, которые показывают существование зависимости между полом и *WCC* в данных, полученных из популяции, где такая зависимость на самом деле отсутствует.

Пример: “отношение числа мальчиков к числу девочек”

Рассмотрим следующий пример, заимствованный и адаптированный из исследования, посвященного статистическим рассуждениям (см. работу Nisbett, et al., 1987). Имеются 2 больницы. В первой из них каждый день рождается 120 детей, во втором только 12. В среднем отношение числа мальчиков, рождающихся в каждой больнице, к числу девочек есть 50/50. Однажды в одной из этих больниц девочек родилось вдвое больше, чем мальчиков. В какой из больниц это событие более вероятно?

Ответ очевиден для статистика, однако, как показывает исследование, он не столь очевиден для дилетанта. Конечно, такое событие гораздо более вероятно для маленькой больницы. Объяснение этого факта состоит в том, что вероятность случайного отклонения (от среднего популяции) возрастает с уменьшением объема выборки.

Почему слабые зависимости могут быть значимо доказаны только на больших выборках

Предыдущий пример показывает, что если зависимость между переменными "объективно" (другими словами, в популяции) мала, не существует иного способа проверить такую зависимость кроме как исследовать выборку достаточно большого объема. Даже если ваша выборка совершенно репрезентативна, эффект не будет статистически значимым, если выборка мала. Аналогично, если зависимость "объективно" (другими словами, в популяции) очень сильная, тогда она может быть обнаружена с высокой значимостью даже на очень маленькой выборке. Рассмотрим следующий иллюстративный пример. Если монета слегка несимметрична, и при подбрасывании орел выпадает чаще решки (например, 60% против. 40%), то 10 подбрасываний монеты было бы не достаточно, чтобы убедить кого бы то ни было, что монета асимметрична, даже если был бы получен совершенно репрезентативный результат, 6 орлов и 4 решки. Не следует ли отсюда, что 10 подбрасываний вообще не могут доказать что-либо? Нет, не следует, потому что если эффект в принципе очень сильный, 10 подбрасываний может быть вполне достаточно. Представьте, что монета настолько несимметрична, что всякий раз, когда вы ее бросаете, выпадает орел. Если вы бросаете такую монету 10 раз, и всякий раз выпадает орел, большинство людей сочтут это убедительным доказательством того, что с монетой что-то не то.

Другими словами, это послужило бы убедительным доказательством того, что в популяции, состоящей из бесконечного числа

Copyright © StatSoft, 1995

подбрасываний этой монеты орел будет встречаться чаще, чем решка. Таким образом, если зависимость сильная, она может быть обнаружена с высоким уровнем значимости даже на малой выборке.

Можно ли отсутствие связей рассматривать как значимый результат?

Чем слабее зависимость между переменными, тем большего объема требуется выборка, чтобы значимо ее обнаружить. Например, представьте как много бросков монеты необходимо сделать, чтобы доказать, что отклонение от равновероятности составляет только .000001%! Таким образом, необходимый минимальный размер выборки возрастает, когда степень эффекта, который нужно доказать, убывает. Когда эффект близок к 0, необходимый объем выборки для его отчетливого доказательства приближается к бесконечности. Другими словами, если зависимость между переменными почти отсутствует, объем выборки, необходимый для ее значимого обнаружения, почти равен объему всей популяции, который предполагается бесконечным. Статистическая значимость представляет вероятность того, что подобный результат был бы получен при проверке всей популяции в целом. Таким образом, все, что получено после тестирования всей популяции было бы, по определению, значимым на наивысшем возможном уровне и это относится ко всем результатам типа “нет связи”.

Как измерить величину зависимости между переменными

Статистиками разработано много различных мер зависимости между переменными. Выбор определенной меры в конкретном исследовании зависит от числа включенных в анализ переменных, используемых шкал измерения, природы зависимостей и т.д. Большинство этих мер, тем не менее, подчиняются одному общему принципу: они пытаются как-то оценить наблюдаемую зависимость, сравнивая ее с “максимальной мыслимой зависимостью” между рассматриваемыми переменными.

Говоря технически, обычный способ выполнить такие оценки заключается в том, чтобы посмотреть как варьируются значения переменных и затем подсчитать, какая часть всей имеющейся вариации может быть объяснена наличием "общей" ("совместной") вариации двух (или более) переменных.

Говоря менее технически, вы сравниваете то “что есть общего в этих переменных”, с тем “что потенциально было бы у них общего, если бы переменные были абсолютно зависимы”. Рассмотрим простой пример. Пусть в вашей выборке, средний показатель (число лейкоцитов) *WCC* равен 100 для мужчин и 102 для женщин. Следовательно, вы могли бы сказать, что отклонение каждого индивидуального значения от общего среднего (101) содержит компоненту связанную с полом субъекта и средняя величина ее равна 1. Это значение, таким образом, представляет некоторую меру зависимости между переменными *Пол* и *WCC*. Конечно, это очень бедная мера, так как она не дает никакой информации о том, насколько велика эта компонента, скажем относительно общего изменения значений *WCC*. Рассмотрим две крайние возможности:

(a) Если все значения *WCC* у мужчин были бы точно равны 100, а у женщин 102, то все отклонения значений от общего среднего в выборке всецело объяснялись бы полом. Поэтому вы могли бы сказать, что пол абсолютно коррелирует с *WCC*, иными словами, 100% наблюдаемых различий между субъектами в значениях *WCC* объясняются полом субъектов.

(b) Если же значения *WCC* лежат в пределах 0-1000, то та же самая разность (2) между средними значениями *WCC* у мужчин и

StatSoft®

Copyright © StatSoft, 1995

женщин, обнаруженная в эксперименте, составляла бы столь малую долю общей вариации, что полученное различие считалось бы пренебрежимо малым. Например, введение в рассмотрение еще одного субъекта могло бы изменить разность или даже изменить ее знак. Поэтому всякая хорошая мера зависимости должна принимать во внимание полную изменчивость индивидуальных значений в выборке и оценивать зависимость по тому, насколько эта изменчивость объясняется изучаемой зависимостью.

Общая конструкция большинства статистических тестов

Так как конечная цель большинства статистических тестов состоит в оценивании зависимости между переменными, большинство статистических тестов следуют общему принципу, объясненному в предыдущем разделе. Говоря техническим языком, эти тесты представляют собой отношение изменчивости, общей для рассматриваемых переменных, к полной изменчивости. Например, такой тест может представлять собой отношение той части изменчивости *WCC,* которая определяется полом к *полной* изменчивости *WCC* (вычисленной для объединенной выборки мужчин и женщин). Это отношение обычно называется отношением *объясненной* вариации к *полной* вариации.

В статистике термин *объясненная вариация* не обязательно означает, что вы даете ей "теоретическое объяснение". Он используется только для обозначения *общей вариации* рассматриваемых переменных, иными словами, для указания на то, что часть вариации одной переменной "объясняется" определенными значениями другой переменной и наоборот.

Как вычисляется статистическая значимость

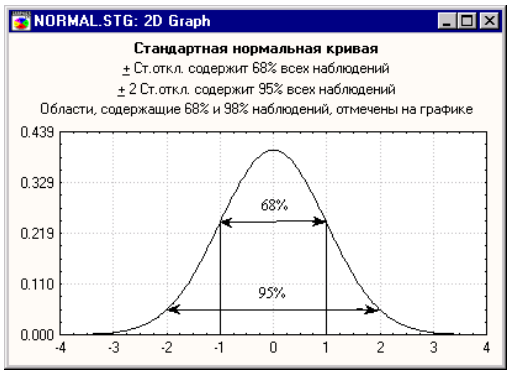
Предположим, вы уже вычислили меру зависимости между двумя переменными (как объяснялось выше). Следующий вопрос, стоящий перед вами: “насколько значима эта зависимость?” Например, является ли 40% объясненной дисперсии между двумя переменными достаточным, чтобы считать зависимость значимой? Ответ: “в зависимости от обстоятельств”. Именно, значимость зависит в основном от объема выборки. Как уже объяснялось, в очень больших выборках даже очень слабые зависимости между переменными будут значимыми, в то время как в малых выборках даже очень сильные зависимости не являются надежными (значимыми). Таким образом, для того чтобы определить уровень статистической значимости, вам нужна функция, которая представляла бы зависимость между “величиной” и “значимостью” зависимости между переменными для каждого объема выборки. Данная функция указала бы вам точно “насколько вероятно получить зависимость данной величины (или больше) в выборке данного объема, в предположении, что в популяции такой зависимости нет.” Другими словами, эта функция давала бы вам уровень значимости (p -уровень), и, следовательно, вероятность ошибочно отклонить предположение об отсутствии данной зависимости в популяции. Эта “альтернативная” гипотеза (состоящая в том, что нет зависимости в популяции) обычно называется *нулевой гипотезой*. Было бы идеально, если бы функция, вычисляющая вероятность ошибки, была линейна и имела только различные наклоны для разных объемов выборки. К сожалению, эта функция существенно более сложная и не всегда точно одна и та же. Тем не менее, в большинстве случаев ее форма известна, и ее можно использовать для определения уровней значимости при исследовании выборок заданного

Copyright © StatSoft, 1995

размера. Большинство этих функций связано с очень важным классом, называемым *нормальным*.

Почему важно “Нормальное распределение”

“Нормальное распределение” важно по многим причинам. В большинстве случаев оно является хорошим приближением функций введенных в предыдущем параграфе. Распределение многих статистик является нормальным или может быть получено из нормальных с помощью некоторых преобразований.



Рассуждая философски, можно сказать, что нормальное распределение представляет собой одну из эмпирически проверенных истин относительно общей природы действительности и его положение может рассматриваться как один из фундаментальных законов природы. Точная форма нормального распределения (характерная “колоколообразная кривая”) определяется только двумя параметрами: *средним* и *стандартным отклонением*.

Характерное свойство нормального распределения состоит в том, что 68% из всех его наблюдений лежат в диапазоне

±1 стандартное отклонение от среднего, а диапазон ±2 стандартных отклонения включает 95% значений. Другими словами, при нормальном распределении, стандартизованные наблюдения, меньшие -2 или большие +2, имеют относительную частоту менее 5% (Стандартизованное наблюдение означает, что из исходного значения вычтено среднее и результат поделен на стандартное отклонение.)

Иллюстрация того, как нормальное распределение используется в статистических рассуждениях

Напомним пример, обсуждавшийся на стр. 1431, когда пары выборок мужчин и женщин выбирались из совокупности, в которой среднее значение *WCC* для мужчин и женщин было в точности одно и то же. Хотя наиболее вероятный результат таких экспериментов (одна пара выборок на эксперимент) состоит в том, что разность между средними *WCC* для мужчин и женщин для каждой пары близка к 0, время от время появляются пары выборок, в которых эта разность существенно отличается от 0. Как часто это происходит? Если объем выборок достаточно большой, то разности “нормально распределены” и зная форму нормальной кривой, вы можете точно рассчитать вероятность случайного получения результатов, представляющих различные уровни отклонения среднего от 0 – значения гипотетического для всей популяции. Если вычисленная вероятность настолько мала, что удовлетворяет принятому заранее уровню статистической значимости (см. стр. 1429), то вы можете сделать лишь один вывод: ваш результат лучше описывает свойства популяции, чем "нулевая гипотеза". Следует помнить, что нулевая гипотеза рассматривается только по техническим соображениям как начальная точка, с которой сопоставляются эмпирические результаты.

Все ли статистики критериев нормально распределены?

Не все, но большинство из них либо имеют нормальное распределение, либо имеют распределение, связанное с нормальным и

StatSoft®
Copyright © StatSoft, 1995

вычисляемое на основе нормального, такое как *t*, *F*, или *хи-квадарт*. Обычно эти критериальные статистики требуют, чтобы анализируемые переменные сами были нормально распределены в совокупности, т.е. удовлетворяли бы "предположению". Многие наблюдаемые переменные действительно нормально распределены, что является еще одним аргументом в пользу того, что нормальное распределение представляет "фундаментальный закон". Проблема может возникнуть, когда пытаются применить тесты, основанные на предположении нормальности, к данным, не являющимся нормальными (смотри критерии нормальности в модулях *Непараметрическая статистика и распределения* или *ANOVA-/MANOVA*). В этих случаях вы можете выбрать одно из двух. Во-первых, вы можете использовать альтернативные "непараметрические" тесты (или так называемые "свободно распределенные тесты", см. главу 11 в этом томе, *Непараметрическая статистика и распределения*). Однако это часто неудобно, потому что обычно эти критерии имеют меньшую мощность и обладают меньшей гибкостью. Как альтернативу, во многих случаях вы можете все же использовать тесты, основанные на предположении нормальности, если уверены, что объем выборки достаточно велик. Последняя возможность основана на чрезвычайно важном принципе, позволяющем понять популярность тестов, основанных на нормальности. А именно, при возрастании объема выборки, форма распределения статистики критерия приближается к нормальной, даже если распределение исследуемых переменных не является нормальным. Этот принцип называется *центральной предельной теоремой*.

Как узнать последствия нарушений предположений нормальности?

Хотя многие утверждения предыдущих параграфов можно доказать математически, некоторые из них не имеют теоретического обоснования и могут быть продемонстрированы только эмпирически, с помощью так называемых экспериментов Монте-Карло. В этих экспериментах большое число выборок генерируется на компьютере, а результаты полученные из этих выборок, анализируются с помощью различных тестов. Этим способом можно эмпирически оценить тип и величину ошибок или смещений, которые вы получаете, когда нарушаются определенные теоретические предположения тестов, используемых вами. Монте-Карловские исследования интенсивно использовались для того, чтобы оценить, насколько тесты, основанные на предположении нормальности, чувствительны к различным нарушениям предположений нормальности. Общий вывод этих исследований состоит в том, что последствия нарушения предположения нормальности менее фатальны, чем первоначально предполагалось. Хотя эти выводы не означают, что предположения нормальности можно игнорировать, они увеличили общую популярность тестов, основанных на нормальном распределении.

Copyright © StatSoft, 1995

ПРЕДМЕТНЫЙ УКАЗАТЕЛЬ

А

Альтернативная гипотеза (элементарные понятия), 1433
априорное обнаружение (элементарные понятия), 1429

В

Величина зависимости между переменными (элементарные понятия), 1429
Величина зависимости между переменными в сравнении с надежностью зависимости (элементарные понятия), 1430
Вероятность ошибки (элементарные понятия), 1429
Высоко значимые результаты (элементарные понятия), 1430

Г

Граница статистически *значимого* результата (элементарные понятия), 1429

З

Зависимые и независимые переменные (элементарные понятия), 1426

И

Интервальные переменные (элементарные понятия), 1428
Исследование корреляций в сравнении. с экспериментальными исследованиями (элементарные Понятия), 1426

К

Количественное представление знания (элементарные понятия), 1428
колоколообразная кривая (элементарные понятия), 1434
Конечная цель всякого исследования (элементарные понятия), 1428

М

Мера уровня значимости (элементарные понятия), 1433
Монте-Карло эксперименты (элементарные понятия), 1435

Н

Надежность зависимости между переменными (элементарные понятия), 1429
Надежность неожиданных находок (элементарные понятия), 1430
Нарушение предположения нормальности (элементарные понятия), 1435
Независимые переменные (элементарные понятия), 1426
Номинальные переменные (элементарные понятия), 1427
Нормально распределенные статистики (элементарные понятия), 1434
Нормальное распределение (элементарные понятия), 1434
Нулевая гипотеза (элементарные понятия), 1433
Нулевой результат (элементарные понятия), 1432

О

Общая конструкция большинства статистических тестов (элементарные понятия), 1433
Общее отличие индивидуальных значений (элементарные понятия), 1433
Объясненная вариация (элементарные понятия), 1433
Относительные переменные (элементарные понятия), 1428
Отношение *объясненной* вариации к *полной* вариации (элементарные понятия), 1433

П

Переменная (определение), 1426
Порядковые переменные (элементарные понятия), 1427
Предположение нормальности (элементарные понятия), 1435
Предположения нормальности (элементарные понятия), 1435
Причинные связи между переменными (элементарные понятия), 1426

Р

Размер выборки (элементарные понятия), 1430
Распределение построенное по выборке (элементарные понятия), 1435
Репрезентативность (элементарные понятия), 1429

С

Связи между переменными, определение (элементарные понятия), 1428
Стандартизованные значения (элементарные понятия), 1434
Статистическая значимость (элементарные понятия), 1429
Статистическая значимость в зависимости от размера выборки (элементарные понятия), 1430

У

Уровень значимости (элементарные понятия), 1429
p-уровень, 1429

Ф

Форма нормального распределения (элементарные понятия), 1434

Ц

Центральная предельная теорема (элементарные понятия), 1435

StatSoft®

Copyright © StatSoft, 1995

Ш

Шкалы измерений (элементарные понятия), 1427

Э

Экспериментальное исследование, 1426
Элементарные понятия статистики, 1426

Copyright © StatSoft, 1995

StatSoft®

Copyright © StatSoft, 1995

Глава 9:

ОСНОВНЫЕ СТАТИСТИКИ И ТАБЛИЦЫ

Оглавление

Вводный обзор 1447
Описательные статистики 1447
Корреляции 1449
t-критерий для независимых выборок 1458
t-критерий для зависимых выборок 1460
Внутригрупповые описательные статистики и корреляции - группировка 1462
Таблицы частот 1465
Таблицы сопряженности и таблицы флагов и заголовков (обзор) 1466
Статистики таблиц сопряженности 1470
Многомерные отклики и дихотомии 1474
Другие критерии значимости 1477
Обзор программы 1479
Примеры 1483
Пример 1: Пошаговое использование системы *STATISTICA* 1483
Пример 2: Группировка и однофакторный дисперсионный анализ 1494
Пример 3: Таблицы частот 1498
Пример 4: Таблицы флагов и заголовков 1502
Пример 5: Таблицы сопряженности 1506
Пример 6: Табулирование многомерных откликов и дихотомий 1509
Диалоговые окна, опции, статистики 1521
Описательные статистики 1522
Корреляционные матрицы 1533
t-критерий для независимых выборок 1537
t-критерий для независимых выборок – сравнение столбцов (переменных) 1539
t-критерий для зависимых выборок 1541
Внутригрупповые описательные статистики и корреляции (группировка) 1541
Внутригрупповые описательные статистики и корреляции - Результаты 1542
Внутригрупповые описательные статистики и корреляции (список таблиц) 1548
Таблицы частот 1553
Таблицы сопряженности и таблицы флагов и заголовков 1558
Задание многомерных откликов/дихотомий 1563
Другие критерии значимости 1567
Предметный указатель 1569
Подробное оглавление см. на следующей странице.

Copyright © StatSoft, 1995

StatSoft®

Подробное оглавление

ВВОДНЫЙ ОБЗОР 1447
Описательные статистики 1447
Обзор 1447
Корреляции 1449
Обзор 1449
t-критерий для независимых выборок 1458
Обзор 1458
t-критерий для зависимых выборок 1460
Обзор 1460
Внутригрупповые описательные статистики и корреляции - группировка 1462
Обзор 1462
Таблицы частот 1465
Обзор 1465
Таблицы сопряженности и таблицы флагов и заголовков (обзор) 1466
Описание совместных распределений частот с помощью кросстабуляций 1466
Таблицы флагов и заголовков 1469
Статистики таблиц сопряженности 1470
Обзор 1470
Статистики, основанные на рангах 1472
Многомерные отклики и дихотомии 1474
Обзор 1474
Многомерные отклики 1474
Многомерные дихотомии 1475
Кросстабуляция многомерных откликов и дихотомий 1476
Парная кросстабуляция переменных с многомерными откликами 1476
Заключительный комментарий 1477
Другие критерии значимости 1477
Критерии значимости 1477

ОБЗОР ПРОГРАММЫ 1479
Описательные статистики 1479
Корреляции 1480
t-критерий для независимых выборок 1480
t-критерий для зависимых выборок 1481
Группировка: внутригрупповые описательные статистики и корреляции 1481
Таблицы частот 1481
Таблицы сопряженности и таблицы флагов и заголовков 1481
Многомерные отклики и дихотомии 1481
Другие критерии значимости 1482
Альтернативные процедуры 1482

StatSoft®

Copyright © StatSoft, 1995

ПРИМЕРЫ 1483
Пример 1: Пошаговое использование системы *STATISTICA* 1483
Инсталляция и запуск программы 1483
Корреляции 1485
Разность между средними (*t*-критерий) 1489
Анализ распределения (описательные статистики) 1492
Пример 2: Группировка и однофакторный дисперсионный анализ 1494
Обзор 1494
Итоговая таблица средних 1495
Однофакторный дисперсионный анализ и *апостериорные* сравнения средних 1496
Внутригрупповые корреляции 1497
Пример 3: Таблицы частот 1498
Обзор 1498
Таблицы частот 1499
Пример 4: Таблицы флагов и заголовков 1502
Обзор 1502
Описание анализа 1502
Чтение таблицы 1503
Статистики 1504
Итог 1506
Пример 5: Таблицы сопряженности 1506
Обзор 1506
Задание анализа 1506
Итог 1509
Пример 6: Табулирование многомерных откликов и дихотомий 1509
Описание файла данных 1509
Задание простых группирующих переменных и переменных с многомерными откликами 1511
Просмотр таблиц частот для факторов с многомерными откликами 1514
Задание многомерной дихотомии 1515
Просмотр таблицы частот для многомерной дихотомии 1516
Кросстабуляция многомерных откликов и дихотомий 1517
Относительные частоты: проценты по числу ответов в сравнении с процентами по числу респондентов 1518
Просмотр трехвходовой таблицы 1519
Заключительное замечание 1520

ДИАЛОГОВЫЕ ОКНА, ОПЦИИ, СТАТИСТИКИ 1521
Стартовая панель 1521
Описательные статистики 1521
Корреляционные матрицы 1521
t-критерий для независимых выборок 1521
t-критерий для зависимых выборок 1521
Группировка: внутригрупповые описательные статистики и внутригрупповые корреляции 1522
Таблицы частот 1522
Таблицы и заголовки 1522

Copyright © StatSoft, 1995

Другие критерии значимости ... 1522
Описательные статистики ... 1522
Переменные ... 1522
Подробные описательные статистики ... 1523
Опции ... 1523
Статистики ... 1523
Распределение ... 1524
Группировка ... 1525
Диаграмма размаха для всех переменных ... 1525
Нормальные вероятностные графики ... 1526
Полунормальные вероятностные графики ... 1527
Нормальные вероятностные графики без тренда ... 1527
2M диаграмма рассеяния (с именами) ... 1527
Матричный ... 1528
3M диаграмма рассеяния (с именами) ... 1528
Поверхность ... 1528
Категоризованные диаграммы размаха ... 1529
Категоризованные графики (взаимодействий) средних ... 1529
Категоризованные гистограммы ... 1530
Категоризованные нормальные графики ... 1530
Категоризованные диаграммы рассеяния ... 1531
3M гистограммы ... 1531
Диалоговое окно: Статистики ... 1531
По умолчанию ... 1532
Все ... 1532
Число наблюдений n ... 1532
Среднее ... 1532
Сумма ... 1532
Медиана ... 1532
Стандартное отклонение ... 1532
Дисперсия ... 1532
Стандартная ошибка среднего ... 1532
95% доверительные границы для среднего ... 1532
Минимум и максимум ... 1532
Нижняя и верхняя квартили ... 1532
Размах ... 1533
Квартильный размах ... 1533
Асимметрия ... 1533
Эксцесс ... 1533
Стандартная ошибка асимметрии ... 1533
Стандартная ошибка эксцесса ... 1533
Другие статистики ... 1533
Корреляционные матрицы ... 1533
Квадратная матрица (один список переменных) ... 1533
Прямоугольная матрица (два списка) ... 1533

StatSoft®

Copyright © StatSoft, 1995

Корреляции ... 1533
Сохранить матрицу ... 1534
Отображать ... 1534
Другие опции ... 1534
2М диаграмма рассеяния (с именами) ... 1535
Матричный ... 1535
Категоризованные диаграммы рассеяния ... 1535
3М диаграмма рассеяния (с именами) ... 1536
Поверхность ... 1536
3М гистограммы ... 1536
t-критерий для независимых выборок ... 1537
Файл данных ... 1537
Переменные ... 1537
Код для группы 1; код для группы 2 ... 1537
T-критерии ... 1537
Опции ... 1538
Диаграммы размаха ... 1538
Категоризованные гистограммы ... 1538
Категоризованные нормальные вероятностные графики ... 1539
Категоризованные нормальные вероятностные графики без тренда ... 1539
Категоризованные диаграммы рассеяния ... 1539
t-критерий для независимых выборок – сравнение столбцов (переменных) ... 1539
Файл данных ... 1540
Переменные (группы) ... 1540
T-критерии ... 1540
Опции ... 1540
Диаграммы размаха ... 1541
t-критерий для зависимых выборок ... 1541
Переменные ... 1541
T-критерии ... 1541
Диаграммы размаха ... 1541
Опции ... 1541
Внутригрупповые описательные статистики и корреляции (группировка) ... 1541
Анализ ... 1542
Переменные ... 1542
Коды для группирующих переменных ... 1542
Внутригрупповые описательные статистики и корреляции - Результаты ... 1542
Итоговая таблица средних ... 1543
Статистики ... 1543
Подробные двувходовые таблицы ... 1543
Маргинальные средние ... 1544
Дисперсионный анализ ... 1544
Апостериорные сравнения средних ... 1544
Внутригрупповые корреляции ... 1545
Переупорядочить факторы в таблице ... 1545

Copyright © StatSoft, 1995

Категоризованные диаграммы размаха 1546
Категоризованные гистограммы 1546
Категоризованные нормальные вероятностные графики 1546
Графики взаимодействий 1547
Категоризованные диаграммы рассеяния 1548
Графики средних и стандартных отклонений 1548
Внутригрупповые описательные статистики и корреляции (список таблиц) 1548
Анализ 1549
Группирующие переменные 1549
Зависимые переменные 1549
Таблицы результатов 1549
Итоговая таблица средних 1549
Подробные двувходовые таблицы 1549
Внутригрупповые корреляции 1549
Дисперсионный анализ 1550
Дополнительные статистики 1550
Опции вычисления внутригрупповых корреляций 1550
Вычислить и показать (*p*)-уровни значимости 1550
Вычислить и показать ковариации 1550
Опции отображения для попарного удаления пропущенных данных (ПД) 1550
Показать число пар N 1550
Показать попарные средние 1551
Показать попарные стандартные отклонения 1551
Апостериорные сравнения средних 1551
Критерий наименьшей значимой разности (НЗР) 1551
Критерий Шеффе 1552
Критерий Ньюмена-Кеулса и критические размахи 1552
Критерий Дункана 1552
Критерий Тьюки достоверно значимой разности (ДЗР) 1552
Критерий Тьюки достоверно значимой разности для выборок неравного размера (неравных N) 1552
Уровень ошибки I рода *альфа* для выделения 1552
Таблицы частот 1553
Переменные 1553
Таблицы частот 1553
Гистограммы 1553
Описательные статистики 1553
Метод категоризации для таблиц и графиков 1554
Опции отображения 1555
Пропущенные данные 1555
Критерии нормальности 1555
Диаграмма размаха для всех переменных 1556
Нормальные вероятностные графики 1556
Полунормальные вероятностные графики 1556
Нормальные вероятностные графики без тренда 1556
3М гистограммы 1556

StatSoft®

Copyright © StatSoft, 1995

Определенные пользователем категории ... 1557
Пример ... 1557
Категория ... 1557
Включить/исключить если ... 1557
Синтаксис условий выбора наблюдений ... 1557
Открыть/Сохранить ... 1558
Просмотреть переменные ... 1558
Открыть все ... 1558
Сохранить все ... 1558
Таблицы сопряженности и таблицы флагов и заголовков ... 1558
Анализ ... 1558
Многовходовые таблицы сопряженности ... 1559
Таблицы флагов и заголовков ... 1559
Использовать только целые коды в выбранных переменных ... 1559
Использовать все выбранные коды ... 1559
Результаты кросстабуляции ... 1559
Просмотреть итоговые таблицы ... 1560
Подробные двувходовые таблицы ... 1560
Таблица флагов и заголовков ... 1561
Отображать длинные метки значений ... 1561
Включить пропущенные данные ... 1561
Отображать выбранные % в отдельных таблицах ... 1561
Статистики для двувходовых таблиц ... 1561
Таблицы ... 1562
Категоризованные гистограммы ... 1562
Графики взаимодействий частот ... 1563
3М гистограммы ... 1563
Задание многомерных откликов/дихотомий ... 1563
Анализ ... 1563
Задать таблицы ... 1564
Парная кросстабуляция ... 1564
Имя многомерного фактора ... 1564
Тип многомерного фактора ... 1564
Длинные метки многомерных факторов ... 1565
Удалить пропущенные данные построчно внутри каждого набора многомерной дихотомии ... 1565
Удалить пропущенные данные построчно внутри каждого набора многомерных откликов ... 1565
Включить пропущенные данные как дополнительную категорию для каждого фактора ... 1565
Просмотр таблиц с многомерными откликами и дихотомиями ... 1565
Просмотреть итоговые таблицы ... 1565
Таблицы частот ... 1566
Подробные двувходовые таблицы ... 1566
Отображать длинные метки значений ... 1566
Отображать выбранные % в отдельной таблице ... 1566
Графики взаимодействий частот ... 1566
Таблицы сопряженности ... 1567

Copyright © StatSoft, 1995

Итоги задания таблицы 1567
Другие критерии значимости 1567
Печатать результаты после каждого вычисления 1568
Различие между двумя коэффициентами корреляции 1568
Различие между двумя средними (нормальное распределение) 1568
Различие между двумя пропорциями 1568
Выделение значимых результатов 1568
ПРЕДМЕТНЫЙ УКАЗАТЕЛЬ 1569

StatSoft®

Copyright © StatSoft, 1995

Глава 9:

ОСНОВНЫЕ СТАТИСТИКИ И ТАБЛИЦЫ

ВВОДНЫЙ ОБЗОР

Статистики, включенные в этот модуль, условно называются *основными статистиками* и рассматриваются в одной группе, т.к. обычно они используются совместно, особенно на начальной, разведочной стадии анализа. Фактически эти статистики являются тестами, которые полезны для самых разнообразных исследований. Каждая основная статистика этого модуля будет подробно рассмотрена в данном введении (предполагается, что читатель освоился с идеями главы *Элементарные понятия статистики,* глава 8). Дальнейшие сведения можно найти в разделе *Примеры* (стр. 1483) данного руководства, а также в учебниках по статистике. Рекомендуем следующие учебники: Kachigan (1986), Runyon and Haber (1976); для более углубленного понимания элементарной теории основных статистик, см. классические книги Hays (1988), Kendall and Stuart (1979). На русском языке см., например, книгу Ивченко Г.И., Медведев Ю.И. “Математическая статистика”, Высшая Школа, 1992.

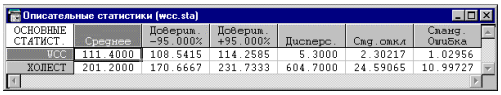
Описательные статистики

Обзор

Цель (упорядочивание данных). Описательные статистики вычисляются отдельно для каждой переменной в файле данных и обеспечивают исследователя такой основной информацией, как среднее, минимальное и максимальное значения, различные меры изменчивости, а также характеристики формы распределения. Например, в приводимых ниже данных содержатся значения двух переменных: число лейкоцитов в крови (переменная *WCC*) и уровень холестерина (переменная *ХОЛЕСТ*), измеренные для каждого из пяти испытуемых (*наблюдение 1* - *наблюдение 5*). Для каждой из этих переменных (т.е. для каждого “столбца” данных) ниже вычислены три типа описательных статистик (среднее, минимум и максимум).

	WCC	ХОЛЕСТ
наблюдение 1	111	220
наблюдение 2	112	201
наблюдение 3	110	185
наблюдение 4	115	170
наблюдение 5	109	230
среднее	111.4	201.2
минимум	109.0	170.0
максимум	115.0	230.0

Эти и другие статистики рассмотрены далее (см. также диалог *Статистики*, стр. 1531, где приведен полный список доступных статистик). Для вычисления описательных статистик можно выбрать все переменные в текущем файле данных. Далее *STATISTICA* построит таблицу описательных статистик, расположив их построчно для каждой переменной.



Меры изменчивости включают дисперсию, стандартное отклонение и стандартную ошибку. Кроме того, доступны различные критерии нормальности (см. диалог *Статистики*, стр. 1522, где дано полное описание имеющихся процедур).

“Истинное” среднее и доверительный интервал. Вероятно, большинство из вас использовало такую описательную статистику, как среднее. Среднее - очень информативная мера “центрального положения ” наблюдаемой

Copyright © StatSoft, 1995

StatSoft®

переменной, если при этом сообщается ее доверительный интервал. Как отмечалось ранее, вам нужны такие статистики (в частности, среднее), которые позволяют сделать вывод относительно популяции в целом. *Доверительный интервал* для среднего представляет собой интервал значений вокруг оценки, где, как вы полагаете, с данным уровнем доверия (см. также *Элементарные понятия статистики,* глава 8), находится “истинное” (неизвестное) среднее совокупности. Границы доверительного интервала – случайные величины, среднее популяции (а именно его мы и оцениваем) – неизвестная величина. Поэтому можно сказать, что доверительный интервал с вероятностью *p* “накрывает” неизвестное значение среднего.

В модуле *Основные статистики и таблицы* можно построить доверительный интервал с любым *p*-уровнем значимости. Например, если среднее выборки равно 23, а нижняя и верхняя границы доверительного интервала с уровнем *p*=.95 равны 19 и 27 соответственно, то можно заключить, что с вероятностью 95% интервал с границами 19 и 27 накрывает среднее популяции. Если вы установите больший уровень доверия, то интервал станет шире, поэтому возрастает вероятность, с которой он накрывает неизвестное среднее популяции, и наоборот; например, хорошо известно, что чем "неопределенней" прогноз погоды (т.е. шире доверительный интервал), тем скорее он будет верным.

Заметим, что ширина доверительного интервала зависит от объема (размера) выборки и разброса данных. Увеличение размера выборки делает оценку среднего более надежной. Увеличение разброса наблюдаемых значений уменьшает надежность оценки (см. *Элементарные понятия статистики*, глава 8). Вычисление доверительных интервалов основывается на предположении нормальности наблюдаемых величин. Если это предположение не выполняется, то оценка может оказаться плохой, особенно для малых выборок. При увеличении объема выборки, скажем, до 100 или более, качество оценки улучшается и без предположения нормальности.

Форма распределения; нормальность. Важным способом описания переменной является форма ее распределения, которая показывает, с какой частотой значения переменной попадают в определенные интервалы. Обычно исследователь интересуется, насколько точно распределение выборки может быть аппроксимировано нормальным (см. также *Элементарные понятия статистики*, глава 8). Простые описательные статистики дают некоторую информацию об этом. Например, если *асимметрия* (показывающая отклонение распределения от симметричного) существенно отличается от 0, то распределение несимметрично, в то время как нормальное распределение абсолютно симметрично. Итак, у симметричного распределения асимметрия равна 0. Асимметрия распределения с длинным правым хвостом положительна. При этом если распределение имеет длинный левый хвост, то его асимметрия отрицательна. Далее, если *эксцесс* (показывающий “пикообразность” распределения) существенно отличен от 0, то распределение или более закруглено в пике, чем нормальное, или, напротив, имеет более острый пик (возможно, имеется несколько пиков). Обычно, если эксцесс положителен, то пик острый, если отрицательный, то пик закругленный. Эксцесс нормального распределения равен 0.

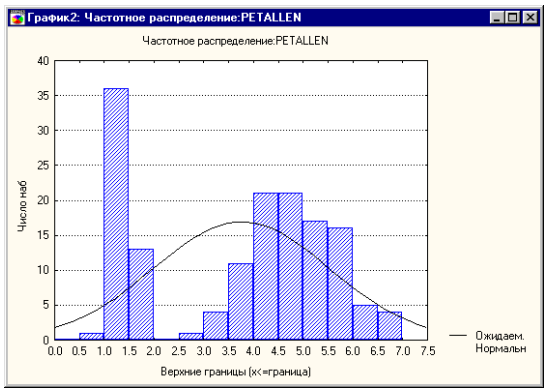
Более точную информацию о форме распределения можно получить с помощью *критериев нормальности* (например, критерия Колмогорова – Смирнова или *W* критерия Шапиро – Уилка). Однако ни один из этих критериев не может заменить визуальную проверку с помощью гистограммы (графика, показывающего частоту попаданий значений переменной в отдельные интервалы).

StatSoft®

Copyright © StatSoft, 1995

В каждом модуле системы *STATISTICA* гистограмму можно построить непосредственно из таблицы результатов, содержащей описательные статистики переменной. Сделайте это, щелкнув правой кнопкой мыши и выбрав в контекстном меню *Быстрые статистические графики* опцию *Гистограмма/нормальн.*

Построенный график позволяет “на глаз” оценить нормальность эмпирического распределения. На гистограмму также накладывается кривая нормального распределения. Гистограмма позволяет *качественно* оценить различные характеристики распределения.



Например, на графике видно, что распределение бимодально (имеет 2 пика). Это может быть вызвано, например, тем, что выборка неоднородна, возможно, извлечена из двух разных популяций, каждая из которых более или менее нормальна. В таких ситуациях, чтобы понять природу наблюдаемых переменных, можно попытаться найти качественный способ разделения выборки на две части.

Другие описательные статистики. Модуль *Основные статистики и таблицы* позволяет выбрать наиболее часто используемые описательные статистики. В дополнение к обычным описательным статистикам, некоторые реже используемые статистики имеются в модуле *Непараметрическая статистика* (например, гармоническое среднее и др.). В этом модуле доступны отличные от нормального распределения, которые можно подогнать к данным. Более специализированные описательные статистики включены в другие модули *STATISTICA*.

Быстрые основные статистики. Заметим, что основные описательные статистики (в том числе для группированных данных) доступны в любом модуле системы *STATISTICA* с помощью процедуры *Быстрые основные статистики* из меню *Анализ* или из панели инструментов *Быстрые основные статистики* (кнопка). Эти возможности подробно описаны в главе 10.

Статистики для выделенного блока данных. Вы можете вычислить итоговые описательные статистики (включая квартили), а также построить графики для выделенного блока данных в любой таблице с помощью процедуры *Статистики блока* (*по строкам* или *по столбцам*), доступной из пункта *Правка* панели инструментов или контекстного меню, вызываемого щелчком правой кнопки мыши. Сначала выделите блок данных, а затем выберите *Статистики блока по строкам* или *Статистики блока по столбцам*, чтобы, соответственно, вычислить описательные статистики для строк или столбцов.

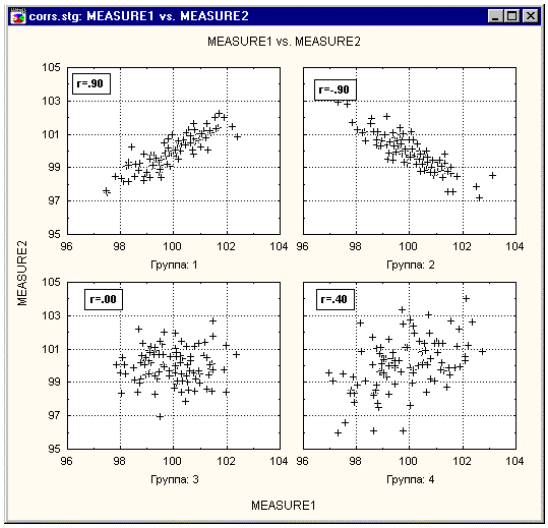
Корреляции

Обзор

Назначение (определение корреляции). Корреляция представляет собой меру зависимости между двумя или более переменными. Используемые шкалы измерений должны быть, как минимум, интервальными (см. *Элементарные понятия статистики*, глава 8). Некоторые другие, непирсоновские, коэффициенты корреляции можно вычислить для менее информативных шкал (см. например, модуль *Непараметрическая статистика*, глава

Copyright © StatSoft, 1995

11). Коэффициенты корреляции изменяются в пределах от –1.00 до +1.00. Значение корреляции –1.00 означает, что переменные связаны линейной зависимостью с отрицательным коэффициентом наклона. Значение +1.00 означает, что переменные связаны линейной зависимостью с положительным наклоном. Значение 0.00 означает отсутствие корреляции.



Наиболее широко используется *Коэффициент корреляции Пирсона r*, называемый также *линейной* корреляцией, т.к. он измеряет степень *линейных* связей. Именно этот тип корреляции доступен в модуле *Основные статистики и таблицы*. Другие виды корреляции будут описаны позже.

Упорядочивание данных. Корреляции можно вычислить между любыми заданными переменными. Например, ниже приведены данные о количестве лейкоцитов (переменная *WCC*) и уровне холестерина (переменная *ХОЛЕСТ*) для каждого из 5 человек (*наблюдение 1 - наблюдение 5*). Корреляция между этими двумя переменными (т.е. столбцами) приводится ниже.

	WCC	ХОЛЕСТ
наблюдение 1	111	220
наблюдение 2	112	201
наблюдение 3	110	185
наблюдение 4	115	170
наблюдение 5	109	230

r = -.73

Если выбранные переменные содержат номинальные данные (см. *Элементарные понятия статистики*, глава 8), то необходимо использовать специальные типы корреляций, имеющиеся, например, в процедуре *Таблицы сопряженности* модуля *Основные статистики и таблицы*, стр. 1466).

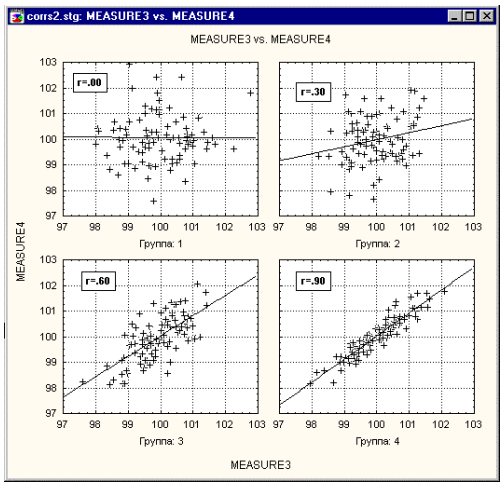
Простая линейная корреляция (Пирсона r). Корреляция Пирсона (далее называемая просто *корреляцией*) предполагает, что две переменные измеряются, по крайней мере, в интервальной шкале (см. *Элементарные понятия статистики*, глава 8), и определяет степень, с какой значения двух переменных пропорциональны друг другу. Важно, что значение коэффициента корреляции не зависит от масштаба измерения. Например, корреляция между ростом и весом будет одной и той же, независимо от того, проводились измерения в *дюймах и футах* или в *сантиметрах и килограммах*. Пропорциональность означает просто линейную зависимость. Корреляция высокая, если на графике зависимость можно представить прямой линией (с положительным или отрицательным угловым коэффициентом).

Эта прямая называется *прямой регрессии* или *прямой наименьших квадратов*. Такое название связано с тем, что сумма квадратов расстояний от наблюдаемых точек до прямой является минимальной из всех возможных. Заметим, что использование квадратов расстояний имеет важные функциональные последствия относительно того, как значения коэффициентов корреляции реагируют на определенное расположение данных (например, на выбросы, что вы увидите позднее).

StatSoft®

Copyright © StatSoft, 1995

Прямая регрессии. Угловой коэффициент прямой регрессии показывает степень зависимости следующим образом. Если масштаб осей одинаков, а корреляция сильно отрицательна или сильно положительна, то наклон прямой составляет около 45° в соответствующем направлении. Если корреляция отсутствует (равна 0), то прямая - горизонтальна, как показано на приведенной ниже диаграмме рассеяния.



Как интерпретировать значения корреляций. Как отмечалось ранее, коэффициент корреляции Пирсона (*r*) представляет собой меру линейной зависимости двух переменных. Если возвести его в квадрат, то полученное значение (коэффициент детерминации r^2) будет представлять долю вариации, общую для двух переменных (иными словами, "степень" зависимости двух переменных). Для того чтобы оценить зависимость между переменными, важно знать как "величину" корреляции, так и ее значимость (см. *Элементарные понятия статистики*, глава 8).

Значимость корреляций. Уровень значимости, вычисленный для каждой корреляции, представляет собой главный источник информации о надежности корреляции. Чтобы легче понять, какие коэффициенты являются значимыми на некотором подходящем уровне, в диалоговом окне *Корреляции Пирсона* модуля *Основные статистики и таблицы* (стр. 1533) можно автоматически выделять значимые коэффициенты корреляции цветом (например, красным).

Как объяснялось выше (см. *Элементарные понятия статистики*, глава 8), значимость определенного коэффициента корреляции будет меняться в зависимости от объема выборок. Критерий значимости основывается на предположении, что распределение остатков (т.е. отклонений наблюдаемых значений от регрессионной прямой) для зависимой переменной *y* является нормальным, с постоянной дисперсией (одинаковой для всех значений независимой переменной *x*).

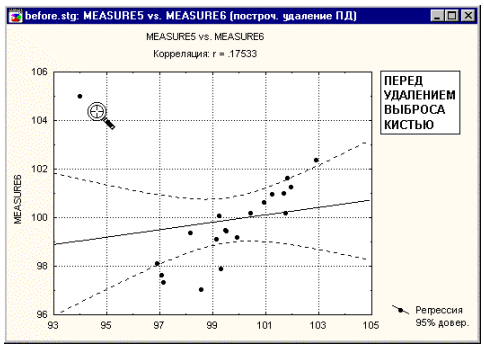
Исследования методом Монте-Карло показали, что нарушение этих условий не абсолютно критично, если размеры выборок не слишком малы, а отклонения от нормальности небольшие. Тем не менее, имеется несколько серьезных опасностей, о которых следует знать.

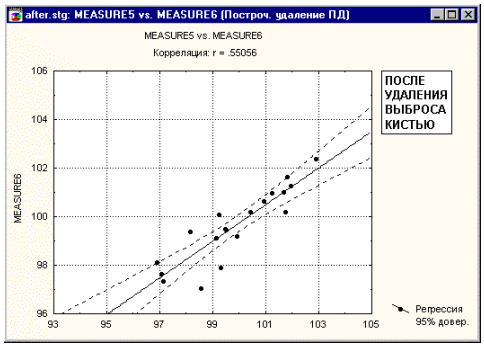
Выбросы. По определению, выбросы являются нетипичными, резко выделяющимися наблюдениями. Так как в определении прямой регрессии используется сумма *квадратов расстояний* наблюдаемых точек до прямой, то выбросы могут существенно повлиять на наклон прямой и, следовательно, на значение коэффициента корреляции. Поэтому единичный выброс (значение которого возводится в квадрат) способен существенно изменить наклон прямой и, следовательно, значение корреляции.

Обычно считается, что выбросы представляют собой случайную ошибку, которую хотелось бы контролировать. К сожалению, не существует общепринятого метода автоматического удаления выбросов (тем не менее, см. следующий параграф). Таким образом, чтобы не быть введенными в заблуждение полученными

Copyright © StatSoft, 1995

значениями, необходимо проверить на *диаграмме рассеяния* каждый важный случай значимой корреляции. Заметим, что графические процедуры в диалоговом окне *Корреляции Пирсона* (стр. 1533), а также средства графического редактирования предоставляют разнообразные способы экспериментирования с данными для интерактивного удаления выбросов (с помощью инструмента "кисть"), позволяя отчетливо увидеть, каким образом выбросы влияют на регрессионную прямую, см. ниже.



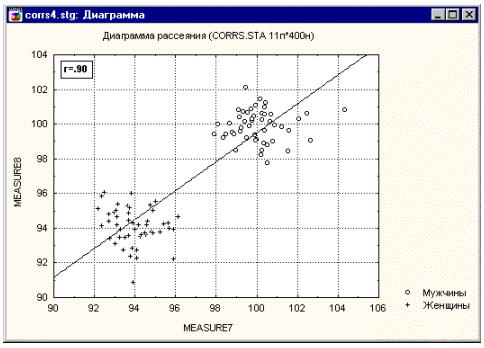


Очевидно, выбросы могут не только искусственно увеличить значение коэффициента корреляции, но также уменьшить реально существующую корреляцию (как показано на графике выше).

Количественный подход к выбросам. Некоторые исследователи применяют численные методы удаления выбросов. Например, исключаются значения, которые выходят за границы ±2 стандартных отклонений (и даже ±1.5 стандартных отклонений) вокруг выборочного среднего. В ряде случаев такая "чистка" данных абсолютно необходима. Например, при изучении реакции в когнитивной психологии, даже если почти все значения экспериментальных данных лежат в диапазоне 300-700 *миллисекунд*, то несколько "странных времен реакции" 10-15 *секунд* совершенно меняют общую картину.

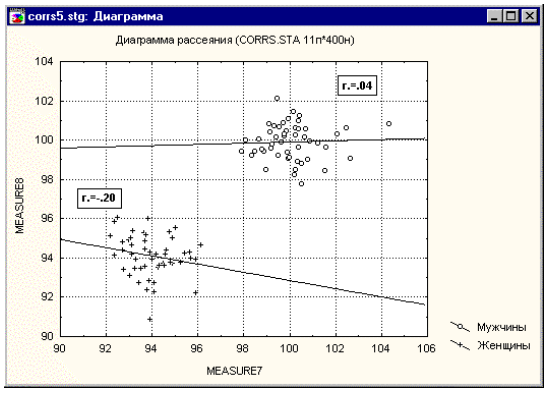
К сожалению, определение выбросов субъективно, и решение должно приниматься индивидуально в каждом эксперименте (с учетом особенностей эксперимента и/или "сложившейся практики" в данной области). Следует заметить, что в некоторых случаях относительная частота выбросов к численности групп может быть исследована и разумно проинтерпретирована с точки зрения самой организации эксперимента.

Корреляции в неоднородных группах. Отсутствие однородности в выборке также является фактором, смещающим (в ту или иную сторону) оценку корреляции. Представьте ситуацию, когда коэффициент корреляции вычислен по данным, которые поступили из двух различных экспериментальных групп, что было проигнорировано при вычислениях. Далее, пусть действия экспериментатора в одной из групп увеличивают значения обеих коррелированных величин и таким образом данные каждой группы сильно различаются на диаграмме рассеяния.



StatSoft®
Copyright © StatSoft, 1995

В подобных ситуациях высокая корреляция может быть следствием разбиения данных на две группы, а вовсе не отражать “истинную” зависимость между двумя переменными, которая может практически отсутствовать (как видно из графиков, построенных для каждой группы).



Если вы допускаете такое явление и знаете, как определить “подмножества” данных, попытайтесь вычислить корреляции отдельно для каждого множества.

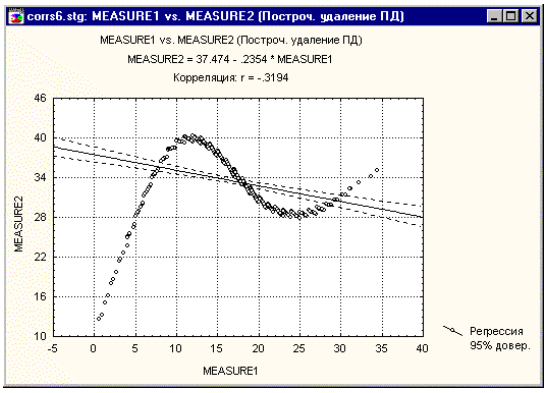
Можно воспользоваться, например, процедурой *Группировка* (см. стр. 1462) в модуле *Быстрые основные статистики* (см. главу 10) или процедурой *Категоризованные диаграммы рассеяния* (стр. 1531). Можно также использовать многочисленные гибкие графические возможности, имеющиеся в меню *Статистические графики* (см. выпадающее меню *Графика*) для того, чтобы проверить корреляции методами РАД (разведочного анализа данных). Первый из двух показанных выше графиков построен с помощью процедуры *Статистические 2М графики - Диаграммы рассеяния*. Этот метод позволяет определить подмножества визуально, при этом все точки рассматриваются как принадлежащие к одному множеству, и подгонка функции производится по всему объему данных. Второй график построен с помощью процедуры *Статистические категоризованные графики - Диаграммы рассеяния*. Этот метод позволяет сгруппировать данные с помощью некоторого заранее определенного критерия. В данном случае, в соответствии с полом испытуемых; функции подогнаны к точкам в заранее определенных множествах, используя параметр *Вместе*.

Если вам неясно, как определить подмножества, попытайтесь применить многомерные методы разведочного анализа, имеющиеся в системе *STATISTICA* (например, *Кластерный анализ*, см. том III).

Нелинейные зависимости между переменными. Другим возможным источником трудностей, связанным с линейной корреляцией *Пирсона r*, является форма зависимости. Корреляция *Пирсона r* хорошо подходит для описания линейной зависимости. Отклонения от линейности увеличивают общую сумму квадратов расстояний от регрессионной прямой, даже если она представляет “истинные” и очень тесные зависимости между переменными.

Еще одной причиной, вызывающей необходимость рассмотрения диаграммы рассеяния для каждого коэффициента корреляции, является нелинейность.

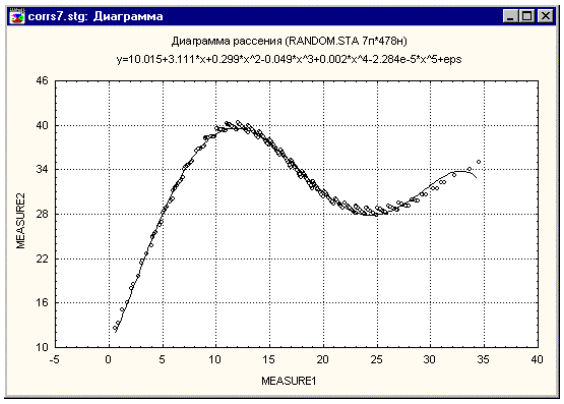
Например, показанная ниже высокая корреляция плохо описывается линейной функцией.



Copyright © StatSoft, 1995

StatSoft®

Однако, как видно на графике ниже, полином пятого порядка достаточно хорошо описывает данные.



Измерение нелинейных зависимостей. Что делать, если корреляция сильная, однако зависимость явно нелинейная (как видно из графика)? К сожалению, не существует простого ответа на данный вопрос, так как не имеется естественного обобщения *коэффициента корреляции Пирсона r* на нелинейный случай. Если кривая монотонна (возрастает или убывает), то можно преобразовать одну или обе переменные, чтобы сделать зависимость линейной, а затем уже вычислить корреляцию между преобразованными величинами.

Часто используется логарифмическое преобразование (для того чтобы проверить возможности данного преобразования, переключитесь с линейной шкалы на логарифмическую в вашей диаграмме рассеяния, используя диалоговое окно *Общая разметка*).

Другая подход состоит в использовании непараметрической корреляции (например, корреляции Спирмена, см. *Непараметрическая статистика и подгонка распределения*, глава 11). Иногда этот метод приводит к успеху, хотя непараметрические корреляции чувствительны только к упорядоченным значениям переменных, например, по определению, она пренебрегает монотонными преобразованиями данных. К сожалению, два самых точных метода исследования нелинейных зависимостей непросты в использовании и требуют хорошего навыка “экспериментирования” с данными. Эти методы состоят в следующем:

(a) нужно попытаться найти функцию, которая наилучшим способом описывает данные. (Удобное для этого средство имеется в диалоговом окне *Задание функции пользователя,* вызываемом с помощью кнопки *Пользовательская* окна *Размещение графика*, которое в свою очередь вызывается щелчком правой кнопки мыши по фону графика и выбором затем опции *Изменить размещение графика* из контекстного меню). После того, как вы определили функцию, можно проверить степень ее согласия с данными при помощи процедур модуля *Нелинейное оценивание* (том III);

(b) вы можете иметь дело с данными, разбитыми некоторой переменной на группы (например, 4 или 5). Определите эту новую переменную как группирующую, а затем используйте дисперсионный анализ *STATISTICA.*

Разведочный анализ корреляционных матриц. Во многих случаях первый шаг анализа данных состоит в вычислении корреляционной матрицы всех переменных и проверке значимых (ожидаемых и неожиданных) зависимостей. После того как это сделано, следует понять общую природу обнаруженной статистической значимости (см. *Элементарные понятия статистики*, глава 8). Иными словами, понять, почему определенные коэффициенты корреляции значимы. Однако следует иметь в виду, если используется несколько критериев, значимые результаты могут появляться “удивительно часто” чисто случайным образом.

Например, коэффициент, значимый на уровне .05, будет встречаться чисто случайно один раз в каждом из 20 подвергнутых исследованию

StatSoft®

Copyright © StatSoft, 1995

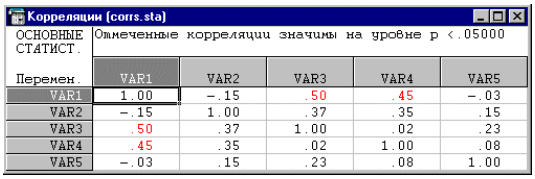
коэффициентов. Нет способа автоматически выделить “истинную” корреляцию. Поэтому следует подходить с осторожностью ко всем, не предсказанным или заранее не запланированным результатам и попытаться соотнести их с другими (надежными) результатами. В конечном счете, самый убедительный способ проверки состоит в проведении повторного экспериментального исследования. Такое положение является общим для всех методов анализа, использующих “множественные сравнения и статистическую значимость”. Эта проблема также обсуждается в описании процедур *Апостериорные сравнения средних* (стр. 1551) и *Группировка* (стр. 1462).

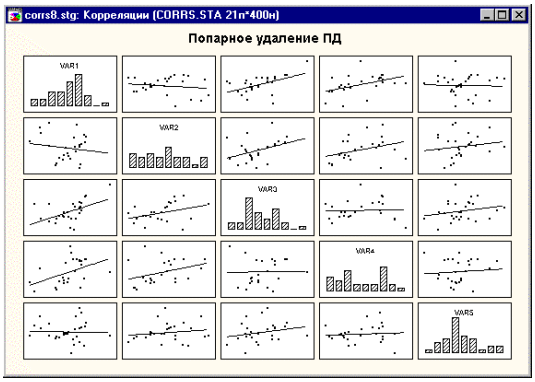
Построчное удаление пропущенных данных в сравнении с попарным удалением. Принятый по умолчанию способ удаления пропущенных данных при вычислении корреляционной матрицы - *построчного удаления* (в нем удаляется вся строка, в которой имеется хотя бы одно пропущенное значение). Этот способ приводит к “правильной” корреляционной матрице в том смысле, что все коэффициенты вычислены по одному и тому же множеству наблюдений. Однако если пропущенные значения распределены случайным образом по переменным, то метод может привести к тому, что в рассматриваемом множестве данных не останется ни одного не исключенного наблюдения (в каждой строке встретится, по крайней мере, одно пропущенное значение).

Чтобы избежать подобной ситуации, используют способ *попарного удаления*. В этом способе учитываются только пропуски в каждой выбранной паре и игнорируются пропуски в других переменных. Корреляция между парой переменных вычисляется по наблюдениям, где нет пропусков. (Очевидно, если в файле содержится лишь две переменные, оба способа совпадают.) Во многих ситуациях, особенно когда количество пропущенных данных относительно мало, скажем 10%, и пропуски распределены достаточно хаотично, этот метод не приводит к серьезным ошибкам. Однако иногда это не так.

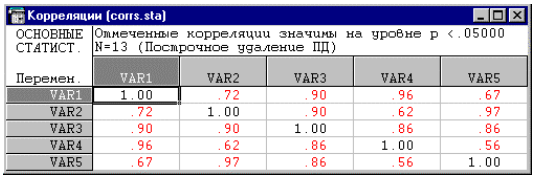
Например, в систематическом смещении (сдвиге) оценки может “скрываться” систематическое расположение пропусков, являющееся причиной различия коэффициентов корреляции, построенных по разным подмножествам (как показано ниже).

Следующая корреляционная матрица вычислена с использованием попарного способа удаления пропущенных данных.

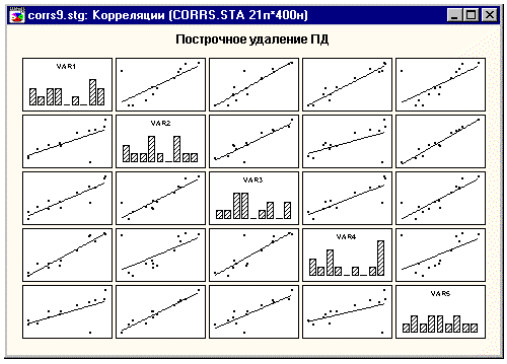




Корреляционная матрица, показанная ниже, вычислена по тем же данным, но с использованием построчного способа удаления пропущенных данных.



Copyright © StatSoft, 1995



Очевидно, эти матрицы различны; вы видите, изменились даже знаки коэффициентов корреляции.

Другая проблема, в дополнении к возможному смещению оценок корреляции, возникает при использовании корреляционной матрицы в других модулях (например, *Множественная регрессия*, *Факторный анализ* или *Кластерный анализ*). В этих модулях предполагается, что используется "правильная" корреляционная матрица с определенным уровнем состоятельности и соответствия различных коэффициентов. Использование матрицы с "плохими" (смещенными) оценками приводит к тому, что программа либо не в состоянии использовать такую матрицу, либо результаты будут ошибочными. Поэтому, если применяется попарный метод исключения пропущенных данных, необходимо проверить имеются ли систематические закономерности в распределении пропусков. Для этого полезна процедура *Диаграммы пропущенных значений и интервалов* (одна из опций меню *Статистические 2М графики*) (см. том II).

Как определить смещения, вызванные попарным удалением пропущенных данных. Процедура *Корреляционные матрицы* строит таблицы результатов с *n* средними и стандартными отклонениями, вычисленными отдельно для каждой пары переменных. Таким образом, оценка каждого коэффициента основана на собственном множестве данных.

Если попарное исключение пропущенных данных не приводит к какому-либо систематическому сдвигу в оценках, то все эти статистики будут похожи на аналогичные статистики, вычисленные при построчном способе удаления пропусков. Если наблюдается значительное различие, то есть основание предполагать наличие сдвига в оценках.

Например, если среднее (или стандартное отклонение) значений переменной A, которое использовалось при вычислении ее корреляции с переменной B, много меньше среднего (или стандартного отклонения) тех же значений переменной A, которые использовались при вычислении ее корреляции с переменной C, то имеются все основания ожидать, что эти две корреляции (A-B и A-C) основаны на разных подмножествах данных, и, таким образом, в оценках корреляций имеется сдвиг, вызванный неслучайным расположением пропусков в значениях переменных.

Попарное удаление пропущенных данных в сравнении с подстановкой среднего значения. Другим общим методом, позволяющим избежать потери наблюдений при построчном удалении, является *замена средним* (замена всех пропущенных значений средним значением соответствующей переменной). Этот метод доступен во многих модулях системы. Вы можете также использовать его (с помощью модуля *Управление данными*, см. глава 7). Подстановка среднего имеет свои преимущества и недостатки по сравнению с попарным способом. Основное преимущество в том, что он дает состоятельные оценки, однако имеет следующие недостатки:

(a) *подстановка среднего* искусственно уменьшает разброс данных, иными словами, чем больше пропущенных данных, тем больше данных, совпадающих со средним значением, искусственно добавленным в данные.

StatSoft®

Copyright © StatSoft, 1995

(b) так как пропущенные данные заменяются искусственно созданными “средними” значениями, то корреляции могут очень сильно уменьшиться.

Ложные корреляции. Хотя, основываясь на коэффициентах корреляции, вы не можете доказать причинной зависимости (см. *Элементарные понятия статистики*, глава 8), однако можете определить *ложные* корреляции, т.е. корреляции, которые связаны с влияниями “других”, остающихся вне вашего поля зрения переменных. Лучше всего понять ложные корреляции на простом примере.

Известно, что существует корреляция между ущербом, причиненным пожаром, и числом пожарных, тушивших пожар. Однако эта корреляция ничего не говорит о том, насколько уменьшатся потери, если будет вызвано меньше число пожарных. Причина в том, что имеется третья переменная (начальный *размер* пожара), которая влияет как на причиненный ущерб, так и на число вызванных пожарных. Если вы будете “контролировать” эту переменную (например, рассматривать только пожары определенной величины), то исходная корреляция (между ущербом и числом пожарных) либо исчезнет, либо, возможно, даже изменит свой знак.

Основная проблема ложной корреляции состоит в том, что вы не знаете, кто является ее агентом. Тем не менее, если вы знаете, где искать, то можно воспользоваться *частными корреляциями,* чтобы контролировать (частично исключенное) влияние определенных переменных. В системе *STATISTICA* процедуры, вычисляющие частные корреляции, включены в модуль *Множественная регрессия* (см. глава 12).

Заметим, что продвинутые методы анализа корреляционных матриц в терминах гипотез об их определенной структуре и причинной зависимости между переменными реализованы в модуле *Моделирование структурными уравнениями (SEPATH)*, том III.

Являются ли коэффициенты корреляции “аддитивными”? Нет, не являются. Например, усредненный коэффициент корреляции, вычисленный по нескольким выборкам, не совпадает “со средней корреляцией” во всех этих выборках. Причина в том, что коэффициент корреляции не является линейной функцией величины зависимости между переменными. Коэффициенты корреляции не могут быть просто усреднены. Если вас интересует средний коэффициент корреляции, следует преобразовать коэффициенты корреляции в такую меру зависимости, которая будет аддитивной.

Например, до того, как усреднить коэффициенты корреляции, их можно возвести в квадрат, получить *коэффициенты детерминации*, которые уже будут аддитивными, или преобразовать корреляции в *z значения Фишера*, которые также аддитивны (что можно сделать с помощью *Z* преобразования, доступного в процедуре *Вероятностный калькулятор*, см. *Быстрые основные статистики*, глава 10).

Как определить, являются ли два коэффициента корреляции значимо различными. Имеется критерий, позволяющий оценить значимость различия между двумя коэффициентами корреляциями (см. *Другие критерии значимости*, стр. 1477). Результат применения критерия зависит не только от величины разности этих коэффициентов, но и от объема выборок и величины самих этих коэффициентов. В соответствии с ранее обсуждаемым принципом, чем больше объем выборки, тем для меньшего эффекта мы можем доказать значимость. Вообще говоря, в соответствии с общим принципом, надежность коэффициента корреляции увеличивается с увеличением его абсолютного значения, относительно малые различия между большими коэффициентами могут быть значимыми. Например, разница .10 между двумя корреляциями может не быть значимой, если коэффициенты равны .15 и .25,

Copyright © StatSoft, 1995

хотя для той же выборки разность 0.10 может оказаться значимой для коэффициентов .80 и .90.

Другие коэффициенты корреляции. Существует другие виды коэффициентов корреляции. Они также включены в модули системы *STATISTICA*. Различные *непараметрические* корреляции включены в модуль *Непараметрическая статистика,* а также вычисляются в процедуре *Таблицы сопряженности*. *Частные корреляции* доступны в модуле *Множественная регрессия*. *Автокорреляции*, т.е. корреляции между значениями самой переменной, сдвинутыми на определенный лаг, включены в модуль *Временные ряды/Прогнозирование* (более специальные меры зависимости между переменными или группами переменных имеются в других модулях системы). Стандартные корреляции Пирсона можно вычислить в различных модулях (например, *Множественная регрессия*, *Факторный анализ*, *Управление данными*). Тетрахорические корреляции можно вычислить в модуле *Надежность/Позиционный Анализ*. Различные метрики доступны в модуле *Кластерный анализ,* меры зависимости между группирующими (номинальными) переменными доступны в процедуре *Таблицы и заголовки* модуля *Основные статистики и таблицы* (см. стр. 1558). Наконец, как уже отмечалось, существуют продвинутые методы анализа корреляционных матриц и их структур (см. *Моделирование структурными уравнениями (SEPATH)*, том III).

t-критерий для независимых выборок

Обзор

Цель, предположения. *t-критерий* является наиболее часто используемым методом, позволяющим выявить различие между средними двух выборок. Например, *t-критерий* можно использовать для сравнения средних показателей в группе пациентов, принимавших определенное лекарство, и в контрольной группе, где принималось безвредное лекарство. Теоретически, *t-критерий* может применяться, даже если размер выборки очень небольшой (например, 10; некоторые исследователи утверждают, что можно исследовать меньшие выборки), и если переменные нормально распределены (внутри групп), а дисперсии наблюдений в группах не слишком различны (см. также *Элементарные понятия статистики*, глава 8).

Предположение о нормальности можно проверить, исследуя распределение (с помощью гистограмм) или применяя критерий нормальности (см. *Описательные статистики).* Равенство дисперсий в двух группах можно проверить с помощью *F* критерия (который включен в таблицу вывода *t*-критерия). Также можно воспользоваться более устойчивым *критерием Левена*. Если условия применимости *t*-критерия не выполнены, можно оценить различие между средними двух групп, используя какую-либо непараметрическую альтернативу *t*-критерия (см. *Непараметрическая статистика и подгонка распределения*, глава 11).

p-уровень значимости *t*-критерия равен вероятности ошибочно принять гипотезу о различии между средними выборок, когда она не верна. Иными словами, вероятности ошибки отвергнуть гипотезу о сходстве средних двух наблюдаемых групп.

Некоторые исследователи предлагают, в случае, когда рассматриваются отличия только в одном направлении (например, переменная X больше (меньше) в первой группе, чем во второй), рассматривать одностороннее *t*-распределение и делить полученный для двустороннего *t-критерия p-уровень* пополам. Другие предлагают всегда работать со стандартным двусторонним *t-критерием.*

StatSoft®

Copyright © StatSoft, 1995

Упорядочивание данных. Чтобы применить *t-критерий* для независимых выборок, требуется, по крайней мере, одна независимая (*группирующая*) переменная (например, *Пол*) и одна зависимая переменная (например, тестовое значение некоторого показателя). Вначале, с помощью специальных значений независимой переменной (эти значения называются *группирующими кодами*), например, *мужчина* и *женщина,* если группирующей переменной является *Пол*, данные разбиваются на две группы. Далее в каждой группе вычисляется среднее значение зависимой переменной. Эти средние сравниваются между собой.

Следующие данные исследуются с помощью *t*-критерия. Мы сравниваем средние значения количества лейкоцитов *WCC* у *мужчин* и *женщин*.

	ПОЛ	WCC
наблюдение 1	мужчина	111
наблюдение 2	мужчина	110
наблюдение 3	мужчина	109
наблюдение 4	женщина	102
наблюдение 5	женщина	104

среднее WCC для мужчин = 110
среднее WCC для женщин = 103

Если вы определите не одну зависимую переменную, а несколько, то программа вычислит последовательность *t*-критериев (один критерий для каждой зависимой переменной).

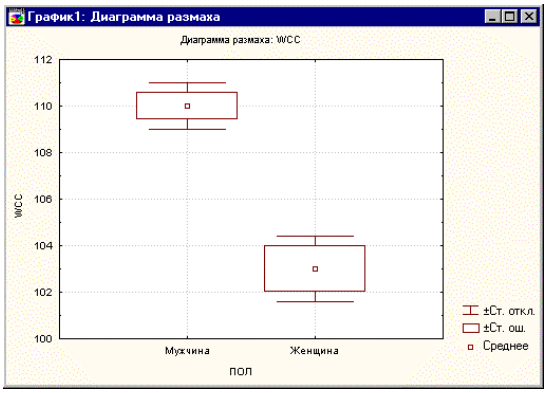
Альтернативная группировка данных. Иногда данные упорядочены так, что каждый столбец или переменная содержат значения, относящиеся к одной группе:

	Муж	Жен
наблюдение 1	111	102
наблюдение 2	110	104
наблюдение 3	109	

Это нестандартное расположение данных, все значения относятся к одной переменной WCC

Процедура *t-критерий для независимых выборок* модуля *Основные статистики и таблицы* может вычислять *t-критерий* и для данных, расположенных таким образом. Однако заметим, что такое расположение нетипично и обычно не рекомендуется, особенно при работе с большими файлами данных. Практически все системы анализа, включая систему *STATISTICA*, предполагают, что данные расположены так, как описывалось в предыдущем абзаце (т.е. имеется отдельная (отдельные) группирующая переменная). Такое расположение позволяет отнести каждого субъекта в файле данных к определенной группе. К данным такого рода удобно применять различные многомерные методы, где требуется вычислять корреляционные матрицы.

Графики *t*-критериев. Анализ данных с помощью *t*-критерия, сравнения средних и меры отклонения от среднего в группах можно производить с помощью *диаграмм размаха*, которые доступны из таблицы результатов простым указанием на выбранный *t-критерий* (используя клавиши управления курсором или мышь).



Различные графики распределения значений зависимой переменной в каждой отдельной группе можно также построить непосредственно из диалога *t-критерий для независимых выборок*. Эти графики позволяют быстро визуально оценить и интуитивно понять степень зависимости между группирующей и зависимой переменными.

Более сложное групповое сравнение. На практике часто приходится сравнивать более двух групп данных (например, имеется *лекарство 1*, *лекарство 2*, и *успокоительное*

Copyright © StatSoft, 1995

лекарство), или сравнивать группы, созданные более чем одной независимой переменной (например, *Пол*, тип *Лекарства* и *Доза*). Тогда следует использовать *Дисперсионный анализ* (см. *ANOVA/MANOVA*, глава 13), который можно рассматривать как обобщение *t*-критерия. Фактически в случае однофакторного сравнения двух групп, дисперсионный анализ дает результаты идентичные *t*-критерию [$t^2(df) = F(1,df)$]. Однако, если план существенно более сложный, ANOVA предпочтительнее *t*-критерия (даже если используется последовательность *t*-критериев, см. главу 13).

Имеются также другие специальные методы группового сравнения в терминах средних, дисперсий, ковариаций. Обратитесь к модулю *Моделирование структурными уравнениями (SEPATH)* (том III).

t-критерий для зависимых выборок

Обзор

Внутригрупповая вариация. Как объяснялось ранее (см. *Элементарные понятия статистики*, глава 8), степень различия между средними в двух группах зависит от *внутригрупповой* вариации (дисперсии) переменных. В зависимости от того, насколько различны эти значения для каждой группы, "грубая разность" между групповыми средними показывает более сильную или более слабую степень зависимости между независимой (*группирующей*) и зависимой переменными.

Например, если среднее *WCC* (число лейкоцитов) равнялось 102 для мужчин и 104 для женщин, то разность только на величину 2 между внутригрупповыми средними будет чрезвычайно важной в том случае, если все значения *WCC* мужчин лежат в интервале от 101 до 103, а все значения *WCC* женщин - в интервале 103 - 105. Тогда можно довольно хорошо предсказать *WCC* (значение зависимой переменной) исходя из пола субъекта (независимой переменной). Однако если та же разность 2 получена из сильно разбросанных данных (например, изменяющихся в пределах от 0 до 200), то ею вполне можно пренебречь. Таким образом, можно сказать, что уменьшение *внутригрупповой вариации* увеличивает чувствительность критерия.

Цель. *t*-критерий для зависимых выборок дает преимущество в том случае, когда важный источник *внутригрупповой вариации* (или *ошибки*) может быть легко определен и исключен из анализа. В частности, это относится к экспериментам, в которых две сравниваемые группы наблюдений основываются на одной и той же выборке наблюдений (субъектов), которые тестировались *дважды* (*до* и *после* лечения) [таким образом, выборки зависимы, отсюда и происходит название критерия]. В таких экспериментах значительная часть внутригрупповой изменчивости (вариации) в обеих группах может быть объяснена индивидуальными различиями субъектов. Заметим, что на самом деле, такая ситуация не слишком отличается от той, когда сравниваемые группы совершенно независимы (см. *t-критерий* для независимых выборок), где индивидуальные отличия также вносят вклад в *дисперсию ошибки*. Однако в случае независимых выборок, вы ничего не сможете поделать с этим, т.к. не сможете определить (или "удалить") часть вариации, связанную с индивидуальными различиями субъектов. Если та же самая выборка тестируется дважды, то можно легко исключить эту часть вариации. Вместо исследования каждой группы отдельно и анализа исходных значений, можно рассматривать просто разности между двумя измерениями (например, "до теста" и "после теста") для каждого субъекта. Вычитая первые значения из вторых (для каждого субъекта) и анализируя затем только эти "чистые (парные) разности", вы исключите ту часть вариации,

StatSoft®

Copyright © StatSoft, 1995

которая является результатом различия в исходных уровнях индивидуумов. Именно так и вычисляется в *t*-критерии для зависимых выборок. В сравнении с *t*-критерием для независимых выборок, такой подход дает всегда "лучший" результат (в частности, критерий становится более чувствительным).

Предположения. Теоретические предположения *t*-критерия для независимых выборок также применимы к критерию зависимых выборок. Это означает, что парные разности должны быть нормально распределены. Если это не выполняется, то можно воспользоваться одним из альтернативных непараметрических критериев (см. модуль *Непараметрическая статистика*).

Упорядочивание данных. Вы можете применять *t*-критерий для зависимых выборок к любой паре переменных в данных. Выбор переменных подобен тому, как при вычислении корреляций (см. *Корреляционные матрицы*, стр. 1533). Однако применение этого критерия мало оправдано, если значения двух переменных несопоставимы. Например, если вы сравниваете среднее *WCC* в выборке пациентов до и после лечения, но используете различные методы вычисления количественного показателя или другие единицы во втором измерении, то высоко значимые значения *t*-критерия могут быть получены искусственно, именно за счет изменения единиц измерения. Следующие данные могли бы быть проанализированы с помощью *t*-критерия для зависимых выборок.

	WCC до	WCC после
наблюдение 1	111.9	113
наблюдение 2	109	110
наблюдение 3	143	144
наблюдение 4	101	102
наблюдение 5	80	80.9
...	...	...

средняя разность между WCC "до" и "после" = 1

Средняя разность между показателями в двух столбцах относительно мала (*d=1*) по сравнению с разбросом данных (от 80 до 143, в первой выборке). Тем не менее, *t-критерий* для зависимых выборок использует только парные разности, "игнорируя" исходные численные значения и их вариацию. Таким образом, величина этой разности *1* будет сравниваться не с разбросом исходных значений, а с разбросом *индивидуальных разностей*, который относительно мал: *0.2* (от *0.9* в наблюдении 5 до *1.1* в наблюдении 1). В этой ситуации разность *1* очень большая и может привести к значимому *t*-значению.

Матрицы *t*-критериев. *t*-критерий для зависимых выборок может быть вычислен для списков переменных и просмотрен далее как матрица. Пропущенные данные при этом обрабатываются либо попарным, либо построчным способом, точно так же как при вычислении корреляционных матриц (см. стр. 1533). Все те предостережения, которые относились к матрицам коэффициентов корреляций, остаются в силе для матриц *t*-критериев: (a) появление артефактов (искусственных результатов) из-за попарного удаления пропусков, (b) возникновение чисто "случайно" значимых результатов.

Более сложные групповые сравнения. Если имеется более двух "зависимых выборок" (например, *до лечения*, *после лечения* и *после лечения -2*), то можно использовать дисперсионный анализ с *повторными измерениями*. Повторные измерения в дисперсионном анализе (ANOVA) можно рассматривать как обобщение *t*-критерия для зависимых выборок, позволяющие увеличить чувствительность анализа.

Например, можно одновременно контролировать не только базовый уровень зависимой переменной, но и другие факторы и включать в план эксперимента более одной зависимой переменной (многомерный дисперсионный анализ MANOVA; более подробно см. *ANOVA/MANOVA*, глава 13).

Copyright © StatSoft, 1995

StatSoft®

В модуле SEPATH реализованы продвинутые методы сравнения зависимых выборок в терминах средних, дисперсий и ковариаций основных структур (см. *Моделирование структурными уравнениями (SEPATH)*, том III).

Внутригрупповые описательные статистики и корреляции - группировка

Обзор

Цель. Процедура разбиения на группы вычисляет описательные статистики и корреляции для *зависимых* переменных в каждой из нескольких групп, определенных одной или большим числом *группирующих* (*независимых*) переменных. Понятие *группирующих* или *кодовых* переменных также объясняется в главе *Управление данными* (см. глава 7).

Группировка данных. В приводимом ниже примере, значения зависимой переменной *WCC* разбиваются на группы кодами двух *независимых* переменных: *ПОЛ* (значения: *мужчины* и *женщины*), и *РОСТ* (значения: *высокий* и *низкий*).

	ПОЛ	РОСТ	WCC
наблюдение 1	мужчина	низкий	101
наблюдение 2	мужчина	высокий	110
наблюдение 3	мужчина	высокий	92
наблюдение 4	женщина	высокий	112
наблюдение 5	женщина	низкий	95
.	.	.	.
.	.	.	.

Результаты группировки можно представить следующим образом (предполагается, что *ПОЛ* - первая независимая переменная, а *РОСТ* - вторая).

По всей выборке Среднее=100 СтдОткл=13 N=120			
Все мужчины Среднее=99 СтдОткл=13 N=60		Все женщины Среднее=101 СтдОткл=13 N=60	
Высок./ мужчины Средн.=98 СтдОткл=13 N=30	Низкие/ мужчины Средн.=100 СтдОткл=13 N=30	Высок./ женщины Средн.=101 СтдОткл=13 N=30	Низкие/ женщины Средн.=101 СтдОткл=13 N=30

Описательные статистики, расположенные в середине таблицы, определяются выбранным порядком независимых переменных. Например, в этой таблице приведены средние значения для “всех мужчин” и “всех женщин”, но не показаны средние для “всех людей высокого роста” или для “всех людей низкого роста”, которые можно вычислить, если определить *РОСТ* как первую группирующую переменную (а не как вторую).

Таким образом, программа группировки позволяет легко “переупорядочить” независимые переменные и увидеть различные “пересечения” данных. Модуль *Основные статистики и таблицы* позволяет не только переупорядочить переменные в таблицах, но также вычислить нужную маргинальную таблицу. Например, вы можете легко вычислить средние для *высоких* и *низких* людей.

Статистические тесты для группированных данных. Группировка часто используется как средство разведочного анализа. Обычный вопрос, который задает исследователь: являются ли группы, созданные независимыми переменными, действительно различными? Если вы интересуетесь различиями средних значений, то подходящим тестом является однофакторный дисперсионный анализ (ANOVA) (*F критерий*). Если интерес представляет различие дисперсий, то можно воспользоваться критерием однородности дисперсии. Процедура группировки системы *STATISTICA* предлагает критерий Левена однородности дисперсии (в настоящее время

StatSoft®

Copyright © StatSoft, 1995

этот критерий практически заменил более старые и менее устойчивые критерии Бартлетта и хи-квадрат; разнообразный набор критериев имеется в модуле *ANOVA/MANOVA*).

Другие близкие методы анализа данных. Хотя в разведочном анализе можно строить классификацию с более чем одной независимой (группирующей) переменной, статистические процедуры, используемые для их анализа, предполагают, что существует только один группирующий фактор (даже если, фактически, результаты получаются комбинированием определенного числа группирующих переменных). Таким образом, эти статистики не обнаруживают и даже не принимают во внимание возможное *взаимодействие* между группирующими переменными.

Например, вполне естественно допустить, что имеются различия между влиянием одной независимой переменной на зависимую переменную на разных уровнях другой независимой переменной (например, высокие люди могут иметь более низкий показатель *WCC,* чем низкие, однако, возможно, это относится только к мужчинам; см. "дерево" данных выше). Вы можете объяснить подобные эффекты, проверяя группировку "визуально" (в таблицах данных и на графиках) и используя различный порядок независимых переменных. Однако величина или значимость таких эффектов не может быть оценена здесь статистически (чтобы сделать это, используйте модуль дисперсионного анализа *ANOVA/MANOVA*, см. глава 13).

Апостериорное сравнение средних. В модуле *Основные статистики и таблицы* имеется обширный набор *апостериорных* критериев для сравнения средних. Обычно после получения статистически значимого результата в дисперсионном анализе (ANOVA), желательно знать, какие средние вызвали наблюдаемый эффект (например, какие группы особенно сильно отличаются друг от друга). Конечно, можно выполнить серию простых *t*-критериев, чтобы сравнить все возможные пары средних. Однако в связи с потенциально большим числом парных сравнений, такая процедура чисто случайно *увеличивает шансы* получить значимый результат. Представьте, вы имеете 20 выборок, по 10 случайных чисел в каждой и вычислили для них средние. Теперь попарно сравните эти средние между собой (всего нужно выполнить 180 сравнений). При этом вы можете ожидать, что на уровне 5% (1/20) примерно 9 разностей (из 180) случайным образом окажутся значимо различными (см. *Элементарные понятия статистики*, глава 8). Далее возьмите выборку с наибольшим средним и сравните с выборкой, в которой среднее является наименьшим. *t-критерий* для независимых выборок будет проверять, значимо или незначимо отличаются эти средние, *предполагая, что имеет дело со случайными выборками.* Иными словами, критерий не подозревает, что мы пытаемся обмануть его и в действительности берем разность максимального и минимального среднего. Очень вероятно, что при большом переборе обязательно найдется высоко значимая разность. Процедуры апостериорного сравнения специально рассчитаны на то, что проводится много попарных сравнений средних этих групп (можно также сказать, что эти процедуры проводят "коррекцию на число выполненных сравнений").

Модуль *ANOVA/MANOVA.* Модуль *ANOVA-/MANOVA* предлагает другие, существенно более тонкие критерии оценки различий между группами, созданными рядом независимых (*группирующих*) переменных, такие, например, как описательные статистики для более чем одной зависимой переменной, корреляции между несколькими зависимыми переменными в группах и т.д.

Группировка – внутригрупповые корреляции. В дополнении к описательным статистикам внутри групп в модуле *Основные статистики и таблицы* можно вычислить

Copyright © StatSoft, 1995

внутригрупповые корреляции. Это позволяет оценить различие между значениями корреляций в разных группах. Процедура позволяет также (с помощью кнопки *Опции*) вычислить ковариации и уровни значимости внутри групп, а также число наблюдений, по которым построена оценка, средние и стандартные отклонения.

Например, используя процедуру *Корреляционные матрицы* в модуле *Основные статистики и таблицы* вы можете получить корреляцию *.07* между переменными *WCC* и *ХОЛЕСТ*. Однако, для того, чтобы получить более точную информацию о зависимости между переменными, следует воспользоваться процедурой *Группировка и однофакторная ANOVA* и вычислить корреляцию между двумя переменными внутри каждой группы (например, получить для *мужчин r = .54,* а для *женщин r = -.25*).

Группировка в сравнении с анализом дискриминантных функций. Группировку можно рассматривать как первый шаг другого типа анализа - *Анализа дискриминантных функций* (см. том III). Аналогично классификации, дискриминантный функциональный анализ исследует различия между группами, построенными с помощью значений (кодов) независимой (группирующей) переменной. Однако в дискриминантном анализе, как правило, одновременно рассматривается более одной независимой переменной, и определяются “типы” (классы) значений этих переменных. Говоря несколько технически, в дискриминантном анализе находят такие линейные комбинации зависимых переменных, которые наилучшим (оптимальным) образом определяют принадлежность наблюдения к определенному классу.

В частности, с помощью дискриминантного анализа можно проанализировать различия между тремя группами людей, выбравших определенные профессии (например, юристы, физики, инженеры), с точки зрения их успехов в школе. Можно утверждать, что этот анализ “объяснит” выбор профессии успехами по определенным предметам (т.е. объяснит, почему выбрана именно эта профессия, а не другая). Таким образом, дискриминантный анализ можно рассматривать как “естественное расширение” простой группировки. Более детальное обсуждение этого вида анализа дано в описании модуля *Дискриминантный анализ* (см. том III).

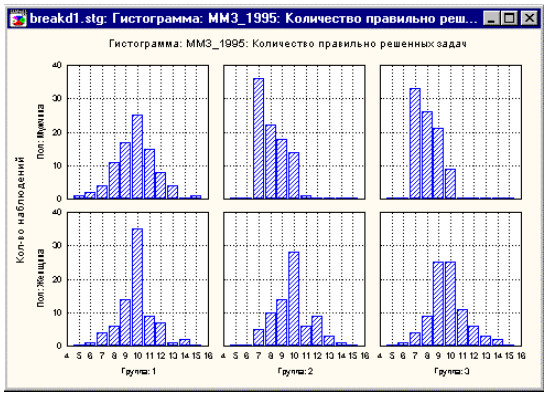
Группировка и таблицы частот. Другой вид анализа, который не может быть непосредственно проведен в диалоге группировка - это сравнения частот (*n*) в различных группах. Часто значения *n* в различных ячейках не равны между собой, потому что отнесение субъекта к определенной группе является следствием некоторых субъективных установок (т.е. здесь нет случайного выбора). Однако, если случайный выбор имеет место, то неравенство частот *n* в различных группах заставляет предположить, что независимые переменные связаны между собой.

Например, кросстабуляция уровней независимых переменных *Возраст* и *Образование* наиболее вероятно не создаст группы равной величины *n*, потому что степень образования различна для разных возрастных групп. Если вы хотите провести такие сравнения, то можете изучить определенные частоты в таблицах группировки и испытать различные способы упорядочивания независимых переменных. Однако, для того, чтобы подвергнуть разности частот статистическому исследованию, следует воспользоваться процедурой *Таблицы сопряженности,* которая предлагает различные способы проверки таблиц частот и таблиц сопряженности. Для продвинутого анализа сложных многовходовых таблиц (таблиц со многими входами) используйте модуль *Логлинейный анализ* (см. том III).

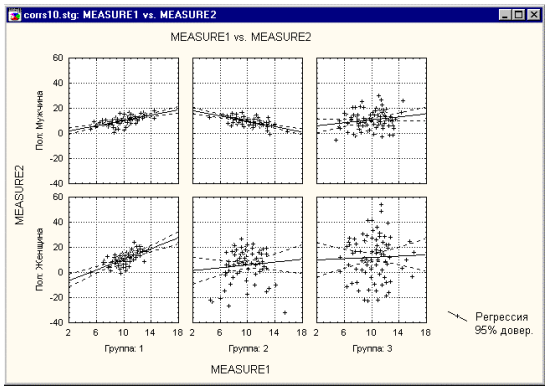
StatSoft®

Copyright © StatSoft, 1995

Графическое представление группировки. Графики часто позволяют обнаружить эффекты (как предполагаемые, так и неожиданные) быстрее, а иногда "лучше", чем численные методы. Диалоговое окно *Результаты* процедуры *Группировка* (стр. 1542) предлагает различные графические возможности для построения графиков средних, распределений, корреляций и т.д. "на пересечении" групп в соответствующих таблицах (например, категоризованные гистограммы, категоризованные вероятностные графики, категоризованные диаграммы размаха).



Категоризованные диаграммы рассеяния позволяют увидеть различия между значениями зависимых переменных, разбитыми на группы с помощью группирующей переменной.



Графические возможности доступны с помощью процедур *Статистические графики* из выпадающего меню *Графика*.

Анализ путей для группированных данных. В системе *STATISTICA* имеются продвинутые методы, позволяющие сравнивать средние, дисперсии, ковариации, корреляции двух или большего числа групп. Этими методами можно также проверить гипотезы о "структуре" зависимостей между переменными в различных группах. Для этого обратитесь к модулю *Структурное моделирование* (том III).

Таблицы частот

Обзор

Цель. Частоты или одновходовые таблицы представляют собой простейший метод анализа категориальных (номинальных) переменных (см. *Элементарные понятия статистики*, глава 8). Часто их используют как одну из процедур разведочного анализа, чтобы просмотреть, каким образом различные группы данных распределены в выборке. Например, изучая зрительский интерес к разным видам спорта (возможно для целей рекламы), вы могли бы представить ответы респондентов в следующей таблице:

FOOTBALL: "Watching football" (sports.sta)

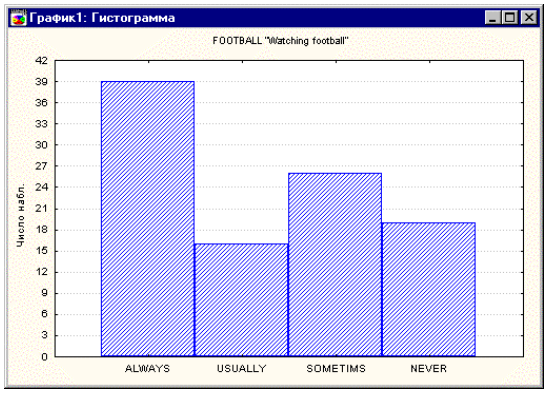
ОСНОВНЫЕ СТАТИСТ.	Частоты	Кумул. Частоты	Процент	Кумул. Процент
ALWAYS	39	39	39.00000	39.0000
USUALLY	16	55	16.00000	55.0000
SOMETIMS	26	81	26.00000	81.0000
NEVER	19	100	19.00000	100.0000
Пропущен	0	100	0.00000	100.0000

Верхняя таблица показывает число, долю и кумулятивную (накопленную) долю респондентов, характеризующих свой интерес к просмотру футбольных матчей в следующей шкале: (1) *Всегда интересуюсь - Always interested*, (2) *Обычно интересуюсь - Usually interested*, (3) *Иногда интересуюсь - Sometimes interested* или (4) *Никогда не интересуюсь - Never interested*. Конечно, система *STATISTICA*

Copyright © StatSoft, 1995

StatSoft®

обеспечивает исследователя разнообразными возможностями, позволяющими описать различные категории респондентов, присутствующие в таблице частот (например, используя “все отличные между собой значения” переменных). Также имеются различные графические возможности, позволяющие визуализировать распределение переменных (см. диалоговое окно *Таблицы частот*, стр. 1553).



Расположение данных. Любая переменная из множества данных может быть проанализирована и представлена в виде таблицы частот. Исследователь может также ввести определенные коды для таблицы, задать интервалы и даже определить ряд логических условий, позволяющих отнести наблюдение к определенной группе.

Приложения. Практически каждый исследовательский проект начинается с построения таблиц частот. Например, в социологических опросах таблицы частот могут отображать число мужчин и женщин, число респондентов из определенной этнической группы и т.д. Ответы, измеренные в определенной шкале (например, в шкале - интерес к футболу) можно также прекрасно свести в таблицу частот. В медицинских исследованиях можно табулировать пациентов с определенными симптомами. В промышленности - частоту выхода из строя элементов, приведших к авариям или отказам всего устройства при испытаниях на прочность (например, для определения какие детали телевизора действительно надежны после эксплуатации в аварийном режиме при большой температуре, а какие нет). Обычно, если в данных имеются категориальные переменные, то для них всегда вычисляются таблицы частот.

Таблицы сопряженности и таблицы флагов и заголовков (обзор)

Описание совместных распределений частот с помощью кросстабуляций

Цель и расположение данных.
Кросстабуляция – это процесс объединения двух (или нескольких) таблиц частот так, что каждая ячейка (клетка) в построенной таблице представляется единственной комбинацией значений кросстабулированных переменных. Таким образом, кросстабуляция позволяет совместить частоты появления наблюдений на разных уровнях рассматриваемых факторов. Исследуя эти частоты, можно определить зависимости между кросстабулированными переменными. Обычно кросстабулируются только группирующие (номинальные) переменные или переменные с относительно небольшим числом значений. Если вы хотите кросстабулировать непрерывную переменную (например, доход), то вначале ее следует *перекодировать,* разбив диапазон изменения на небольшое число интервалов (например, низкий, средний, высокий; см. *Панель инструментов и Строка состояния*, глава 2; см. также глава 3).

Таблицы 2 x 2. Простейшая форма кросстабуляции - это таблица 2 x 2, в которой значения двух переменных “пересечены” (сопряжены) и каждая переменная принимает только два значения, т.е. имеет два уровня (поэтому таблица и называется "2 на 2"). Рассмотрим поясняющий пример.

StatSoft®
Copyright © StatSoft, 1995

Предположим, проводится простое исследование, в котором мужчин и женщин спрашивают, какой напиток они предпочитают (газированную воду марки *A* или газированную воду марки *B*); файл данных показан ниже:

	ПОЛ	ГАЗ.ВОДА
наблюдение 1	МУЖЧИНА	A
наблюдение 2	ЖЕНЩИНА	B
наблюдение 3	ЖЕНЩИНА	B
наблюдение 4	ЖЕНЩИНА	A
наблюдение 5	МУЖЧИНА	B
	.	.
	.	.

Результаты кросстабуляции выглядят следующим образом.

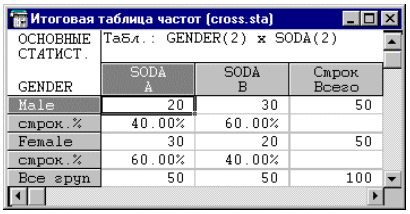
	ГАЗ.ВОДА:A	ГАЗ.ВОДА:B	
ПОЛ:МУЖЧИНА	20(40%)	30(60%)	50 (50%)
ПОЛ:ЖЕНЩИНА	30(60%)	20(40%)	50 (50%)
	50(50%)	50(50%)	100(100%)

Каждая ячейка таблицы содержит единственную комбинацию значений двух кросстабулированных переменных (в строке - указана переменная *ПОЛ,* в столбце - переменная *ГАЗ.ВОДА*). Каждая ячейка стоит на пересечении столбца и строки. Числа в каждой ячейке, на пересечении определенной строки и определенного столбца, показывают, сколько наблюдений соответствует данным значениям. Посмотрите на таблицу. К ней легко привыкнуть. Таблица показывает, что женщины больше мужчин предпочитают газированную воду марки *A*, мужчины больше предпочитают марку *B*. Таким образом, пол и предпочтение могут быть зависимыми (позже будет показано, как эту зависимость измерить).

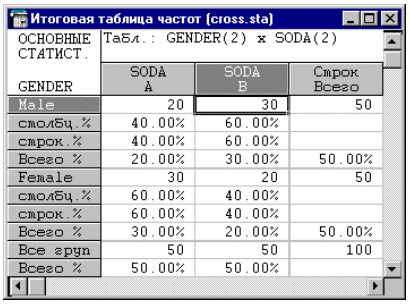
Маргинальные частоты. Значения, расположенные на краях таблицы – это просто одномерные таблицы частот для *всех* рассматриваемых переменных. Эти значения важны, т.к. позволяют оценить распределение частот в отдельных столбцах и строках. Например, 40% и 60% мужчин и женщин (соответственно), выбравших марку *A* (см. первый столбец таблицы), не могли бы показать какой-либо связи между переменными *ПОЛ* и *ГАЗ.ВОДА – Soda*, если бы маргинальные частоты переменной *ПОЛ* были также 40% и 60%. В этом случае они просто отражали бы разную долю мужчин и женщин, участвующих в опросе. Таким образом, различие в распределении частот в строках (или столбцах) отдельных переменных и в соответствующих маргинальных частотах дают информацию о зависимости кросстабулированных переменных.

Проценты по столбцам, по строкам и кумулятивные проценты. Приведенный пример показывает, что для оценки зависимости между кросстабулированными переменными, необходимо сравнивать маргинальные доли и индивидуальные доли в столбцах и строках. Такие сравнения легче провести с процентами.

Процедура *Итоговые таблицы* позволяет выдать кросстабулированные частоты в таблице результатов вместе с числом наблюдений, попавших в ячейку, процентами в столбцах и строках, а также суммарными процентами.



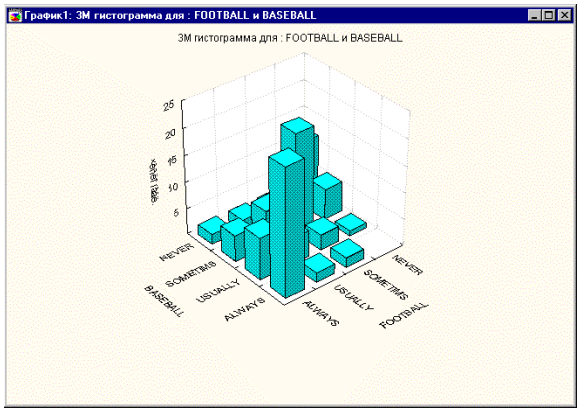
Можно построить итоговую объединенную таблицу, в которой каждая ячейка содержит эти числа.



Copyright © StatSoft, 1995

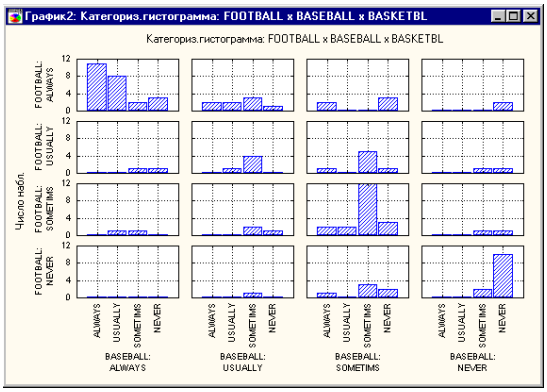
StatSoft®

Графическое представление кросстабуляций. Отдельные строки и столбцы таблицы удобно представить в виде графиков. Полезно также отобразить целую таблицу на отдельном графике. Имеется несколько способов сделать это с помощью процедуры *Таблицы сопряженности*. Таблицы с двумя входами можно визуально представить *3М гистограммой*.

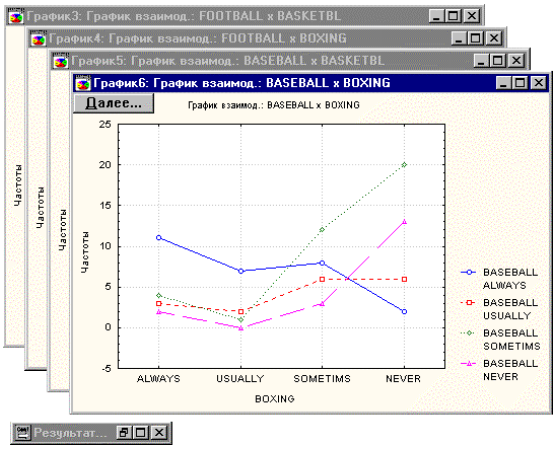


Другой способ визуализации таблиц сопряженности - построение *категоризованной гистограммы*, в которой каждая переменная представлена индивидуальными гистограммами, разбитыми на каждом уровне другой переменной (см. ниже).

Преимущество 3М гистограммы в том, что она позволяет представить на одном графике таблицу полностью. Достоинство категоризованного графика в том, что он дает возможность точно оценить специфические частоты в каждой ячейке.



Обратитесь к диалогу *Результаты кросстабуляции* (стр. 1559) для более полного представления возможностей графических процедур. Отметим, если выбрано более одной таблицы и выбрана какая-либо графическая опция, то на экране появится “каскад” графиков для каждой таблицы.



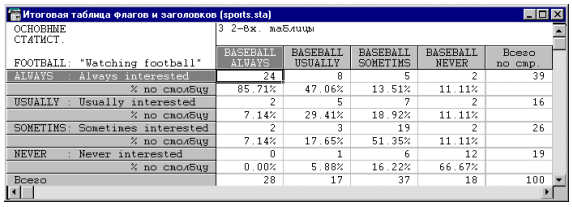
Нажмите кнопку *Далее* (см. верхний левый угол графика) и продолжите отображение графиков на экране. Нажатие “плавающей” кнопки (с пиктограммой *Cont* - “далее”) с именем диалогового окна, из которого эта последовательность графиков была вызвана (см. нижний левый угол экрана, показанного выше) вызовет прерывание процесса отображения графиков.

StatSoft

Copyright © StatSoft, 1995

Таблицы флагов и заголовков

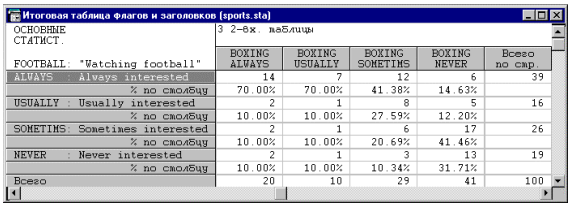
Таблицы флагов и заголовков или, кратко, таблицы *заголовков* позволяют отобразить несколько двумерных таблиц в сжатом виде как одну таблицу. Этот тип таблиц легко объяснить на примере файла интересов к спорту.



Итоговая таблица Флагов и заголовков (sports.sta)

ОСНОВНЫЕ СТАТИСТ. / FOOTBALL: "Watching football"	3 2-8x. таблицы				
	BASEBALL ALWAYS	BASEBALL USUALLY	BASEBALL SOMETIMS	BASEBALL NEVER	Всего по стр.
ALWAYS : Always interested	24	8	5	2	39
% по столбцу	85.71%	47.06%	13.51%	11.11%	
USUALLY : Usually interested	2	5	7	2	16
% по столбцу	7.14%	29.41%	18.92%	11.11%	
SOMETIMS: Sometimes interested	2	3	19	2	26
% по столбцу	7.14%	17.65%	51.35%	11.11%	
NEVER : Never interested	0	1	6	12	19
% по столбцу	0.00%	5.88%	16.22%	66.67%	
Всего	28	17	37	18	100

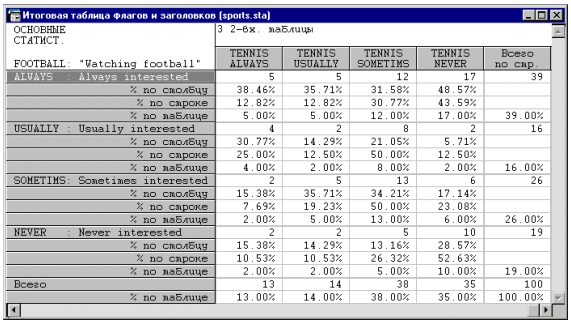
Интерпретация таблиц заголовков. В данной таблице результатов представлены три двувходовые таблицы, в которых интерес к *Футболу - Football* сопряжен с интересом к *Бейсболу - Baseball*, *Теннису - Tennis* и *Боксу - Boxing*. Таблица содержит информацию о процентах по столбцам, поэтому суммы по строкам равны 100%. Например, число в левом верхнем углу таблицы результатов (*85.71*) показывает, что 85.71 процентов всех респондентов ответили, что им *всегда интересно* смотреть футбол и *всегда интересно* смотреть баскетбол. Рассмотрите первый столбец приведенной таблицы. Вы видите, например, что имеется 2 респондента, обычно интересующихся футболом и всегда интересующихся баскетболом. Также 2 (других) респондента иногда интересуются футболом и всегда интересуются баскетболом. Нет ни одного респондента, которому был бы всегда интересен баскетбол и никогда не интересен футбол. Аналогично интерпретируются другие столбцы. Если вы прокрутите таблицу вправо, то увидите, что процент тех, кому *всегда интересно* смотреть футбол и *всегда интересно* смотреть теннис равен 38.46; для бокса этот процент составляет 70.0 (см. таблицы ниже).



Итоговая таблица Флагов и заголовков (sports.sta)

ОСНОВНЫЕ СТАТИСТ. / FOOTBALL: "Watching football"	3 2-8x. таблицы				
	BOXING ALWAYS	BOXING USUALLY	BOXING SOMETIMS	BOXING NEVER	Всего по стр.
ALWAYS : Always interested	14	7	12	6	39
% по столбцу	70.00%	70.00%	41.38%	14.63%	
USUALLY : Usually interested	2	1	8	5	16
% по столбцу	10.00%	10.00%	27.59%	12.20%	
SOMETIMS: Sometimes interested	2	1	6	17	26
% по столбцу	10.00%	10.00%	20.69%	41.46%	
NEVER : Never interested	2	1	3	13	19
% по столбцу	10.00%	10.00%	10.34%	31.71%	
Всего	20	10	29	41	100

Проценты в столбце (*Всего по строке*), показанные после каждого набора переменных, всегда связаны с общим числом наблюдений. В диалоговом окне *Результаты кросстабуляции* имеется множество процедур, позволяющих построить таблицы заголовков в различных форматах. Например, можно одновременно отображать число наблюдений в ячейках, строках, столбцах и общие проценты в одной и той же таблице.



Итоговая таблица Флагов и заголовков (sports.sta)

ОСНОВНЫЕ СТАТИСТ. / FOOTBALL: "Watching football"	3 2-8x. таблицы				
	TENNIS ALWAYS	TENNIS USUALLY	TENNIS SOMETIMS	TENNIS NEVER	Всего по стр.
ALWAYS : Always interested	5	5	12	17	39
% по столбцу	38.46%	35.71%	31.58%	48.57%	
% по строке	12.82%	12.82%	30.77%	43.59%	
% по таблице	5.00%	5.00%	12.00%	17.00%	39.00%
USUALLY : Usually interested	4	2	8	2	16
% по столбцу	30.77%	14.29%	21.05%	5.71%	
% по строке	25.00%	12.50%	50.00%	12.50%	
% по таблице	4.00%	2.00%	8.00%	2.00%	16.00%
SOMETIMS: Sometimes interested	2	5	13	6	26
% по столбцу	15.38%	35.71%	34.21%	17.14%	
% по строке	7.69%	19.23%	50.00%	23.08%	
% по таблице	2.00%	5.00%	13.00%	6.00%	26.00%
NEVER : Never interested	2	2	5	10	19
% по столбцу	15.38%	14.29%	13.16%	28.57%	
% по строке	10.53%	10.53%	26.32%	52.63%	
% по таблице	2.00%	2.00%	5.00%	10.00%	19.00%
Всего	13	14	38	35	100
% по таблице	13.00%	14.00%	38.00%	35.00%	100.00%

Многовходовые таблицы с контрольными переменными. Когда кросстабулируются только две переменные, результирующая таблица называется *двувходовой (двумерной)*. Конечно, общую идею кросстабулирования можно обобщить на большее число переменных. В примере с “газированной водой”, добавим третью переменную с информацией о штате, в котором проводилось исследование (*Небраска* или *Нью-Йорк*).

Copyright © StatSoft, 1995

StatSoft®

	ПОЛ	ГАЗ.ВОДА	ШТАТ
наблюдение 1	МУЖЧИНА	A	НЕБРАСКА
наблюдение 2	ЖЕНЩИНА	B	НЬЮ-ЙОРК
наблюдение 3	ЖЕНЩИНА	B	НЕБРАСКА
наблюдение 4	ЖЕНЩИНА	A	НЕБРАСКА
наблюдение 5	МУЖЧИНА	B	НЬЮ-ЙОРК
.	.	.	.
.	.	.	.

Кросстабуляция этих 3-х переменных представлена в следующей таблице:

	ШТАТ: НЬЮ-ЙОРК			ШТАТ: НЕБРАСКА		
	ГАЗ.ВОДА: A	ГАЗ.ВОДА: B		ГАЗ.ВОДА: A	ГАЗ.ВОДА: B	
П:МУЖЧИНА	20	30	50	5	45	50
П:ЖЕНЩИНА	30	20	50	45	5	50
	50	50	100	50	50	100

Теоретически любое число переменных может быть кросстабулировано в одной многовходовой таблице. Однако на практике возникают сложности с проверкой и "пониманием" таких таблиц, если они содержат более четырех переменных. (Хотя процедура *Кросстабуляция* позволяет строить таблицы очень большой сложности, рекомендуется анализировать зависимости между факторами в таких таблицах с помощью более продвинутых методов, например, использовать *Логлинейный анализ*, том III.)

Графическое представление многовходовых таблиц. Как и во всех видах анализа в *STATISTICA* имеется большое число графиков, помогающих интерпретировать многовходовые таблицы. Вы можете построить "дважды категоризованные" гистограммы, 3М гистограммы или линейные графики, позволяющие свести частоты для более чем 3-х факторов в один график, наборы (каскады) графиков. Обратитесь к диалоговому окну *Результаты кросстабуляции* (стр. 1559) для более подробного описания этих графических процедур.

Статистики таблиц сопряженности

Обзор

Таблицы сопряженности позволяет определить зависимость между кросстабулированными переменными. Следующая таблица отчетливо показывает очень сильную зависимость между двумя переменными: переменная *ВОЗРАСТ* (*ВЗРОСЛЫЙ* или *РЕБЕНОК*) и переменная предпочитаемый сорт *ПЕЧЕНЬЕ* (сорт *A* или сорт *B*).

	ПЕЧЕНЬЕ:A	ПЕЧЕНЬЕ:B	
ВОЗРАСТ:ВЗРОСЛЫЙ	50	0	50
ВОЗРАСТ:РЕБЕНОК	0	50	50
	50	50	100

Из этой таблицы видно, что все взрослые выбирают печенье *A*, а все дети печенье *B*. В данном случае, нет никаких оснований сомневаться в надежности этого факта. Невозможно поверить, что данная структура частот носит чисто случайный характер. Мало кто усомнится, что между предпочтениями детей и взрослых имеется отчетливое различие. Однако в реальной обстановке зависимости между переменными значительно слабее, и поэтому возникает вопрос, как их измерить и оценить надежность (статистическую значимость).

Далее обсуждаются самые общие меры зависимости между *двумя* группирующими переменными. Методы, используемые для анализа зависимостей между *более чем двумя* переменными в кросстабуляциях высокого порядка обсуждаются в модуле *Логлинейный анализ* (см. том III).

Таблицы сопряженности с числом переменных до 6 (6-ти мерные таблицы) можно построить в модуле *Основные статистики и таблицы* автоматически. Таблицы практически неограниченной сложности можно построить с помощью задания условий выбора наблюдений

StatSoft®

Copyright © StatSoft, 1995

(см. глава 3). Все меры зависимости между кросстабулированными переменными приводятся для двумерных таблиц, даже если эти таблицы являются слоями большой многовходовой таблицы (см. описание диалогового окна *Результаты кросстабуляции,* см. стр. 1559).

Критерий *хи-квадрат* Пирсона. *Хи-квадрат* Пирсона – это наиболее простой критерий проверки значимости зависимостей между группирующими переменными. Критерий Пирсона основывается на том, что в двувходовой таблице *ожидаемые* частоты при гипотезе, что между переменными нет зависимости, можно непосредственно вычислить.

Представьте, что опрошено 20 мужчин и 20 женщин относительно выбора газированной воды (марка *A* или марка *B*). Если между выбором и полом нет зависимости, то естественно *ожидать* равного выбора марки *A* и марки *B* для каждого пола. Критерий *хи-квадрат* становится высоко значимым при отклонении реально наблюдаемых частот в таблице от ожидаемых, иными словами, когда выбор мужчин и женщин различен.

Значение статистики *хи-квадрат* и ее уровень значимости зависит от общего числа наблюдений и количества ячеек в таблице. В соответствии с принципами, обсуждаемыми в разделе *Элементарные понятия статистики* (глава 8), относительно малые отклонения наблюдаемых частот от ожидаемых будет доказывать значимость, если число наблюдений велико.

Имеется только одно существенное ограничение использования критерия *хи-квадрат* (кроме очевидного предположения о случайном выборе наблюдений) - ожидаемые частоты должны быть не очень малы. Аргумент в пользу этого предположения состоит в том, что *хи-квадрат* по своей природе тестирует *вероятности* в каждой ячейке; и когда ожидаемые частоты в ячейках, становятся, например, меньше 5, эти вероятности нельзя оценить с достаточной точностью. Дальнейшие обсуждения см. в работах Everitt (1977), Hays (1988), или Kendall and Stuart (1979).

Критерий *Хи-квадрат*, основанный на методе максимального правдоподобия (Критерий МП хи-квадрат). *Максимум правдоподобия хи-квадрат* предназначен для проверки тех же самых гипотез, что и статистика *Хи-квадрат* Пирсона. Однако его вычисление основано на методе максимального правдоподобия. На практике статистика МП хи-квадрат очень близка по величине к обычной статистике Пирсона Хи-квадрат. Подробнее об этой статистике можно прочитать в работах Bishop, Fienberg, and Holland (1975) или Fienberg (1977). В разделе *Логлинейный анализ* (том III) она обсуждается более подробно.

Поправка Йетса. Аппроксимация статистики *хи-квадрат* для таблиц 2x2 с малыми частотами может быть улучшена понижением абсолютного значения разностей между ожидаемыми и наблюдаемыми частотами на величину 0.5 перед возведением в квадрат (*поправка Йетса*). Поправка Йетса, делающая оценку более умеренной, обычно применяется в тех случаях, когда таблицы содержат только малые частоты, таким образом, некоторые ожидаемые частоты становятся меньше 10 (дальнейшее обсуждение см. в Conover, 1974; Everitt, 1977; Hays, 1988; Kendall and Stuart, 1979 и Mantel, 1974).

Точный критерий Фишера. Этот критерий применим только для таблиц 2x2. Критерий основан на следующем рассуждении. Даны маргинальные частоты в таблице, предположим, что оба фактора в таблице независимы. Зададимся вопросом: какова вероятность получения наблюдаемых в таблице частот, исходя из маргинальных? Эта вероятность вычисляется *точно* подсчетом всех таблиц, которые можно построить, исходя из маргинальных частот. Таким образом, критерий

Copyright © StatSoft, 1995

StatSoft®

Фишера вычисляет *точную* вероятность появления наблюдаемых частот при нулевой гипотезе. Вычисляются односторонние и двусторонние вероятности.

Макнемара хи-квадрат. Этот критерий применяется, когда частоты в таблице 2x2 представляют *зависимые* выборки. Например, наблюдения одних и тех же индивидуумов до и после эксперимента. Вы можете подсчитывать число студентов, имеющих минимальные успехи по математике в начале и в конце семестра. Вычисляются два значения *хи-квадрата*: *A/D* и *B/C*. *A/D хи-квадрат* проверяет гипотезу о том, что частоты в ячейках *A* и *D* (верхняя левая, нижняя правая) одинаковы. *B/C хи-квадрат* проверяет гипотезу о равенстве частот в ячейках *B* и *C* (верхняя правая, нижняя левая).

Коэффициент *Фи*. *Фи-квадрат* представляет собой меру зависимости между двумя группирующими переменными в таблице 2x2. Его значения изменяются от *0* (нет зависимости между факторами; *хи-квадрат = 0.0*) до *1* (абсолютная зависимость между двумя факторами в таблице). Детали см. в Castellan and Siegel (1988, стр. 232).

Тетрахорическая корреляция. Эта статистика вычисляется (и применяется) только для таблиц сопряженности 2x2. Если таблица 2x2 может рассматриваться как результат (искусственного) разбиения двух непрерывных переменных на два класса, то коэффициент тетрахорической корреляции будет оценивать зависимость между двумя этими переменными.

Коэффициент сопряженности *C*. Коэффициент сопряженности представляет собой основанную на статистике *хи-квадрат* меру зависимости между двумя группирующими переменными (предложенную Пирсоном, создателем критерия *хи-квадрат*). Преимущество этого коэффициента перед обычным *хи-квадрат* в том, что он легче интерпретируется, т.к. диапазон его изменения от *0* до *1* (где *0* означает полную независимость). Недостаток в том, что верхний предел "ограничен" размером таблицы; *C* может достигать значения 1 только, если число классов не ограничено (см. 1956, стр. 201).

Интерпретация мер сопряженности. Существенный недостаток мер сопряженности (описанных выше) связан с трудностью их интерпретации в обычных терминах вероятности или "доли вариации", как в случае коэффициента корреляции *r* Пирсона (см. *Корреляции*, стр. 1449). Поэтому не существует общепринятой меры зависимости между классами.

Статистики, основанные на рангах

Во многих случаях классы, используемые в кросстабуляции, содержат информацию о ранговом упорядочивании объектов; иными словами, имеются измерения лишь в *ординальной (порядковой)* шкале (см. *Элементарные понятия статистики*, глава 8). Предположим, вы опросили некоторое множество респондентов для того, чтобы выяснить их отношение к некоторым видам спорта. Затем представили измерения в 4-х точечной шкале со следующими метками: (1) *всегда - always*, (2) *обычно - usually*, (3) *иногда - sometimes* и (4) *никогда - never interested*. Очевидно, что ответ *иногда интересуюсь - sometimes interested* показывает меньший интерес, чем *обычно интересуюсь- usually interested*, *обычно интересуюсь - usually interested* меньший интерес, чем *всегда интересуюсь - always interested* и т.д.

Таким образом, можно упорядочить (ранжировать) респондентов по их отношению, например, к футболу. Для таких переменных имеются свои типы корреляции, позволяющие численно выразить зависимости между ними.

Статистика *R* Спирмена. *R* корреляцию Спирмена можно интерпретировать так же, как и корреляцию Пирсона (*r* Пирсона) в терминах объясненной доли дисперсии (принимая во

StatSoft®

Copyright © StatSoft, 1995

внимание, что статистика Спирмена вычислена по рангам). Предполагается, что переменные измерены как минимум в *порядковой* шкале. Всестороннее обсуждение статистики *R* Спирмена, ее мощности и эффективности можно найти, например, в книгах Gibbons (1985), Hays (1981), McNemar (1969), Siegel (1956), Siegel and Castellan (1988), Kendall (1948), Olds (1949), или Hotelling and Pabst (1936). См. также *Непараметрическая статистика и подгонка распределения* (глава 11).

***Tau* Кендалла.** Статистика *may* Кендалла эквивалентна *R* Спирмена при выполнении некоторых основных предположений. Также эквивалентны их мощности. Однако обычно значения *R* Спирмена и *may* Кендалла различны, потому что они отличаются как своей внутренней логикой, так и способом вычисления. Siegel and Castellan (1988) выразили соотношение между этими двумя мерами следующим неравенством:

-1 ≤ 3*тау Кендалла - 2*R Спирмена ≤ 1

Более важно то, что статистики Кендалла *may* и Спирмена *R* имеют различную интерпретацию: в то время как статистика *R* Спирмена может рассматриваться как прямой аналог статистики *r* Пирсона, вычисленный по рангам, статистика Кендалла *may* скорее основана на вероятности. Более точно, проверяется, что имеется различие между вероятностью того, что наблюдаемые данные расположены в том же самом порядке для двух величин и вероятностью того, что они расположены в другом порядке. См. также модуль *Непараметрическая статистика* (глава 11).

Обычно вычисляется два варианта статистики *may* Кендалла: may_b и may_c. Эти меры различаются только способом обработки совпадающих рангов. В большинстве случаев их значения довольно похожи. Если возникают различия, то, по-видимому, самый безопасный способ - рассматривать наименьшее из двух значений. Kendall (1948, 1975), Everitt (1977), и Siegel and Castellan (1988) детально обсуждают *may* Кендалла.

Коэффициент *d* Соммера: d(X|Y), d(Y|X). Статистика *d* Соммера представляет собой несимметричную меру зависимости между двумя переменными. Эта статистика близка к may_b (см. Siegel and Castellan, 1988, стр. 303-310).

***Гамма*.** Если в данных имеется много совпадающих значений, статистика *Гамма* предпочтительнее статистики R Спирмена или *may* Кендалла. С точки зрения основных предположений, статистика *Гамма* эквивалентна статистике R Спирмена или тау Кендалла. Ее интерпретация и вычисления более похожи на статистику тау Кендалла, чем на статистику R Спирмена.

Говоря кратко, *Гамма* представляет собой также *вероятность*; точнее, разность между вероятностью того, что ранговый порядок двух переменных совпадает, минус вероятность того, что он не совпадает, деленную на единицу минус вероятность совпадений. Таким образом, статистика *Гамма* в основном эквивалентна тау Кендалла, за исключением того, что совпадения явно учитываются в нормировке. Детальное обсуждение статистики *Гамма* можно найти в Goodman and Kruskal (1954, 1959, 1963, 1972), Siegel (1956) и Siegel and Castellan (1988). См. также модуль *Непараметрическая статистика* (глава 11).

Коэффициенты неопределенности: S(X,Y), S(X|Y), S(Y|X). Эти коэффициенты измеряют *стохастическую зависимость,* имеющуюся между переменными. Понятие *стохастическая зависимость* берет начало в теоретико-информационном подходе к анализу таблиц частот и пользователь может обратиться к соответствующим руководствам (см. Kullback, 1959; Ku and Kullback, 1968; Ku, Varner, and Kullback, 1971; см. также Bishop, Fienberg, and Holland, 1975, стр. 344-348). Статистика *S(Y, X)* относится к симметричной зависимости,

Copyright © StatSoft, 1995

StatSoft®

статистики *S(X|Y)* и *S(Y|X)* выражают несимметричную зависимость.

Многомерные отклики и дихотомии

Обзор

Переменные типа многомерных откликов или многомерных дихотомий возникают в ситуациях, когда исследователя интересуют не только “простые” частоты событий, но также некоторые (часто неструктурированные) качественные свойства этих событий. Типичным примером является опрос общественного мнения, где вопросы, по крайней мере, частично имеют так называемые "открытые концы" (не подразумевают однозначного ответа), и респондент делает выбор из неограниченного (или очень большого) списка ответов. Вопрос состоит в том, как разумным образом закодировать эти ответы. Непосредственно сделать это нельзя, т.к. для одного наблюдения переменные могут принимать несколько значений. Природу многомерных переменных (факторов) лучше всего понять на примерах. Специальные процедуры табулирования таких данных обсуждаются в описании диалогового окна *Многомерные отклики/дихотомии* (см. стр. 1563).

Многомерные отклики

Представьте, что в процессе большого маркетингового исследования, вы попросили покупателей назвать 3 лучших, с их точки зрения, безалкогольных напитка. Обычный вопрос может выглядеть следующим образом:

```
Напишите ниже три ваших любимых
безалкогольных напитка:
1:_________     2:_________     3:_________
```

Анкета содержит от 0 до 3 ответов. Очевидно, список напитков может быть очень большим. Ваша цель - свести результаты в таблицу, в которой, например, будет подсчитан процент респондентов, предпочитающих определенный напиток.

Следующий шаг после получения анкет - занесение ответов в файл данных. Предположим, в ответах упоминалось 50 различных напитков. Вы могли бы, конечно, создать 50 переменных - одну для каждого напитка, рассмотреть респондентов как наблюдения (строки таблицы), ввести код *1* для респондента и переменной, если он предпочитают данный напиток (*0* если нет); например:

	COKE	PEPSI	SPRITE	....
наблюдение 1	0	1	0	
наблюдение 2	1	1	0	
наблюдение 3	0	0	1	
	.	.	.	
	.	.	.	

Такой метод кодирования откликов, т.е. приписывания им конкретных значений, очевидно, “расточителен”. Заметим, что каждый респондент дает максимум 3 ответа; однако для кодирования используется 50 переменных. (Если вы интересуетесь только тремя напитками, то такой метод кодирования будет успешным. Чтобы табулировать предпочтения в выборе напитка, следует рассмотреть 3 переменные как одну *многомерную дихотомию*; см. ниже.)

Кодирование многомерных откликов.

Более разумным является следующий подход. Введите 3 переменные и определите схему кодирования для 50 напитков. Затем введите соответствующие коды (*альфа* метки) для значений переменных и получите таблицу вида.

	Ответ_1	Ответ 2	Ответ_3
набл. 1	COKE	PEPSI	JOLT
набл. 2	SPRITE	SNAPPLE	DR PEPPER
набл. 3	PERRIER	GATORADE	MOUNTAIN DEW
.	.	.	.
.	.	.	.
.	.	.	.

Теперь, чтобы получить число респондентов, предпочитающих определенный напиток, рассмотрите переменные *Ответ1 - Ответ3* как

StatSoft®

Copyright © StatSoft, 1995

переменную с многомерным откликом. Само название переменной показывает, что она принимает многомерные значения. Таблица значений такой переменной имеет вид:

N=500 Категория	Число	Процент Ответов	Процент наблюдений
COKE	44	5.23	8.80
PEPSI	43	5.11	8.60
MOUNTAIN	81	9.62	16.20
PEPPER	74	8.79	14.80
...	...	...	...
Всего Ответов	842	100.00	168.40

Интерпретация таблиц частот с многомерными откликами. Итак, общее число респондентов в опросе *n=500*. Заметьте, что числа в первой колонке таблицы не составляют в сумме 500, как можно было бы ожидать, а равны 842. Вы поймете, почему это так, если вспомните, что каждый респондент может дать несколько ответов. Число, приведенное внизу первого столбца (на границе таблицы) - это общее число *ответов.* Каждый респондент может сделать до 3-х ответов (написать внизу 3 названия напитков), поэтому общее число ответов в действительности больше числа респондентов.

Возвращаясь к примеру, видим, что первое наблюдение (*Coke*, *Pepsi*, *Jolt*) "дает" три вклада в таблицу частот: в категорию *Coke*, в категорию *Pepsi* и один раз в категорию *Jolt*.

Вторые и третьи столбцы таблицы содержат проценты относительного числа ответов (второй столбец) так же, как все респонденты (третий столбец). Таким образом, вход *8.80* в первой строке и в последнем столбце таблицы означает, что 8.8% всех респондентов выбрали *Coke* первым, вторым или третьим пунктом ответа.

Повторные идентичные отклики. Как учитывать повторяющиеся ответы в одной и той же анкете? В отличие от других популярных программ, строящих таблицы для многомерных откликов, процедура *Кросстабуляция* в модуле *Основные статистики и таблицы,* по умолчанию, игнорирует одинаковые отклики. Например, если респондент ответил: *Jolt*, *Jolt, Jolt*, *STATISTICA* учтет *Jolt* только один раз. Следовательно, этот респондент в таблице частот будет учтен только один раз в группе *Jolt* (так же как в таблицах сопряженности; см. ниже), иными словами, в эту группу будет добавлена единица, а не тройка.

Многомерные дихотомии

Предположим, вас интересуют только *Coke*, *Pepsi* и *Sprite*. Как отмечалось, одним из способов кодирования является следующий:

	COKE	PEPSI	SPRITE
наблюдение 1		1	
наблюдение 2	1	1	
наблюдение 3			1
...	...	...	...

Здесь каждая переменная используется для одного напитка. Код *1* будет введен в таблицу в переменную всякий раз, когда соответствующий респондент указал ее в своем ответе. Заметим, что каждая переменная является *дихотомией*, т.к. принимает только два значения: "*1*" и "*не 1*" (можно ввести *1* и *0*, но так обычно не делается, можно просто рассматривать *0* как пустую ячейку или пропуск). Когда табулируются такие значения, вы получите итоговую таблицу очень похожую на ту, что была показана раньше для переменных с многомерными откликами; из нее вы можете вычислить число и процент респондентов (и ответов) для каждого напитка. Таким образом, вы компактно представили три переменные *Coke*, *Pepsi*, *Sprite* одной переменной (*Безалкогольные напитки*) - *многомерной дихотомией*. Заметьте, для кодирования 3-х напитков использовано 3 одномерные дихотомии, для кодирования 10 напитков, понадобится 10 одномерных дихотомий и т.д.

Copyright © StatSoft, 1995

Кросстабуляция многомерных откликов и дихотомий

Процедура *Кросстабуляция* модуля *Основные статистики и таблицы* позволяет определить простые группирующие переменные (например, *ПОЛ: МУЖЧИНА* или *ЖЕНЩИНА*), многомерные отклики и многомерные дихотомии. Все эти типы переменных можно использовать в таблицах сопряженности. Например, вы можете "сопрячь" многомерную дихотомию *Безалкогольные напитки* (закодированную, как описано выше) с многомерным откликом *Любимая система быстрого питания* (со многими категориями, например, *ГАМБУРГЕРЫ*, *ПИЦЦА* и т.д.), а также с простой группирующей переменной *ПОЛ.*

Как и в таблице частот для обычных переменных, в таблице частот для многомерных переменных, можно вычислить проценты и маргинальные суммы или по общему числу респондентов или по общему числу ответов (откликов). Например, рассмотрим следующего гипотетического респондента:

ПОЛ	COKE	PEPSI	SPRITE	Еда1	Еда2
ЖЕНЩИНА	1	1		РЫБА	ПИЦЦА

Этот респондент - *ЖЕНЩИНА,* - назвал *Coke* и *Pepsi* своими любимыми напитками и *Рыбу* и *Пиццу* любимыми блюдами. В полной таблице сопряженности этот респондент будет представлен следующими наборами:

		Еда				Общее Число Ответов
ПОЛ	Напиток	ГАМ-БУРГ.	РЫБА	ПИЦЦА	...	
ЖЕНЩИНА	COKE		X	X		2
	PEPSI		X	X		2
	SPRITE					
МУЖЧИНА	COKE					
	PEPSI					
	SPRITE					

Данный респондент учитывается в таблице 4 раза. Дополнительно, он будет считаться дважды в столбце *ЖЕНЩИНА-COKE* маргинальных частот, если этот столбец запрошен для представления общего числа откликов. Если пользователь запрашивает маргинальные суммы, вычисленные как общее число респондентов, тогда этот респондент будет учитываться только один раз.

Парная кросстабуляция переменных с многомерными откликами

Особенность процедуры табулирования многомерных переменных состоит в их попарном рассмотрении. Лучше всего показать это на простом примере. Предположим, проводится обследование нынешних и бывших домовладений респондента. Вы попросили респондента описать три последних дома, которыми он владел (включая тот, которым он владеет в данный момент). Естественно, для некоторых из респондентов нынешний дом является самым первым (до этого они не приобретали дома в частную собственность). Другие владели домами раньше. Для каждого дома респондента просят написать количество квартир и число жильцов – членов семьи. Ниже показано, как ответ одного респондента (скажем, наблюдение *112*) может быть введен в файл данных:

№ набл	Комнаты 1	2	3	Число жильцов 1	2	3
112	3	3	4	2	3	5

Респондент имел три дома: первый из 3-х комнат, второй также из 3-х комнат, третий из 4-х комнат. Количество членов семьи также росло: в первом доме жило 2 человека, во втором - 3, в третьем - 5.

Пусть вы хотите кросстабулировать число комнат с числом жильцов для всех респондентов (например, чтобы понять, как количество комнат связано с числом жильцов). Один из способов – создать 3 различные таблицы с двумя входами; одну таблицу для одного дома. Вы можете также рассмотреть два фактора в этом исследовании (*Число комнат, Число жильцов*) как переменные со многими откликами. Однако,

StatSoft®

Copyright © StatSoft, 1995

очевидно, нет никакого смысла в приведенном примере с респондентом *112* учитывать значения *3* и *5* в ячейке *Комнаты - Жильцы* в таблице сопряженности (которые вы могли бы учитывать, если бы рассматривали два эти фактора как одинарные переменные с многомерными откликами). Другими словами, вы хотите игнорировать комбинацию жильцов в третьем доме с числом комнат в первом. Скорее всего, вам нужно рассматривать переменные *попарно*; вы хотели бы рассмотреть число комнат в первом доме вместе с числом жильцов в первом доме, число комнат во втором доме вместе с числом жильцов в нем и т.д. Так именно и происходит, когда программа выполняет *парную* кросстабуляцию многомерных переменных.

Заключительный комментарий

Иногда при создании сложных таблиц сопряженности с переменными типа многомерные отклики и дихотомии возникает следующий вопрос (в ваших вычислениях): какую "выбрать дорогу" или как точно будут учитываться наблюдения в файле данных. Лучший способ проверить, как программа строит соответствующую таблицу – рассмотреть простой пример, и по нему ясно увидеть, каким образом учитывается каждое наблюдение (какой вносит оно вклад). В разделе Примеры (стр. 1483) используется именно такой метод для того, чтобы показать, как вычисляются данные для таблиц, включая переменные типа многомерные отклики и многомерные дихотомии.

Другие критерии значимости

Критерии значимости

В главе *Элементарные понятия статистики* (глава 8) описывается логика, лежащая в основе критериев значимости. Для краткости, предположим, что имеется две группы - 50 мужчин и 50 женщин, в которых вы измеряете высоту каждого человека. Ваша цель – проверить являются или нет мужчины выше женщин в рассматриваемой совокупности, откуда была получена выборка (например, среди студентов). Подходящим статистическим критерием здесь мог бы быть *t-критерий для независимых выборок* (см. стр. 1458). Не вдаваясь в детали, заметим, что *t*-статистика представляет собой разность средних в двух группах, стандартизованную специальным образом, чтобы учесть изменчивость данных.

Теперь, предположим, вы повторили исследования несколько раз, делая повторную выборку из 50 мужчин и 50 женщин из популяции, измеряя их рост и вычисляя *t*-статистику для каждой повторной выборки. После повторения исследований 1000 раз вы можете построить график распределения *t*-статистики на основании 1000 "экспериментов"; это распределение называется *выборочным распределением*. К счастью, сделав некоторые предположения о природе и распределении интересующей вас переменной (*РОСТ*, в данном случае), вы можете сделать вывод, не прибегая к рассмотрению повторной выборки 1000 раз. В данном случае, если *РОСТ* нормально распределен и нет разности стандартизованных средних, вы можете рассмотреть процентные точки *t*-распределения, с помощью таблиц или с помощью *Вероятностного калькулятора* (см. *Быстрые основные статистики*, глава 10)

Продолжая пример, если только в одной выборке значение *t*-статистики будет очень "необычно" или неправдоподобно (например, с $p<.01$), то вы можете заключить (после того как будете уверены, что вариация в двух группах одна и та же), что предположение о равенстве средних не верно. Таким образом, вы отвергаете гипотезу, что женщины имеют такой же рост, как и мужчины.

Процедура *Другие критерии значимости* позволяет вычислить наиболее общие критерии

Copyright © StatSoft, 1995

значимости. В частности, проверить значимость различия в средних, корреляции, процентилях (для независимых выборок). Пользователь может выбрать соответствующие статистики (т.е. средние, коэффициенты корреляции или процентили) и другие параметры (например, объем выборок). За деталями обратитесь к описанию процедуры *Другие критерии значимости* (стр. 1567). (Заметим, что в *STATISTICA* также имеется общий *Вероятностный калькулятор* для вычисления различных непрерывных функций распределения, в том числе: *F*, *t*, *хи-квадрат*, *Вейбулла*, и т.д.; обратитесь к главе 10, *Быстрые основные статистики* за дополнительной информацией.)

StatSoft®

Copyright © StatSoft, 1995

ОБЗОР ПРОГРАММЫ

Модуль *Основные статистики и таблицы* предоставляет пользователю несколько наиболее употребляемых методов анализа, табулирования и тестирования данных. С помощью этого модуля можно обработать большие объемы данных, например, вычислить и исследовать корреляционную матрицу размером до 4,092x4,092 с попарным удалением пропусков. Заметим, что с помощью *Менеджера мегафайлов* могут быть вычислены корреляционные матрицы практически неограниченного размера (до 32,000x32,000) (см. том III). Как и в других модулях *STATISTICA*, все результаты в модуле *Основные статистики и таблицы* можно вывести на графики.

Щелкнув, например, на нужном результате в таблице вывода правой кнопкой мыши, получите доступ к гистограмме частот с наложенным графиком нормальной плотности, диаграмме рассеяния с прямой регрессии и доверительными интервалами или к диаграмме размаха. Можно создать разнообразные категоризованные графики для визуального анализа группировок, категоризованные диаграммы рассеяния, 2M и 3M диаграммы рассеяния, матричные графики, категоризованные диаграммы и многие другие графики, позволяющие проводить разведочный анализ данных. Большое разнообразие дополнительных графиков доступно с помощью процедур *Статистические графики* из выпадающего меню *Графика*. Краткий обзор процедур, доступных в модуле, дан ниже.

Описательные статистики

Процедура *Описательные статистики* вычисляет следующие статистики: средние, медианы, стандартные отклонения, доверительные интервалы для средних (с заданными уровнями доверия), эксцессы, асимметрию и их стандартные ошибки. Другие описательные статистики (порядковые статистики, мода, определенные пользователем процентили, гармонические и геометрические средние и т.д.), а также всесторонний набор процедур подгонки доступен в модуле *Непараметрическая статистика* (глава 11) и *Анализ процессов* (том IV). Самые разнообразные непрерывные распределения могут быть подогнаны к данным с помощью опций *Статистические 2M графики - Гистограммы*, *Графики квантиль-квантиль*, *Графики вероятность-вероятность* (см. том II).

Различные специальные описательные статистики доступны в модулях *Временные ряды/Прогнозирование* (том III), *Анализ выживаемости* (том III) и других модулях. Например, модуль *Временные ряды/Прогнозирование* содержит специальные статистики для описания данных, упорядоченных по времени, в модуле *Анализ выживаемости* доступны статистики для цензурированных данных.

Заметим, что основные описательные статистики (включая медиану и квартили) могут быть вычислены в каждом модуле системы *STATISTICA* с помощью опций *Быстрые основные статистики*, доступных из выпадающего меню *Анализ* или из панели инструментов *Быстрые основные статистики* с помощью кнопки (детали см. в главе 10); описательные статистики (и графики) для выбранных блоков данных в любой таблице результатов или таблице исходных данных вычисляются с помощью *Статистик блока* (*по столбцам* или *по строкам*), доступных из выпадающего меню *Правка* или меню, вызываемого правой кнопкой мыши (см. *Окно таблицы исходных данных*, глава 3).

Вычисление медианы и квартилей (25%, 50%, 75% процентили). Когда распределение значений нельзя точно разбить напополам и на

Copyright © StatSoft, 1995

четверти, имеется несколько приближенных способов вычислить медиану и квартили (а также процентили). Определенный метод вычисление этих статистик можно выбрать в диалоговом окне *STATISTICA Параметры по умолчанию: общие настройки* (см. глава 3; это диалоговое окно доступно с помощью выпадающего меню *Сервис* (опция *Общие настройки*).

Корреляции

Эта процедура вычисляет коэффициенты корреляции Пирсона. В модуле *Основные статистики и таблицы* можно вычислить квадратные и прямоугольные корреляционные матрицы, а также корреляционные матрицы с показанным числом парных удалений *n* и уровнями значимости (матрицы в расширенном формате). Значимость разностей между коэффициентами корреляции можно проверить с помощью процедуры *Другие критерии значимости* (см. ниже).

Имеется много различных типов коэффициентов корреляции, доступных в других модулях *STATISTICA*. Разнообразные *непараметрические* коэффициенты корреляции (например, *R* Спирмена, *тау* Кендалла, *Гамма*, и др.) включены в модуль *Непараметрическая статистика*; эти коэффициенты также вычисляются в процедуре *Таблицы сопряженности* данного модуля (стр. 1466), *частные корреляции* включены в модуль *Множественная регрессия* (том III), разнообразные *автокорреляции* (т.е. корреляции между переменной и ее сдвигом на определенный лаг) включены в модуль *Временные ряды/Прогнозирование* (том III).

Возможность вычисления стандартной корреляции Пирсона имеется также в других модулях (например, *Множественная регрессия, Факторный анализ*, *Менеджер мегафайлов*; см. том III). Заметим, что корреляционные матрицы для сгруппированных данных, могут быть вычислены в каждом модуле системы *STATISTICA* с помощью опций *Быстрые основные статистики*, доступных из выпадающего меню *Анализ* или с помощью кнопки на панели инструментов *Быстрые основные статистики* (детали см. в главе 10).

Тетрахорические корреляции можно вычислить в модуле *Надежность/позиционный анализ* (том III). Разнообразные меры расстояний вычисляются в модуле *Кластерный анализ* (том III), меры связи между группирующими (номинальными) переменными доступны с помощью процедуры *Кросстабуляция* в модуле *Основные статистики и таблицы* (см. стр. 1558).

Модуль *Нелинейное оценивание* (том III) содержит различные предопределенные нелинейные регрессионные модели для анализа зависимых группирующих переменных (пробит и логит анализ); пользователь может также задать любой тип регрессии в этом модуле. Модуль *Анализ выживаемости* (том III) содержит специализированные регрессионные программы для обработки данных типа времен жизни (опция *Регрессионные модели* в стартовой панели данного модуля); некоторые из этих моделей являются непараметрическими (свободными от параметров) по своей природе (например, регрессионная модель Кокса пропорциональных интенсивностей). Продвинутые методы анализа корреляционных матриц в терминах причинной зависимости между переменными доступны в модуле *Моделирование структурными уравнениями (SEPATH)* (том III).

t-критерий для независимых выборок

Эта процедура позволяет сравнивать средние двух групп (образованных наблюдениями одной переменной); значимость разности между групповым средним и средним популяции или

StatSoft®

Copyright © StatSoft, 1995

ожидаемым средним может быть проверена с помощью процедуры *Другие критерии значимости* (см. ниже). В модуле *Непараметрическая статистика* (глава 11) также доступны различные непараметрические критерии сравнения групп; методы сравнения групп (частично) цензурированных наблюдений доступны в модуле *Анализ выживаемости*.

t-критерий для зависимых выборок

Эта процедура позволяет сравнить средние переменных, определенных на одном и том же множестве наблюдений (субъектов, индивидуумов; *зависимых* выборок). Заметим, различные непараметрические критерии для сравнения распределений переменных доступны в модуле *Непараметрическая статистика* (глава 11).

Группировка: внутригрупповые описательные статистики и корреляции

Процедуры вычисляют различные описательные статистики (например, средние, стандартные отклонения, корреляции и т.д.) переменных, разбитых на группы с помощью одной или нескольких группирующих переменных (например, *Пол* и *Место жительства*). Программа выполнит также однофакторный дисперсионный анализ выбранных переменных, что особенно удобно для исследования однофакторных планов с очень большим числом групп (например, при числе групп больше 200, как часто бывает в сельскохозяйственных исследованиях).

Заметим, что вы можете вычислить различные описательные статистики (включая медианы и квартили), а также корреляционные матрицы переменных, разбитых на группы значениями группирующей переменной, в каждом модуле системы *STATISTICA* с помощью опций *Быстрые основные статистики*, доступных из выпадающего меню *Анализ* или из панели инструментов *Быстрые основные статистики* с помощью кнопки (детали см. в главе 10).

Таблицы частот

Данная процедура позволяет вычислить таблицы частот (и гистограммы). В этих таблицах представляются частоты попадания значений переменной (наблюдений) в разные классы (приводятся численные или численно буквенные значения и их метки). *STATISTICA* предлагает различные процедуры для определения категорий (классов) в таблицах частот (например, целые интервалы, определенные коды и т.д.). Пользователь может табулировать данные с помощью определенных условий, заданных в виде логических выражений. Заметим, что таблицы частот можно также вычислить с помощью опций *Быстрые основные статистики*, из выпадающего меню *Анализ* или панели инструментов *Быстрые основные статистики* с помощью кнопки (детали, см. в главе 10).

Таблицы сопряженности и таблицы флагов и заголовков

Это процедуры позволяют кросстабулировать данные (таблицы с числом входов до 6; многовходовые таблицы более высокого уровня можно строить, используя условия выбора) и построить разнообразные таблицы сопряженности. Здесь также доступно большое количество статистик (например, критерии *хи-квадрат*, *фи-квадрат*, *Гамма* и т.д.). Заметим, что модуль *Логлинейный анализ* позволяет строить и анализировать сложные многовходовые таблицы.

Многомерные отклики и дихотомии

Модуль *Основные статистики и таблицы* имеет разнообразные возможности построения итоговых таблиц для переменных с

Copyright © StatSoft, 1995

StatSoft®

многомерными откликами, а также для многомерных дихотомий. Обычно группирующие переменные или факторы делят выборку на непересекающиеся (эксклюзивные) группы, например, группу *мужчин* и *женщин*. Очевидно, достаточно только одной группирующей переменной, чтобы закодировать пол субъекта. Однако в некоторых исследованиях категории не исключают друг друга (пересекаются).

Например, в маркетинговых исследованиях респонденту можно задать вопрос о трех самых любимых безалкогольных напитках. Предположим, 60 различных напитков присутствует в ответах, которые можно закодировать тремя группирующими переменными (первые три предпочтения). В этом случае категории, очевидно, не являются взаимно исключающими. Действительно, человек может отметить 3 различных напитка, как предпочтительные. Следовательно, если наблюдение – это субъект, то для трех различных группирующих переменных это наблюдение является общим (не эксклюзивным). Такие группирующие переменные называют переменными с *многомерными откликами* (*многомерные дихотомии* по существу, схожи с ними). Эти переменные легко анализировать в модуле *Основные статистики и таблицы*.

Другие критерии значимости

Используйте эту процедуру, чтобы вычислить статистическую значимость различий между двумя независимыми выборками (группами) в средних, в коэффициентах корреляции и в процентилях. Можно использовать как односторонний, так и двусторонний критерии.

Альтернативные процедуры

Подробные описательные статистики (включая корреляционные матрицы) можно вычислить в каждом модуле системы *STATISTICA* с помощью опций *Быстрые основные статистики* (доступных из выпадающего меню *Анализ* или из панели инструментов *Быстрые основные статистики* с помощью кнопки (детали см. в главе 10). *Быстрые основные статистики* позволяют вычислить статистики для данных, разбитых на группы группирующей переменной. Итоговые описательные статистики (включая квартили) и графики для значений, содержащихся в любой таблице результатов (или таблице исходных данных) можно вычислить с помощью *Статистик блока* (*по строкам* или *по столбцам*), доступной из выпадающего меню *Правка* или контекстного меню, вызываемого правой кнопкой мыши. Описательные и категоризованные графики можно построить в каждом модуле *STATISTICA* с помощью опций *Статистические 2М графики*, *Статистические 3М графики* и *Статистические категоризованные графики* (см. том II). Заметим, что различные непараметрические критерии сравнения групп доступны в модуле *Непараметрическая статистика* (глава 11); методы для сравнения групп (частично) цензурированных наблюдений доступны в модуле *Анализ выживаемости*.

StatSoft®

Copyright © StatSoft, 1995

ПРИМЕРЫ

Пример 1: Пошаговое использование системы *STATISTICA*

Рассматривается пример, основанный на файле данных *Adstudy.sta,* который находится на дисках *STATISTICA*.

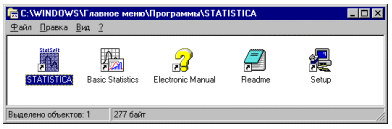
Инсталляция и запуск программы

Если вы еще не установили систему *STATISTICA*, пожалуйста, следуйте инструкциям, показанным ниже:

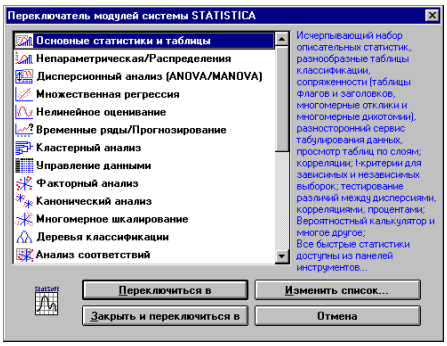
Для того чтобы установить систему *STATISTICA* при загруженном Windows, вставьте Диск 1 в дисковод A: и из меню *Файл* выберите *Запустить*. Наберите *A:SETUP* (вы так же можете использовать *B:SETUP*, если это отвечает конфигурации вашей системы) и следуйте инструкциям программы установки.

STATISTICA содержит полное *электронное руководство* (нажмите F1 в любой момент анализа, чтобы получить к ней доступ).

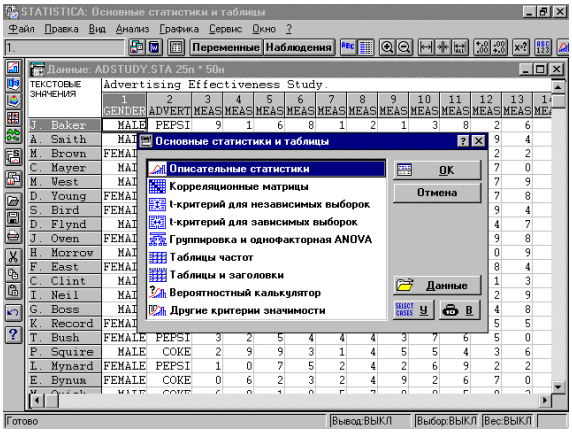
Когда вы впервые устанавливаете систему *STATISTICA*, программа *Setup* создаст группу приложений с названием *STATISTICA* и установит пиктограммы *Переключателя модулей* (см. ниже пиктограмму *STATISTICA*, первую в группе), модуля *Основные статистики и таблицы* и некоторых других программ (например, *Help, Setup*; вы можете также установить пиктограммы для других часто используемых модулей при помощи *Создать - Ярлык* из меню *Проводник*).



Запуск модуля *Основные статистики и таблицы*. Запустите программу, щелкнув на пиктограмме *Основные статистики и таблицы* в группе программ системы *STATISTICA* или щелкнув на пиктограмме *STATISTICA* и выбрав процедуру *Основные статистики и таблицы* из *Переключателя модулей* (см. ниже).



Программа откроется (в полноэкранном режиме), появится последний использовавшийся файл (*Adstudy.sta*, если программа используется в первый раз) и стартовая панель модуля *Основные статистики и таблицы*.

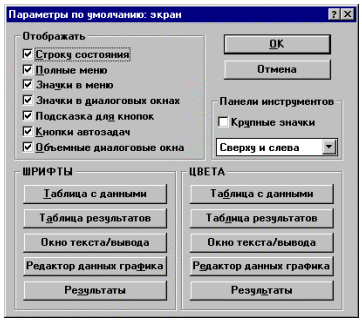


Copyright © StatSoft, 1995

StatSoft®

Настройка появления *STATISTICA*. Можно настроить практически любой вид появления системы *STATISTICA*. Глава *Основные соглашения* (глава 1) содержит подробное описание настроек. Например, можно настроить, чтобы система *STATISTICA* запустилась не в полноэкранном режиме, убрала стартовую панель или автоматически открыла окно вывода текста/графиков (см. в выпадающем меню *Сервис* команду *Общие настройки*).

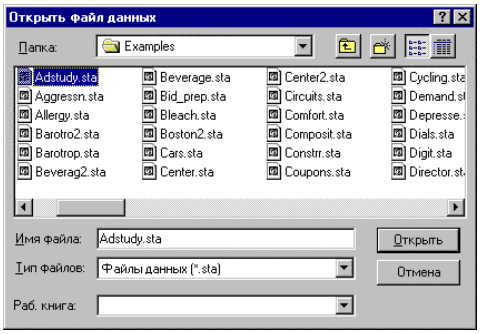
При первом запуске *STATISTICA*, возможно, вы захотите изменить цвет или формат различных установок изображения. Из выпадающего меню *Сервис*, выберите *Экран* для вызова диалогового окна *Параметры по умолчанию: экран*.



Изменения, которые вы внесете в этом диалоговом окне, будут применены ко всей вашей работе с системой *STATISTICA*, и сохранятся даже после выключения компьютера (выбранные установки запоминаются в файле *Statist.ini*; см. главу 1). Если вы используете маленький экран в режиме высокого разрешения, вы можете установить параметр *Крупные значки*; что приведет к увеличению кнопок панели инструментов в два раза. Также вы можете выбрать оптимальный формат панели инструментов, задав размер экрана вашего компьютера в области *Панели инструментов* диалогового окна. По умолчанию, панель инструментов отображается *Сверху и слева* экрана. Наконец, вы можете изменить цвета, шрифты и размер шрифтов, которые используются для отображения таблиц результатов, таблиц исходных данных, текстовых окон и т.д. После того, как вы сделали желаемые настройки, нажмите *OK* для продолжения.

Выбор файла данных. Текущий файл данных отображается в виде таблицы исходных данных. Если файл *Adstudy.sta* не открылся, вы можете выполнить команду *Открыть данные* из выпадающего меню *Файл* (или щелкнуть по кнопке *Открыть данные* на панели инструментов таблицы исходных данных) и выбрать файл *Adstudy.sta* в диалоговом окне *Открыть файл данных* (по умолчанию, в поддиректории *Примеры*).

Поддиректория *Примеры* содержит примеры файлов данных, используемых для иллюстрации возможностей системы *STATISTICA*.

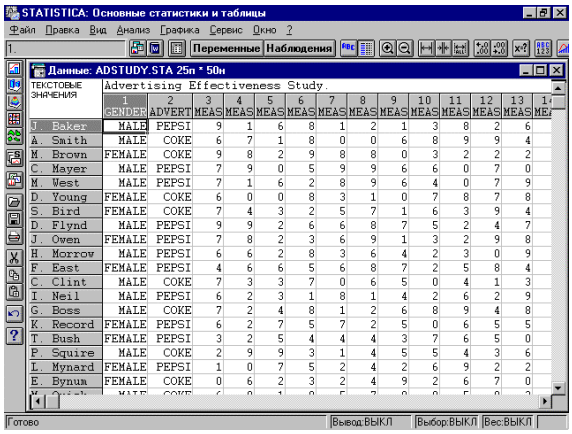


Файл *Adstudy.sta* содержит 25 переменных и 50 наблюдений. Эти (фиктивные) данные были собраны путем социологического опроса в рекламном исследовании, где женщины и мужчины оценивали качество двух реклам. Пол респондента кодировался с помощью переменной 1 (*ПОЛ - GENDER*: *1=МУЖЧИНА - MALE, 2=ЖЕНЩИНА - FEMALE*). Каждому респонденту случайным образом предлагался на рассмотрение один из двух рекламных роликов (*ПРЕДПОЧТЕНИЕ - ADVERT*: *1=Coke®, 2=Pepsi®*). Затем они оценивали привлекательность соответствующего продукта по 23 различным шкалам (с *Меры1 - Measur1* до

StatSoft®
Copyright © StatSoft, 1995

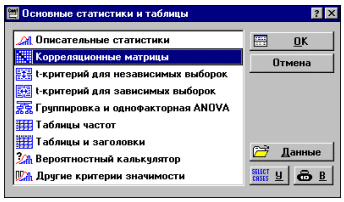
Меры23 - Measur23). В каждой из этих шкал респонденты могли дать ответы между 0 и 9.

Таблицы исходных данных. Когда файл с данными открылся, он отображается в виде таблицы исходных данных. Большинство результатов системы *STATISTICA* отображается в таблицах, похожих по внешнему виду на таблицы исходных данных, называемых таблицами результатов. Однако таблица исходных данных отличается от всех таблиц результатов тем, что она имеет особые возможности по управлению данными (см. *Окно таблицы исходных данных*, глава 3).

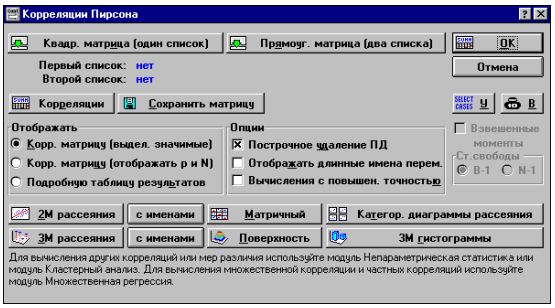


Корреляции

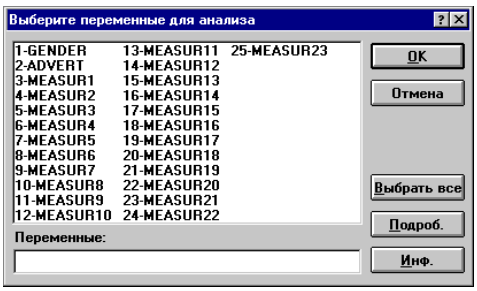
Сначала проверим, не коррелируют ли между собой оценки в различных шкалах (другими словами, не измеряют ли некоторые шкалы одни и те же свойства). В стартовой панели *Основные статистики и таблицы*, выберите процедуру *Корреляционные матрицы* и щелкните *OK* (или можете просто дважды щелкнуть на процедуре *Корреляционные матрицы*).



После выбора этой процедуры откроется диалоговое окно *Корреляции Пирсона*.



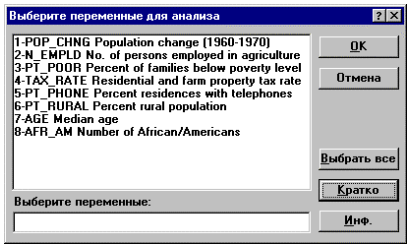
Вы можете выбрать переменные как из одного списка (т.е. матрица будет квадратной), так и из двух списков (прямоугольная матрица). Для нашего примера, щелкните на кнопке *Квадр. матрица (один список)*, и появится стандартное диалоговое окно для выбора переменных.



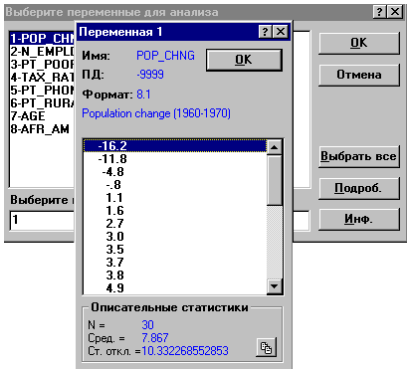
Просмотр и выбор переменных. Заметим, что если вы раньше экспериментировали с таблицей и выделили блок значений, то переменные, выделенные в блоке, будут выбраны автоматически. Окно выбора переменной поддерживает разные способы выделения переменных и предлагает различные короткие пути и возможности просмотра содержания файла данных. Например, вы можете нажать кнопку *Подробно*, чтобы расширить список переменных и посмотреть длинные метки переменных, формулы или связи.

Copyright © StatSoft, 1995

StatSoft®



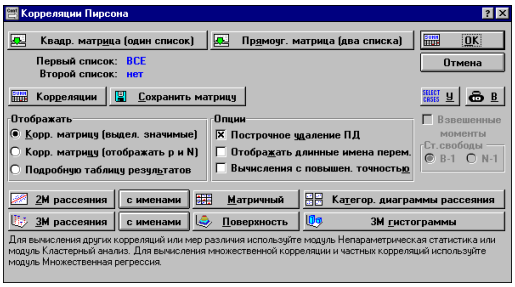
Если вы нажмете кнопку *Инф.*, то просмотрите список всех значений переменной и ее описательные статистики. Нажав кнопку *Копировать* в нижней части диалогового окна, вы скопируете описательные статистики на панель инструментов.



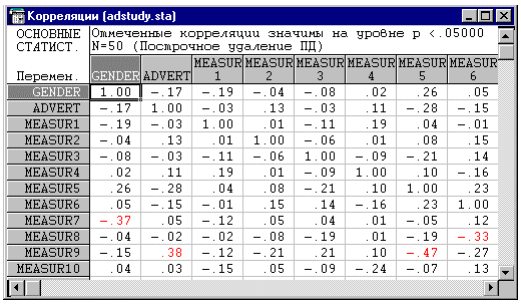
Переменные могут быть выбраны следующими способами. Вы можете:

- Щелкнуть по кнопке *Выбрать все* для выбора всех переменных, перечисленных в этом диалоговом окне,
- Выделить мышью область тех переменных, которые вы хотите выбрать,
- Ввести номера переменных или интервал номеров в поле ввода внизу списка (или * для выбора всех переменных),
- Выбрать разрывный список переменных, удерживая клавишу CTRL во время выделения переменных мышью (непрерывный список переменных можно выделить, если щелкнуть на первом и последнем элементах непрерывного блока переменных, удерживая при этом клавишу SHIFT).

Заметим, что можно отменить выделение отдельных переменных, нажав клавишу CTRL и щелкнув на переменной. Для нашего примера, выберите все переменные в файле данных и щелкните *OK*, чтобы вернуться в диалоговое окно *Корреляции Пирсона*.



В этом диалоговом окне имеется несколько процедур (см. стр. 1533); однако, для нашего примера, просто щелкните *OK*, принимая выбор, сделанный по умолчанию, и вызывая таблицу с результатами корреляции.

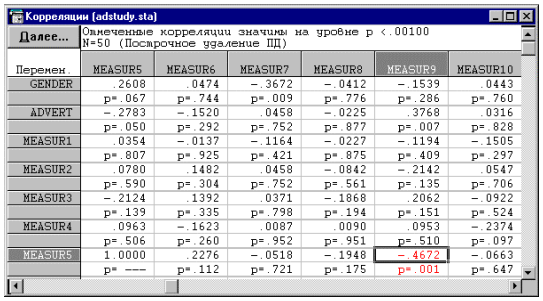


Выделение значимых корреляций. Вы можете указать уровень значимости (*альфа*) (*.05*, по умолчанию) для выделения значимых коэффициентов корреляции в таблице результатов. Чтобы изменить уровень *альфа*, щелкните по кнопке *Параметры* на панели инструментов таблицы результатов и откройте диалоговое окно *Уровень значимости*. Введите

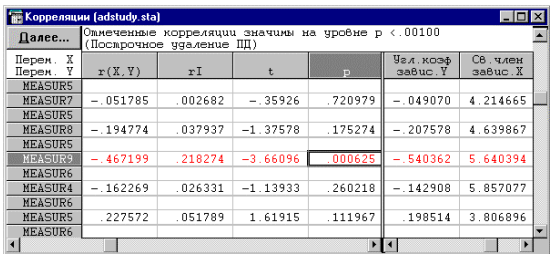
StatSoft®
Copyright © StatSoft, 1995

в это диалоговое окно уровень значимости *.001* и щелкните *OK*.

После выхода из этого диалогового окна, появится новая таблица результатов с корреляциями, в которой все соответствующие корреляции будут выделены. Теперь стало очень легко обнаружить эти высокие корреляции (например, корреляция между *Мерой5 - Measur5* и *Мерой9 - Measur9* равна *-.47*). Такая высокая корреляция показывает, что две шкалы оценок могут измерять одну и ту же характеристику зрительного восприятия рекламы (хотя одна мера этой характеристики возрастающая, а другая - убывающая). Две опции из диалогового окна *Корреляции Пирсона* позволяют получить таблицу данных с коэффициентами корреляции, а также более подробными статистиками (например, *p*-значение, число пар *N*, r^2, *t*-значения, и т.д.). Когда вы выберете установку *Корр. матрицу (отображать p и N)*, вместе с коэффициентами корреляции будут также выведены *p*-значения и число пар *N*.



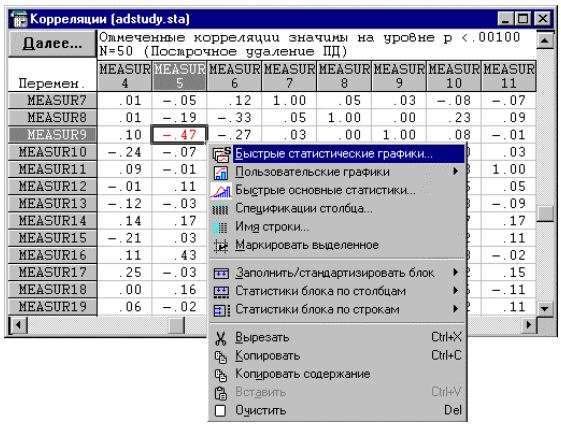
Выбор опции *Подробную таблицу результатов* в диалоговом окне *Корреляции Пирсона* возможен только при выборе 20 или меньше переменных для анализа, так как для каждой корреляции автоматически будет выводиться большое количество информации. После выбора этой опции будет построена таблица результатов, содержащая соответствующие описательные статистики, коэффициенты корреляции, *p*-значения и число пар *N*, а также наклон и отрезок регрессии для каждой переменной.



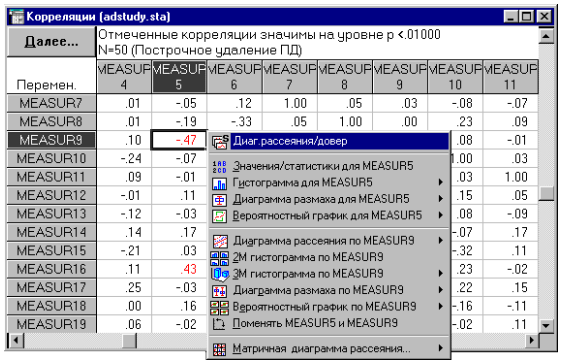
Эту опцию следует использовать только для отдельных корреляций (но не для подробного анализа), потому что в этом формате для каждого коэффициента корреляции будут заняты 22 ячейки таблицы результатов; таким образом, для матрицы корреляций 20x20 получится таблица результатов с 8,800 ячейками. Вы можете увидеть выше, что корреляция между *Мерой5 - Measur5* и *Мерой9 - Measur9*, действительно, значима (*p=.0006*), это означает, что ошибка, связанная с принятием результата составляет 6 из 10,000. То есть, если делать случайные выборки такого же размера из совокупности, в которой эти две переменные не коррелируют, то только шесть раз из каждых 10,000 получим выборку, в которой корреляция будет -.47 или меньше (см. *Элементарные понятия статистики*, глава 8).

Построение графиков для таблицы результатов. Чтобы визуализировать значения корреляций между переменными, можно построить график корреляций. Если щелкнуть по соответствующему коэффициенту корреляции (*-.47*) правой кнопкой мыши, то появится контекстное меню.

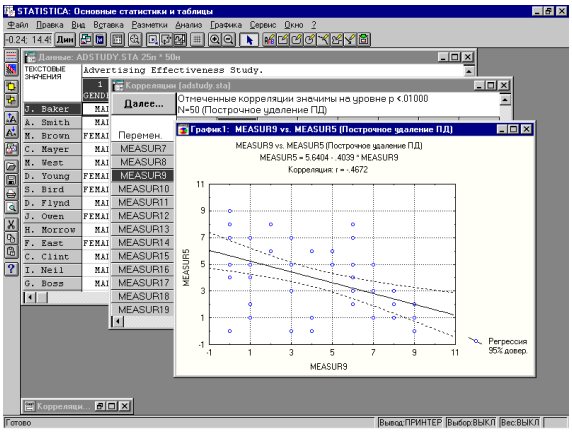
Copyright © StatSoft, 1995



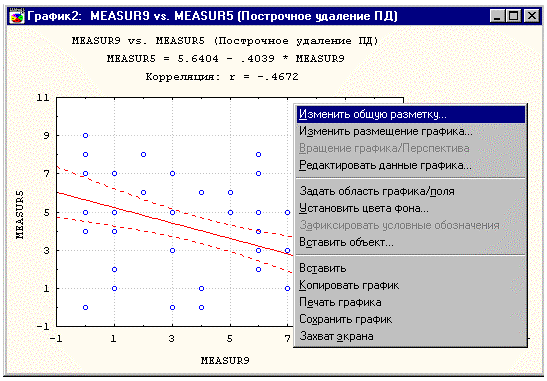
Теперь перейдите в подменю *Быстрые статистические графики* и выберите *Диаг. рассеяния/довер*.



Будет построен график с параметрами, заданными по умолчанию (диаграмма рассеяния для выбранного коэффициента корреляции с прямой регрессии, 95%-ая доверительная полоса и уравнение регрессии в заголовке).

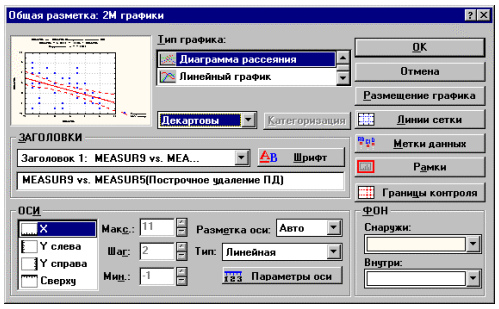


Настройка графика. Заметьте, что теперь, при активизации окна с графиком (см. выше), панель инструментов изменилась. Эта панель инструментов (которая сопровождает все окна с графиками) отличается от панелей инструментов таблиц исходных данных и таблиц результатов (которые похожи друг на друга). Она содержит множество настроек и инструментов для рисования. Все эти настройки также доступны из выпадающих меню (например, для того, чтобы записать их в макрос). Доступ к большинству из них также возможен из контекстных меню, появляющихся при щелчке мышью по соответствующей части графика. Если вы щелкнете где-нибудь на свободном месте снаружи осей графика, появится меню глобальных опций.

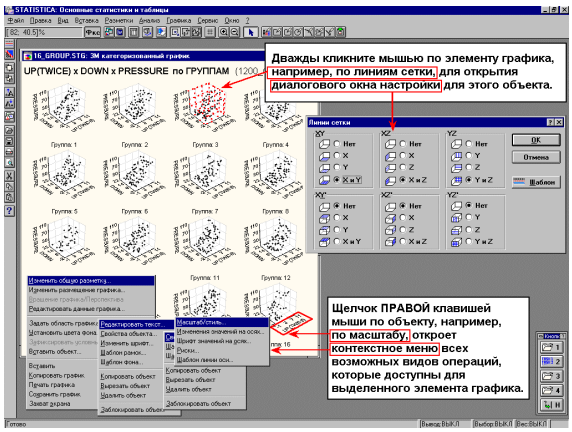


StatSoft®
Copyright © StatSoft, 1995

Большинство основных настроек формата графика доступно в диалоговом окне *Общая разметка* (см. выше первую опцию контекстного меню).

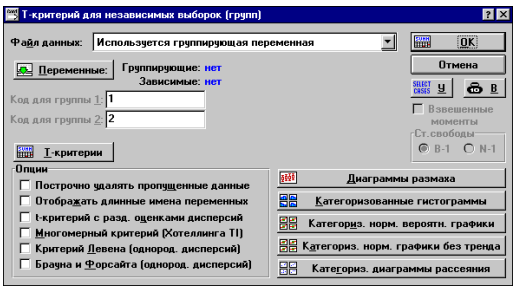


Однако большинство элементов графика может быть настроено выбором специальных установок в их индивидуальных контекстных меню, появляющихся при щелчке по элементу графика (например, по легенде, заголовке, значению шкалы и т.д.) правой кнопкой мыши (см. *Общие свойства и параметры графиков*, том II). Ниже проиллюстрированы основные соглашения по использованию мыши для настройки графиков.

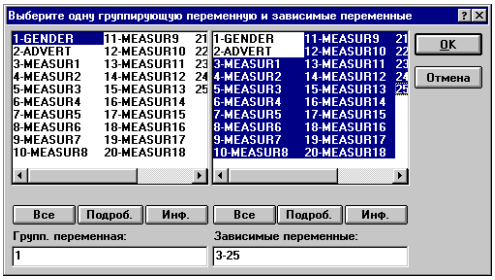


Разность между средними (*t*-критерий)

На следующем шаге анализа будет исследована возможность различия между моделью ответов мужчин и женщин. Подробнее, мужчины могут в некоторых шкалах давать более высокие или низкие оценки по сравнению с женщинами. *t*-критерий для независимых выборок может использоваться для идентификации такой возможной разницы. Выборки мужчин и женщин будут сравнены относительно среднего их оценок по каждой шкале. Вернитесь к стартовой панели *Основные статистики и таблицы* и щелкните на процедуре *t-критерий для независимых выборок*, чтобы открыть диалоговое окно *T-критерий для независимых выборок (групп)*.



Выбор независимых (группирующих) и зависимых переменных. Щелкните по кнопке *Переменные*, чтобы открыть стандартное диалоговое окно для выбора переменных. Здесь вы можете выбрать и независимые (группирующие), и зависимые переменные. Для нашего примера выберите (выделите) переменную *Пол - Gender* как независимую переменную и переменные от 3 до 25 (содержащие ответы) в качестве зависимых переменных.

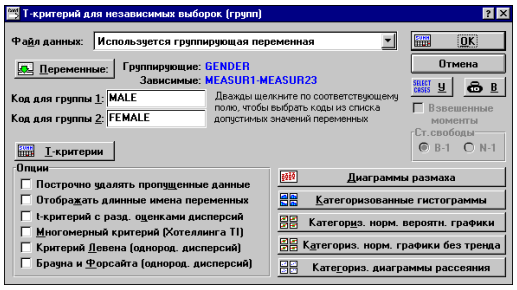


Щелкните *OK* в этом диалоговом окне, чтобы вернуться в диалоговое окно *T-критерий для*

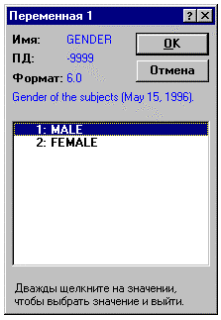
Copyright © StatSoft, 1995

StatSoft

независимых выборок (групп), где отобразится ваш выбор.



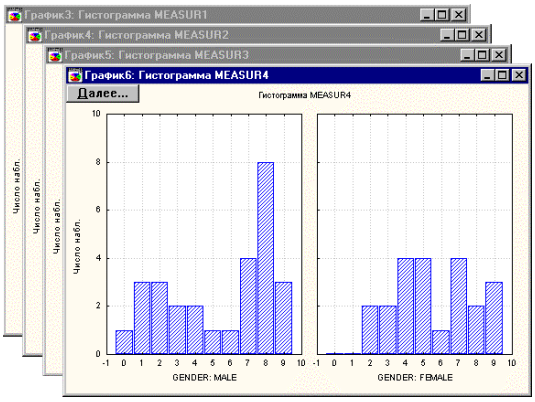
Поскольку вы выбрали группирующую переменную, *STATISTICA* автоматически предложит коды, используемые в этой переменной для идентификации групп, которые будут сравниваться (в нашем случае коды *МУЖЧИНА - MALE* и *ЖЕНЩИНА - FEMALE*). Вы можете дважды щелкнуть по каждому из этих полей ввода для вызова окна *Значения переменной* (см. *Окно таблицы исходных данных*, глава 3), где вы можете посмотреть и выбрать коды для каждой группы. Для быстрого выбора значения из списка (и выхода из диалогового окна), щелкните дважды на желаемом значении.



Из диалогового окна *Т-критерий для независимых выборок (групп)* доступно также множество других процедур (каждая из этих процедур обсуждается на стр. 1537). При помощи графических процедур этого диалогового окна вы можете изучать графики функций распределений переменных.

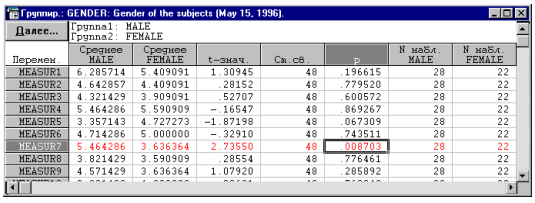
Например, щелкните по кнопке *Диаграммы размаха*, чтобы получить каскад категоризованных по группирующей переменной графиков, один график для каждой зависимой переменной (см. также *график t-критерия* ниже).

Аналогично щелкните по кнопке *Категоризованные гистограммы* для получения каскада категоризованных (по группирующей переменной) гистограмм (см. также, стр. 1492).



Чтобы посмотреть функции распределения для переменных из группы, также можно построить категоризованные нормальные вероятностные графики, нормальные вероятностные графики без тренда и диаграммы рассеяния.

Теперь щелкните *OK* для вывода таблицы результатов.



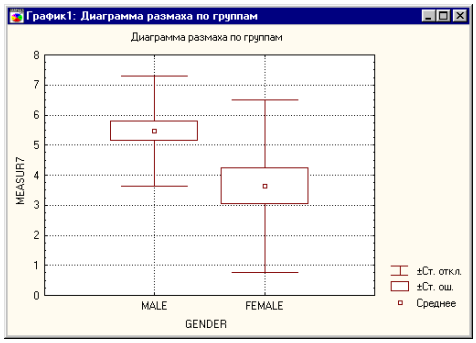
Перемен	Среднее MALE	Среднее FEMALE	t-знач.	Ст. св.	p	N набл. MALE	N набл. FEMALE
MEASUR1	6.285714	5.409091	1.30945	48	.196615	28	22
MEASUR2	4.642857	4.409091	.28152	48	.779520	28	22
MEASUR3	4.321429	3.909091	.52707	48	.600572	28	22
MEASUR4	5.464286	5.590909	-.16547	48	.869267	28	22
MEASUR5	3.357143	4.727273	-1.87198	48	.067309	28	22
MEASUR6	4.714286	5.000000	-.32910	48	.743511	28	22
MEASUR7	5.464286	3.636364	2.73550	48	.008703	28	22
MEASUR8	3.821429	3.590909	.28554	48	.776461	28	22
MEASUR9	4.571429	3.636364	1.07920	48	.285892	28	22

Просмотр результатов *t*-критерия. Самым быстрым способом изучения таблицы является просмотр пятого столбца (содержащего *p*-уровни) и определение того, какие из *p*-значений меньше установленного уровня значимости *.05* (см. *Элементарные понятия статистики*, глава

StatSoft®
Copyright © StatSoft, 1995

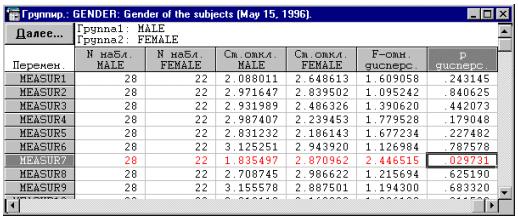
8). Для большинства зависимых переменных средние по двум группам (*МУЖЧИНЫ - MALES* и *ЖЕНЩИНЫ - FEMALES*) очень близки. Единственная переменная, для которой *t*-критерий соответствует установленному уровню значимости *.05*, - это *Measur7*, для нее *p-уровень* равен *.0087*. Как показывают столбцы, содержащие средние значения (см. две первые колонки) для мужчин, эта переменная принимает в среднем существенно большие значения – в выбранной шкале измерений для мужчин она равна *5.46*, а для женщин *3.63*. При этом нельзя исключить вероятность того, что полученная разница на самом деле отсутствует, и получилась лишь в результате случайного совпадения (см. ниже), хотя это выглядит маловероятным, поскольку такая разница в среднем может возникнуть только в 9 случаях из тысячи. Мы рассмотрим полученные результаты несколько позже, а сейчас построим и рассмотрим графики (выбранные по умолчанию) для наших данных.

График *t*-критерия. Графиком по умолчанию для этих таблиц результатов является *диаграмма размаха*. Для построения этой диаграммы щелкните правой кнопкой мыши в любом месте строки, соответствующей зависимой переменной (например, на среднем для *Measur7*). В открывшемся контекстном меню выберите постройку графика *Диаграмма размаха* из подменю *Быстрые статистические графики*. Далее, выберите опцию *Среднее/ст.ош./ст.откл.* окна *Диаграмма размаха* и нажмите *OK* для построения графика.



Разность средних на графике выглядит существенно больше и не может быть объяснена только на основании изменчивости исходных данных. Однако на графике стало заметным еще одно неожиданное отличие. Дисперсия для группы женщин намного больше дисперсии для группы мужчин (посмотрите на прямоугольники). Если дисперсии в двух группах существенно отличаются, то нарушается одно из требований для использования *t*-критерия (см. *Вводный обзор*, стр. 1458), и разность средних должна рассматриваться особенно внимательно. Кроме того, дисперсия обычно коррелирована со средним значением, т.е. чем больше среднее, тем больше дисперсия. Однако в данном случае наблюдается нечто противоположное. В такой ситуации опытный исследователь предположил бы, что распределение переменной *Measur7*, возможно, не является нормальным (для мужчин, женщин, или для тех и других). Однако сначала рассмотрим критерий разности дисперсий для того, чтобы проверить, является ли наблюдаемое на графике отличие действительно заслуживающим внимания.

Критерий разности дисперсий. Вернемся к таблице результатов и прокрутим ее вправо до тех пор, пока не станут видны результаты *F*-критерия. Значение *F*-критерия действительно соответствует указанному уровню значимости *.05*, что означает существенную разность дисперсий переменной *Measur7* в группах *МУЖЧИНЫ - MALES* и *ЖЕНЩИНЫ - FEMALES*. Однако, значимость наблюдаемой разности дисперсий достаточно близка к граничному уровню значимости (ее *p-уровень* равен *.029*).



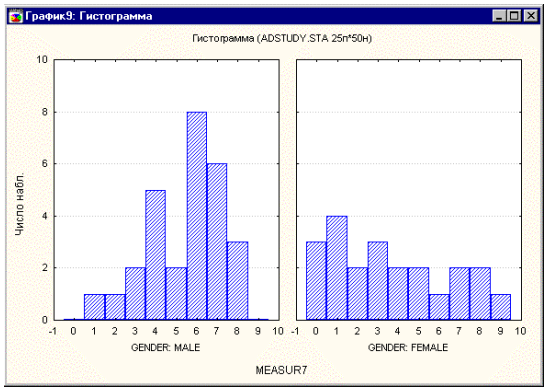
Перемен	N набл. MALE	N набл. FEMALE	Ст. откл. MALE	Ст. откл. FEMALE	F-отн. дисперс	p дисперс
MEASUR1	28	22	2.088011	2.648613	1.609058	.243145
MEASUR2	28	22	2.971647	2.839502	1.095242	.840625
MEASUR3	28	22	2.931989	2.486326	1.390620	.442073
MEASUR4	28	22	2.987407	2.239453	1.779528	.179048
MEASUR5	28	22	2.831232	2.186143	1.677234	.227482
MEASUR6	28	22	3.125251	2.943920	1.126984	.787578
MEASUR7	28	22	1.835497	2.870962	2.446515	.029731
MEASUR8	28	22	2.708745	2.986622	1.215694	.625190
MEASUR9	28	22	3.155578	2.887501	1.194300	.683320

Copyright © StatSoft, 1995

Большинство исследователей посчитало бы один этот факт не достаточным для признания недействительным *t*-критерия разности средних, дающего высокий уровень значимости для этой разности ($p = .0087$). Далее мы рассмотрим распределение переменной *Measur7*, категоризованной по независимой переменной *Пол - Gender*.

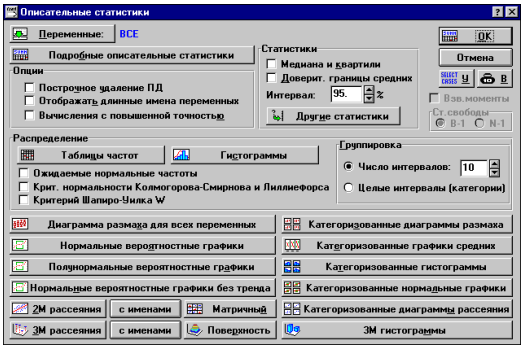
Категоризованная гистограмма. Щелкните правой кнопкой мышки по таблице результатов и выберите в открывшемся контекстном меню постройку графика *2М гистограмма по Gender* из подменю *Быстрые статистические графики*. В открывшемся после этого диалоговом окне можно выбрать коды для группирующей переменной.

В нашем примере нажмите *OK*, чтобы автоматически принять два кода для разделения респондентов на мужчин и женщин и построения категоризованной гистограммы.



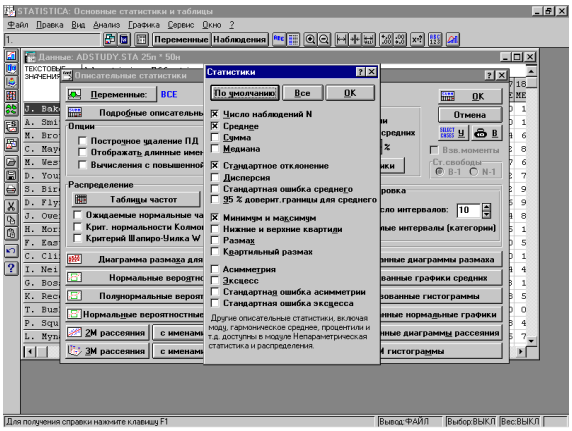
Анализ распределения (описательные статистики)

Теперь вернитесь в стартовую панель модуля *Основные статистики и таблицы* и нажмите кнопку *Описательные статистики*, чтобы открыть соответствующее окно. В этом окне нажмите кнопку *Переменные* и выберите все переменные, содержащиеся в файле данных. После нажатия *OK* в окне выбора переменных, диалоговое окно *Описательные статистики* будет выглядеть следующим образом:



Это окно позволяет получать таблицы результатов с разнообразными описательными статистиками, а также строить таблицы частот. Еще здесь можно построить большое количество графиков для проведения визуального анализа распределения выбранных переменных.

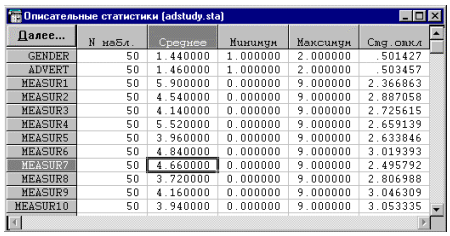
По умолчанию таблицы результатов из окна *Описательные статистики* содержат средние значения, число наблюдений *N*, стандартные отклонения, а также минимальные и максимальные значения для выбранных переменных. Нажмите кнопку *Другие статистики*, чтобы открыть окно *Статистики*, в котором можно выбрать типы вычисляемых статистик (см. стр. 1531).



StatSoft®

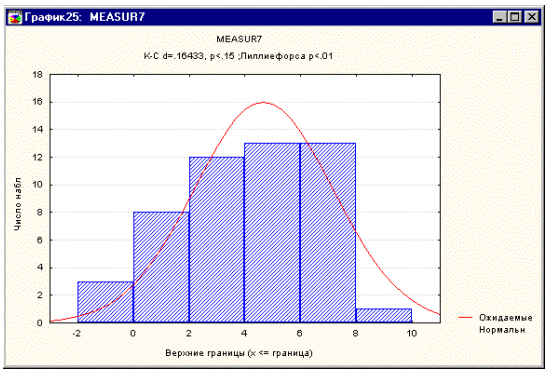
Copyright © StatSoft, 1995

В нашем примере оставьте выбор статистик, сделанный по умолчанию, и нажмите кнопку *Подробные описательные статистики* для построения таблицы результатов.

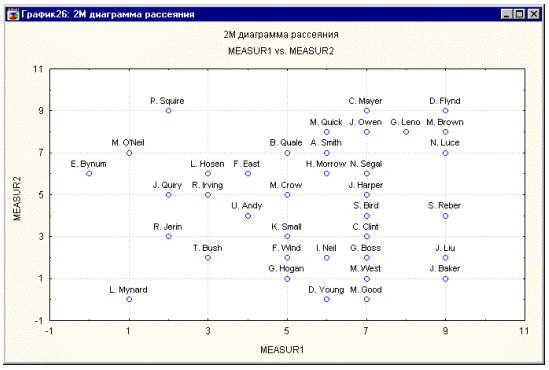


Графиком по умолчанию для этой таблицы результатов является гистограмма с наложенной на нее нормальной кривой. Этот график обычно используется для визуальной проверки нормальности исходных данных. Для построения графика щелкните правой кнопкой мыши в любом месте таблицы результатов (например, на среднем значении переменной *Measur7*), и в появившемся контекстном меню выберите построение графика *Гистограмма/нормальн* из подменю *Быстрые статистические графики*.

Такая же гистограмма может быть построена нажатием кнопки *Гистограммы* в разделе *Распределение* окна *Описательные статистики*. Этот раздел также позволяет анализировать распределение частот для каждой выбранной переменной (при этом происходит построение по одной таблице результатов или гистограмме на каждую переменную). В этом окне также возможно вычисление некоторых специальных критериев нормальности и использование настроек категоризации изучаемых данных (см. также стр. 1522).

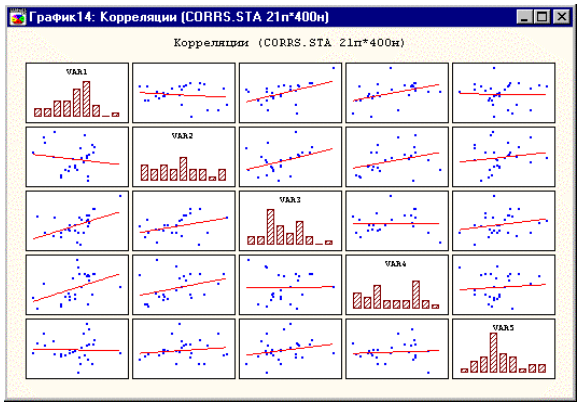


Графические процедуры. Окно *Описательные статистики* предлагает большое количество графических процедур для проведения визуального анализа распределений переменных, а также корреляций между ними. Например, нажмите кнопку *2М рассеяния (с именами)*, чтобы получить наглядное представление о характере зависимости между двумя переменными.

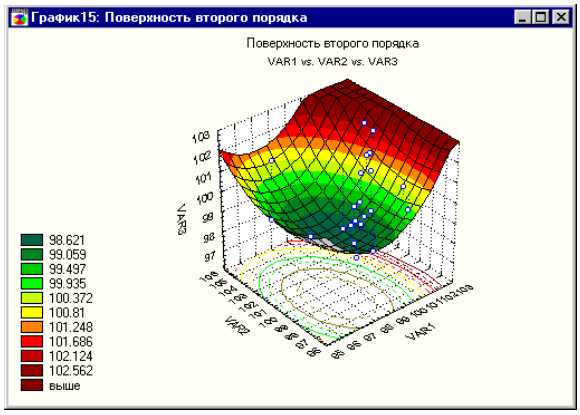


При использовании опции *с именами* программа располагает на диаграмме имена наблюдений рядом с соответствующими им точками. Также вы можете построить матрицу диаграмм рассеяния, нажав кнопку *Матричный*.

Copyright © StatSoft, 1995

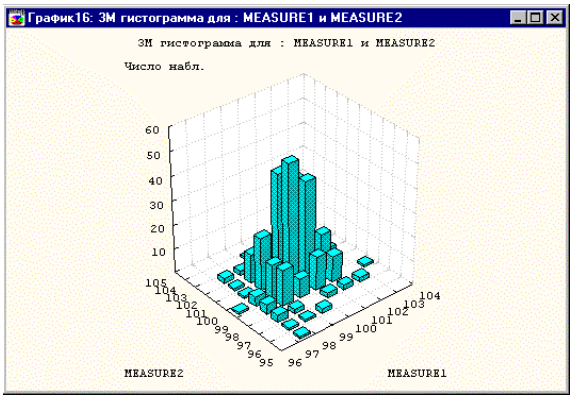


Кнопка *Поверхность* позволяет строить поверхность в пространстве (по умолчанию второго порядка), приближающую значения выбранных переменных.



Также возможно построение категоризованных диаграмм размаха, гистограмм, диаграмм рассеяния и вероятностных графиков.

Наконец, возможно построение *3М гистограммы двух переменных* для изучения двумерного распределения выбранных переменных.



Этот график обычно используется для описательных (т.е. не аналитических) целей, а также при проведении разведочного анализа данных; однако, иногда он может быть полезен при проверке нормальности распределения.

Пример 2: Группировка и однофакторный дисперсионный анализ

Обзор

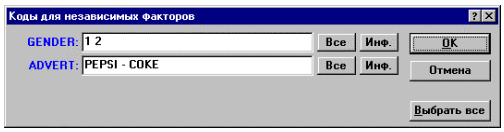
Вы можете легко вычислить разнообразные описательные статистики (например, средние, стандартные отклонения, корреляции, и т.д.) для данных, разбитых на группы одной или несколькими группирующими переменными (например, переменными *Пол - Gender* и *Реклама - Region*). Можно также выполнить однофакторный дисперсионный анализ с помощью процедуры *Группировка и однофакторная ANOVA*, доступной из стартовой панели модуля *Основные статистики и таблицы*. Будем использовать тот же файл данных *Adstudy.sta*. После выбора процедуры *Группировка и однофакторная ANOVA* в стартовой панели, нажмите кнопку *Переменные* в диалоговом окне *Внутригрупповые описательные статистики и корреляции*, и выберите две группирующие переменные *ПОЛ - GENDER* (*МУЖЧИНА - MALE* и *ЖЕНЩИНА -*

StatSoft

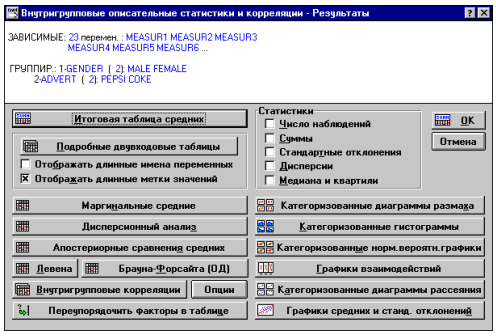
Copyright © StatSoft, 1995

FEMALE) и *ПРЕДПОЧТЕНИЕ - ADVERT* (тип предпочтения - какой ролик предпочли: *COKE* или *PEPSI*).

Щелкните по кнопке *Коды для группирующих переменных* и выберите коды для группирующих переменных в диалоговом окне *Коды для независимых факторов.*



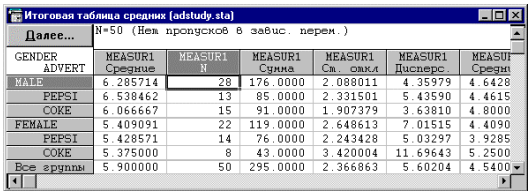
Чтобы выбрать все коды переменной, можно либо ввести номера кодов в соответствующем поле ввода, либо нажать кнопку *Все*, - либо поставить * в соответствующем поле ввода. Щелкнув по кнопке *Выбрать все* в этом диалоговом окне, вы выберете все коды для каждой переменной. Нажатие *OK* без задания каких-либо значений эквивалентно определению всех значений для всех переменных. Нажмите *OK* здесь и в диалоговом окне *Внутригрупповые описательные статистики и корреляции* для того, чтобы открыть диалоговое окно *Внутригрупповые описательные статистики и корреляции - Результаты*.



Диалог *Внутригрупповые описательные статистики и корреляции* предоставляет различные процедуры и настройки для внутригруппового анализа данных (анализ данных внутри групп). Цель такого анализа – лучшее понимание различий между группами.

Итоговая таблица средних

Вы можете выбрать нужные статистики для того, чтобы отобразить их на экране в *Итоговой таблице средних* или *Подробных двувходовых таблицах*. В этом примере выберите все пять возможных статистик. В настройках области *Подробные двувходовые таблицы* можно также выбрать длинные имена переменных и метки, если они имеются в файле данных. Затем щелкните по кнопке *Подробные двувходовые таблицы,* чтобы увидеть таблицу результатов.



GENDER ADVERT	MEASUR1 Средние	MEASUR1 N	MEASUR1 Сумма	MEASUR1 Ст. откл	MEASUR1 Дисперс.	MEASUR1 Средн
MALE	6.285714	28	176.0000	2.088011	4.35979	4.6428
PEPSI	6.538462	13	85.0000	2.331501	5.43590	4.4615
COKE	6.066667	15	91.0000	1.907379	3.63810	4.8000
FEMALE	5.409091	22	119.0000	2.648613	7.01515	4.4090
PEPSI	5.428571	14	76.0000	2.243428	5.03297	3.9285
COKE	5.375000	8	43.0000	3.420004	11.69643	5.2500
Все группы	5.900000	50	295.0000	2.366863	5.60204	4.5400

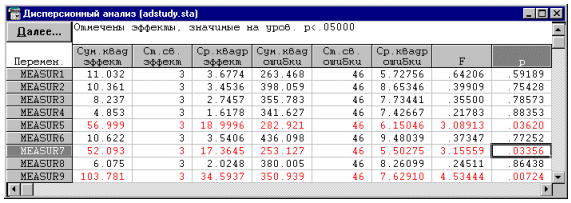
В показанной выше таблице результатов приведены описательные статистики для переменных, разбитых на группы (прокрутите таблицу, чтобы увидеть результаты для остальных переменных). Например, рассматривая средние внутри каждой группы, вы можете заметить, что имеется слабое различие в средних переменной *Measur1* между группами *МУЖЧИНЫ - MALES* и *ЖЕНЩИНЫ - FEMALES*. Теперь исследуем, как различаются средние переменной *Measur1* внутри групп *МУЖЧИНА - MALE* и *ЖЕНЩИНА - FEMALE*. Как вы можете заметить, имеется очень малое различие между группами *PEPSI* и *COKE* в пределах одного пола; таким образом, группы, получающиеся разделением по полу, кажутся достаточно однородными.

Copyright © StatSoft, 1995

StatSoft®

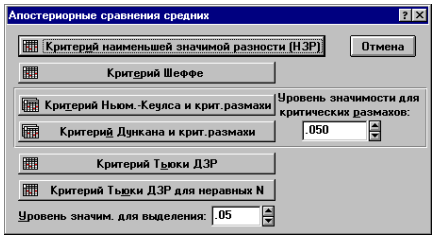
Однофакторный дисперсионный анализ и *апостериорные* сравнения средних

Можно легко проверить значимость этих различий в средних с помощью кнопки *Дисперсионный анализ*. Щелкните на эту кнопку и вы увидите результаты однофакторного дисперсионного анализа для каждой зависимой переменной.

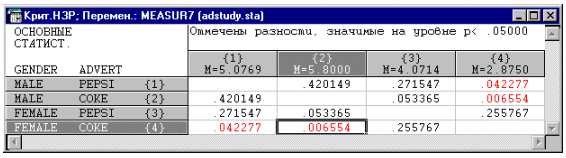


Вы видите, что для переменных *Measur5*, *Measur7*, и *Measur9* процедура однофакторного *Дисперсионного анализа* дала статистически значимые результаты на уровне *p<.05*. Эти результаты показывают, что различие в средних значимо. Возвратитесь в диалоговое окно результатов и нажмите кнопку *Апостериорные сравнения средних* для того, чтобы оценить значимость различий между средними конкретных групп. Прежде всего, нужно выбрать переменную (переменные) для сравнений. В этом примере выберите переменную *Measur7.*

После того, как вы нажмете *OK* в окне выбора переменной, на экране появится диалоговое окно *Апостериорные сравнения средних*.

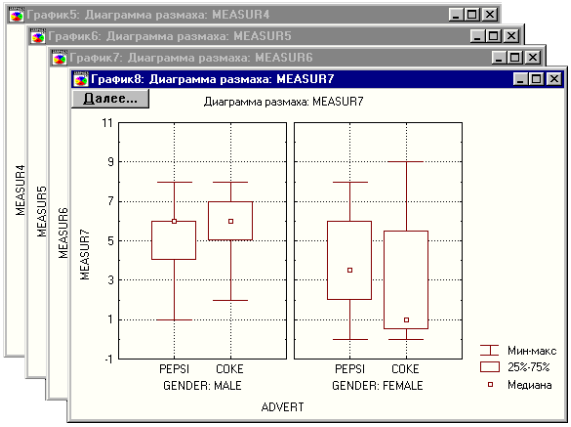


В этом окне можно выбрать несколько *апостериорных* критериев (см. обсуждение этих критериев на стр.1551). В данном примере выберем *Критерий наименьшей значимой разности (НЗР)*.



Критерий *НЗР* эквивалентен *t*-критерию для независимых выборок, основанному на *N* сравниваемых группах. Ранее *t*-критерий для независимых выборок (см. *Пример 1,* стр. 1490) показал, что имеется значимое различие между ответами *МУЖЧИН - MALES* и ответами *ЖЕНЩИН - FEMALES* для переменной *Measur7*. Используя процедуру *Группировка и однофакторная ANOVA*, мы видим (см. таблицу результатов), что значимое различие имеется только для респондентов, выбравших *COKE*.

Графическое представление результатов. Эти различия можно увидеть графически с помощью большого числа графических опций в диалоговом окне *Внутригрупповые описательные статистики и корреляции - Результаты*.



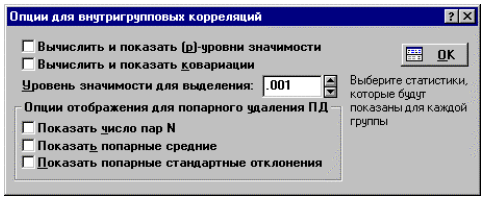
Например, для того, чтобы сравнить распределения выбранных переменных внутри групп, щелкните на кнопку *Категоризованные диаграммы размаха* и выберите опцию *Медиана/кварт./размах* из диалогового окна

StatSoft®
Copyright © StatSoft, 1995

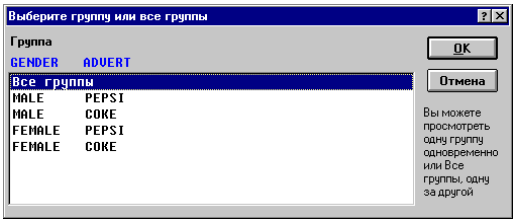
Диаграмма размаха. Когда вы нажмете *OK* в этом диалоговом окне, программа построит каскад диаграмм размаха. Как вы видите на графике, для переменной *Measur7* в распределении значений для группы *ЖЕНЩИНА - COKE - FEMALE - COKE* и для группы *МУЖЧИНА - COKE - MALE - COKE* имеются различия.

Внутригрупповые корреляции

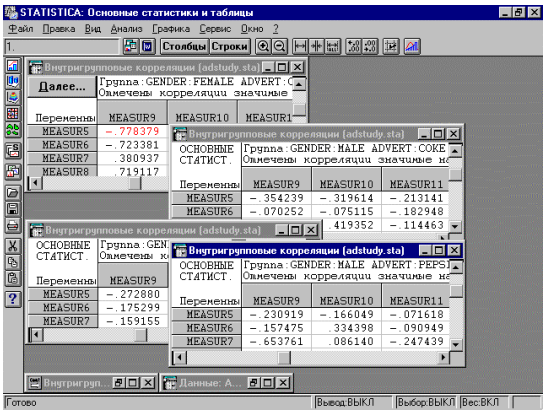
Теперь рассмотрим корреляции между переменными внутри групп. Вернитесь к диалоговому окну *Внутригрупповые описательные статистики и корреляции - Результаты* и нажмите кнопку *Внутригрупповые корреляции*. Заметьте, что в этом диалоговом окне имеются опции различного представления корреляционных матриц на экране. Нажмите кнопку *Опции,* расположенную рядом с кнопкой *Внутригрупповые корреляции* (см. стр. 1548).



В этом примере измените параметр *Уровень значимости для выделения.* Задайте его равным *.001* и щелкните *OK,* чтобы вернуться в диалоговое окно *Внутригрупповые описательные статистики и корреляции - Результаты*. Нажмите кнопку *Внутригрупповые корреляции* и откройте диалоговое окно *Выберите группу или все группы*, в котором можно выбрать одну группу (или *Все группы*) для корреляционных матриц.



В *Примере 1* (см. стр. 1485) была построена корреляционная матрица, в которой корреляция между переменными *Measur5* и *Measur9* (*r = -.47*) была высоко значимой (*p<.001*). Процедура *Группировка и однофакторная ANOVA* позволит вам провести дальнейшее исследование этой значимой корреляции с помощью вычисления корреляции *внутри* групп. В диалоговом окне *Выберите группу или все группы* дважды щелкните на строке *Все группы,* чтобы получить следующие 4 корреляционные матрицы:



Как видите, корреляции в отдельных группах заметно отличаются (например, корреляция очень высокая в группе *ЖЕНЩИНА/COKE - FEMALE/COKE* и намного ниже в остальных трех группах). Нигде корреляция между переменными *Measur5* и *Measur9* не является значимой на уровне *.001*; однако если вы измените *альфа* уровень на *.05* с помощью кнопки *Опции* диалогового окна *Внутригрупповые описательные статистики и корреляции - Результаты* и снова нажмете

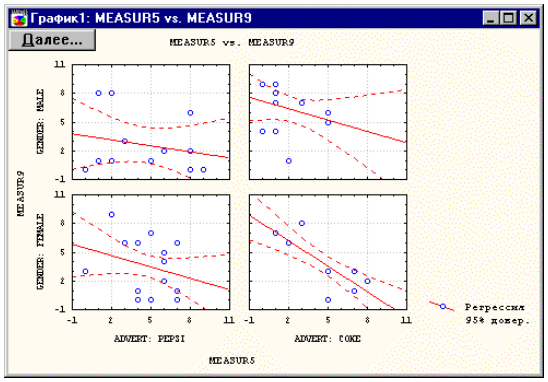
Copyright © StatSoft, 1995

StatSoft®

кнопку *Внутригрупповые Корреляции*, то увидите, что корреляция между *Measur5* и *Measur9* значима на этом уровне (*p=.02*) для группы *ЖЕНЩИНА/COKE - FEMALE/COKE*. Никакие другие корреляции не значимы на уровне *.05*.

Заметим, что можно использовать *Другие критерии значимости* (см. стр. 1567), чтобы проверить различие между коэффициентами корреляции.

Категоризованные диаграммы рассеяния. Внутригрупповые корреляции можно представить графически, используя команду *Категоризованные диаграммы рассеяния* в диалоговом окне *Внутригрупповые описательные статистики и корреляции - Результаты*. Нажав эту кнопку, вы сможете выбрать переменные для графиков. Выберите *Measur5* в первом списке и *Measur9* во втором. Далее нажмите *OK*, чтобы построить график.

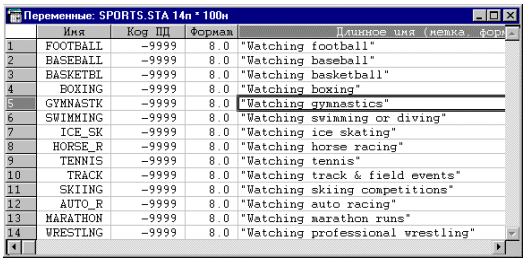


Из данного графика отчетливо видна сильная отрицательная корреляция между переменными *Measur5* и *Measur9* для группы *ЖЕНЩИНА/COKE - FEMALE/COKE*.

Пример 3: Таблицы частот

Обзор

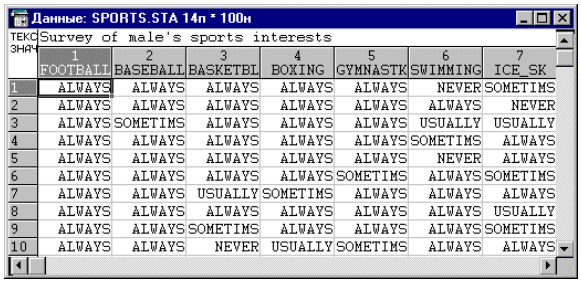
Следующий пример основан на (гипотетических) данных опроса общественного мнения. Предположим, вы владеете несколькими “спортивными барами” в своем городе. В каждом имеется два или три больших телевизора, показывающие разные спортивные соревнования. Так как вы не можете показать все соревнования одновременно (они идут по разным каналом ТВ), вам хотелось бы знать, какие виды спорта наиболее интересные для посетителей. Проводился опрос 100 покупателей (в различные моменты времени в разных барах) относительно их интереса к разным видам спорта. Каждый респондент получил список из 14 видов спорта с просьбой определить свой интерес к ним по 4-х точечной шкале: (1) *Всегда интересуюсь - Always interested,* (2) *Обычно интересуюсь - Usually interested*, (3) *Иногда интересуюсь - Sometimes interested* и (4) *Никогда не интересуюсь - Never interested.* Ниже показаны спецификации всех 14-ти переменных таблицы (для этого щелкните по кнопке *Переменные* и выберите *Все спецификации*).



	Имя	Код ПД	Формат	Длинное имя (метка, фор
1	FOOTBALL	-9999	8.0	"Watching football"
2	BASEBALL	-9999	8.0	"Watching baseball"
3	BASKETBL	-9999	8.0	"Watching basketball"
4	BOXING	-9999	8.0	"Watching boxing"
5	GYMNASTK	-9999	8.0	"Watching gymnastics"
6	SWIMMING	-9999	8.0	"Watching swimming or diving"
7	ICE_SK	-9999	8.0	"Watching ice skating"
8	HORSE_R	-9999	8.0	"Watching horse racing"
9	TENNIS	-9999	8.0	"Watching tennis"
10	TRACK	-9999	8.0	"Watching track & field events"
11	SKIING	-9999	8.0	"Watching skiing competitions"
12	AUTO_R	-9999	8.0	"Watching auto racing"
13	MARATHON	-9999	8.0	"Watching marathon runs"
14	WRESTLNG	-9999	8.0	"Watching professional wrestling"

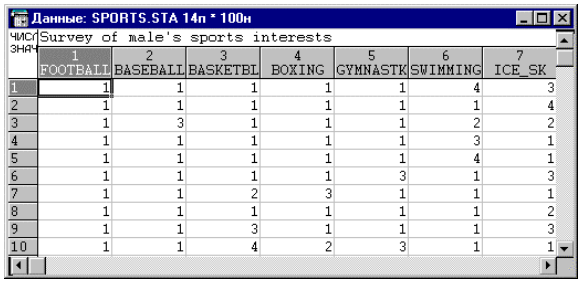
Файл данных. Данные опроса записаны в файл *Sports.sta*. Каждая переменная в файле соответствует определенному виду спорта (см. выше), строка – ответам одного респондента. Для каждого вида спорта выбор респондентов отмечался определенной позицией на 4-х точечной шкале.

StatSoft
Copyright © StatSoft, 1995



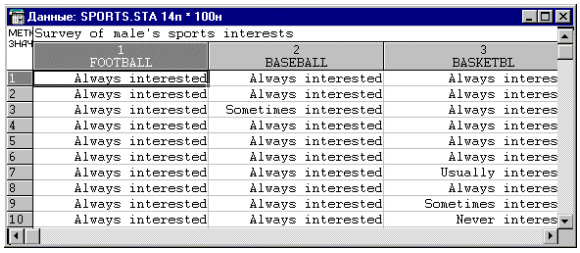
	1 FOOTBALL	2 BASEBALL	3 BASKETBL	4 BOXING	5 GYMNASTK	6 SWIMMING	7 ICE_SK
1	ALWAYS	ALWAYS	ALWAYS	ALWAYS	ALWAYS	NEVER	SOMETIMS
2	ALWAYS	ALWAYS	ALWAYS	ALWAYS	ALWAYS	ALWAYS	NEVER
3	ALWAYS	SOMETIMS	ALWAYS	ALWAYS	ALWAYS	USUALLY	USUALLY
4	ALWAYS	ALWAYS	ALWAYS	ALWAYS	ALWAYS	SOMETIMS	ALWAYS
5	ALWAYS	ALWAYS	ALWAYS	ALWAYS	ALWAYS	NEVER	ALWAYS
6	ALWAYS	ALWAYS	ALWAYS	ALWAYS	SOMETIMS	ALWAYS	SOMETIMS
7	ALWAYS	ALWAYS	USUALLY	SOMETIMS	ALWAYS	ALWAYS	ALWAYS
8	ALWAYS	ALWAYS	ALWAYS	ALWAYS	ALWAYS	ALWAYS	USUALLY
9	ALWAYS	ALWAYS	SOMETIMS	ALWAYS	ALWAYS	ALWAYS	SOMETIMS
10	ALWAYS	ALWAYS	NEVER	USUALLY	SOMETIMS	ALWAYS	ALWAYS

Можно щелкнуть по кнопке *Отображение числовых/текстовых значений* панели инструментов таблицы исходных данных, чтобы переключиться в численное представление значений переменных в таблице.



	1 FOOTBALL	2 BASEBALL	3 BASKETBL	4 BOXING	5 GYMNASTK	6 SWIMMING	7 ICE_SK
1	1	1	1	1	1	4	3
2	1	1	1	1	1	1	4
3	1	3	1	1	1	2	2
4	1	1	1	1	1	3	1
5	1	1	1	1	1	4	1
6	1	1	1	1	3	1	3
7	1	1	2	3	1	1	1
8	1	1	1	1	1	1	2
9	1	1	3	1	1	1	3
10	1	1	4	2	3	1	1

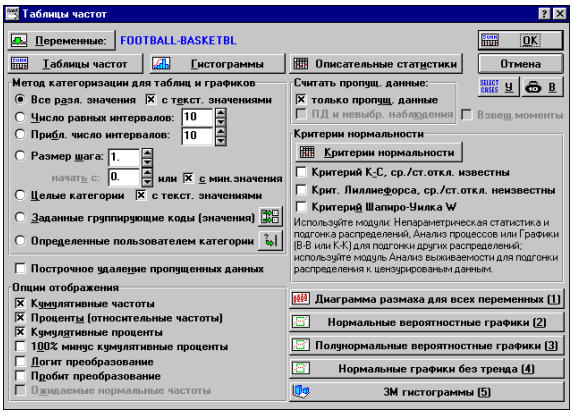
Заметим, что вы можете включить режим просмотра меток (в частности, увидеть на экране метки, записи или комментарии, относящиеся к определенным значениям). Это достигается нажатием на кнопку *Отображение числовых/текстовых значений* при одновременном нажатии клавиши CTRL.



	1 FOOTBALL	2 BASEBALL	3 BASKETBL
1	Always interested	Always interested	Always interes
2	Always interested	Always interested	Always interes
3	Always interested	Sometimes interested	Always interes
4	Always interested	Always interested	Always interes
5	Always interested	Always interested	Always interes
6	Always interested	Always interested	Always interes
7	Always interested	Always interested	Usually interes
8	Always interested	Always interested	Always interes
9	Always interested	Always interested	Sometimes interes
10	Always interested	Always interested	Never interes

Таблицы частот

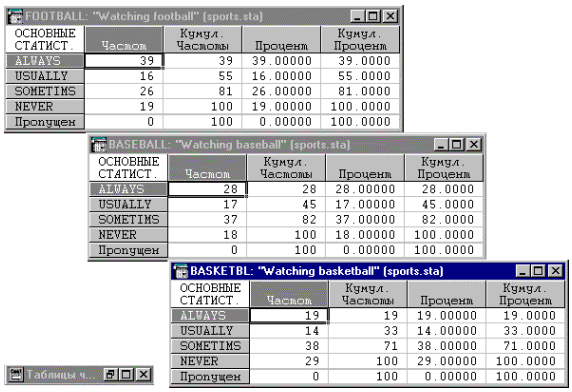
Прежде всего, посмотрим на респондентов, интересующихся самыми популярными видами спорта: футболом, бейсболом, баскетболом. Из стартовой панели *Основные статистики и таблицы* выберите процедуру *Таблицы частот,* чтобы открыть диалоговое окно *Таблицы частот*. В этом окне щелкните по кнопке *Переменные* и выберите первые три переменные (*Футбол - Football*, *Бейсбол - Baseball*, *Баскетбол - Basketball*). Диалоговое окно *Таблицы частот* появится на экране в следующем виде:



Это диалоговое окно предлагает множество настроек, позволяющих изменять вид и группировку в таблицах частот, а также проверять нормальность распределения, в том числе и графическими способами. Эти настройки обсуждаются на стр. 1553. В этом примере используем принятый по умолчанию метод группировки (в частности, *Все различные значения, с текстовыми значениями*) и опции отображения (*Кумулятивные частоты*, *Проценты (относительные частоты)*, *Кумулятивные проценты*), как показано в диалоговом окне выше. Теперь нажмите на кнопку *Таблицы частот,* чтобы увидеть каскад таблиц результатов, одну для каждой выбранной переменной.

Copyright © StatSoft, 1995

StatSoft®



FOOTBALL: "Watching football" (sports.sta)

ОСНОВНЫЕ СТАТИСТ.	Частот.	Кумул. Частоты	Процент	Кумул. Процент
ALWAYS	39	39	39.00000	39.0000
USUALLY	16	55	16.00000	55.0000
SOMETIMS	26	81	26.00000	81.0000
NEVER	19	100	19.00000	100.0000
Пропущен	0	100	0.00000	100.0000

BASEBALL: "Watching baseball" (sports.sta)

ОСНОВНЫЕ СТАТИСТ.	Частот.	Кумул. Частоты	Процент	Кумул. Процент
ALWAYS	28	28	28.00000	28.0000
USUALLY	17	45	17.00000	45.0000
SOMETIMS	37	82	37.00000	82.0000
NEVER	18	100	18.00000	100.0000
Пропущен	0	100	0.00000	100.0000

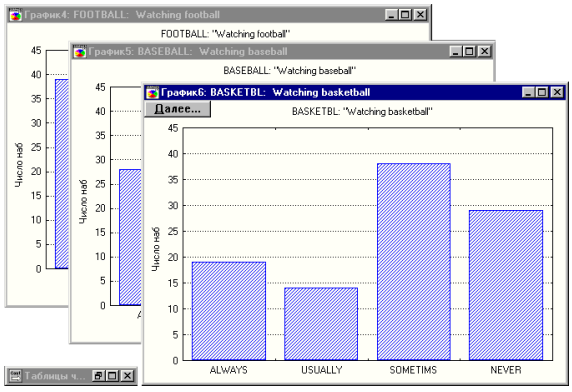
BASKETBL: "Watching basketball" (sports.sta)

ОСНОВНЫЕ СТАТИСТ.	Частот.	Кумул. Частоты	Процент	Кумул. Процент
ALWAYS	19	19	19.00000	19.0000
USUALLY	14	33	14.00000	33.0000
SOMETIMS	38	71	38.00000	71.0000
NEVER	29	100	29.00000	100.0000
Пропущен	0	100	0.00000	100.0000

Таблицы ч...

Как вы видите переменной *Футбол - Football*, 39% респондентов отметили, что они *всегда интересуются - always interested* футболом, 16% - *обычно интересуются - usually interested* и т.д. Всего 81% респондентов попали в категории *всегда - always*, *обычно - usually*, *иногда интересуюсь - sometimes interested* и только 19% сказали - *никогда не интересуюсь - never interested*.

Построение гистограмм. Заметим, что можно легко построить гистограммы всех выбранных переменных, если вернуться обратно в диалоговое окно *Таблицы частот* и нажать кнопку *Гистограммы*. Каскад гистограмм, по одной гистограмме для каждой выбранной переменной, мгновенно появится на экране.



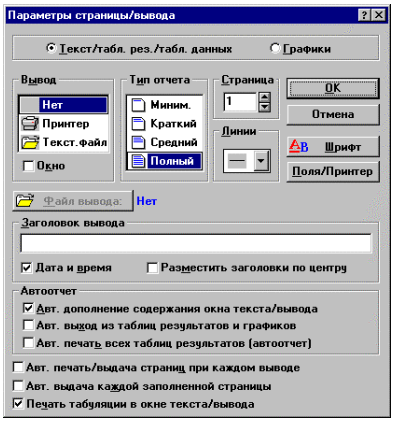
Посмотрите таблицу результатов для второй переменной (*Бейсбол - Baseball*). Ее можно выбрать из выпадающего меню *Окно* для того, чтобы поместить на передний план окна.

BASEBALL: "Watching baseball" (sports.sta)

ОСНОВНЫЕ СТАТИСТ.	Частот.	Кумул. Частоты	Процент	Кумул. Процент
ALWAYS	28	28	28.00000	28.0000
USUALLY	17	45	17.00000	45.0000
SOMETIMS	37	82	37.00000	82.0000
NEVER	18	100	18.00000	100.0000
Пропущен	0	100	0.00000	100.0000

Из таблицы вы видите, что большинство респондентов (37%) только иногда интересуются бейсболом по ТВ. Однако, 28% респондентов сказали, что всегда интересуются.

Печать таблиц в пакетном режиме. В системе *STATISTICA* вы можете распечатать (или сохранить в файле) результаты анализа либо автоматически (когда содержимое каждой выводимой на экран таблицы результатов одновременно направляется на принтер и/или в *Окно текста/вывода*), либо вручную (когда пользователь сам выбирает какую таблицу результатов или часть какой таблицы результатов распечатать). Перед тем, как распечатать результаты анализа, программа попросит вас уточнить направление вывода (т.е. *Текст. файл*, *Принтер*, *Нет*, и/или *Окно*) в окне *Параметры страницы/вывода* (выберите установку *Параметры страницы/вывода* в выпадающем меню *Файл*, настройку *Принтер* в выпадающем меню *Сервис* или дважды щелкните на поле *Вывод* строки состояния; описание этого диалогового окна см. в главе 3).

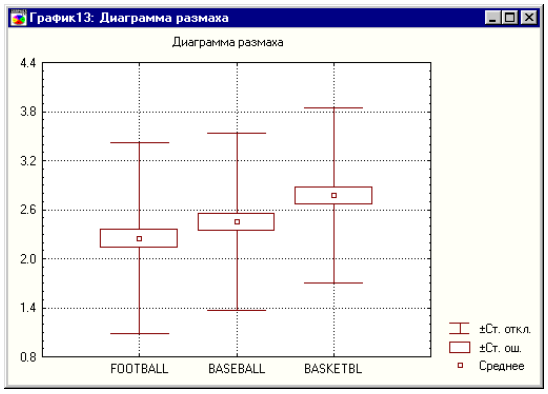


StatSoft®

Copyright © StatSoft, 1995

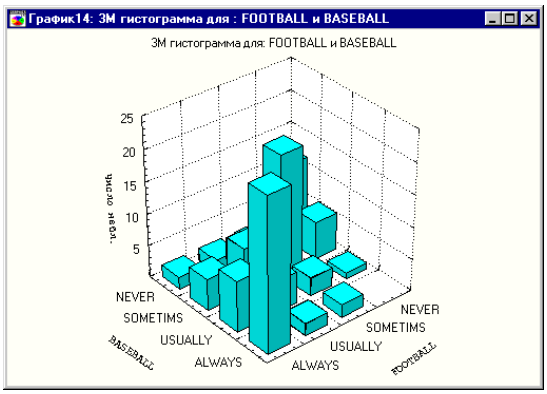
В этом окне можно также определить дополнительную информацию для печати вместе с таблицей результатов. Доступны следующие формы выводимого отчета: *Минимальный, Краткий, Средний или Полный*. О каждой из этих четырех форм отчета рассказывается в главе 3 (в описании диалогового окна *Параметры страницы/вывода*). Если в окне *Параметры страницы/вывода* была выбрана настройка *Авт. печать всех таблиц результатов (автоотчет)*, то дополнительная информация (количество которой определяется установленным в этом же окне форматом отчета), а также все результаты анализа будут автоматически выведены на принтер или в файл (в зависимости от того, выбрана ли установка *Окно* в левой верхней части этого диалогового окна). Этот режим печати полезен, если вы хотите получить полную сводку всех результатов, выведенных на экран в процессе анализа.

Графические процедуры. Практически все эти результаты могут быть отображены на графиках с помощью графических процедур, доступных в данном окне. Прежде всего, щелкните по кнопке *Диаграмма размаха для всех переменных* и в появившемся диалоговом окне выберите *Средние/ст.ош./ст.откл.* и затем нажмите *OK,* чтобы построить график.



Разброс ответов для каждой переменной кажется очень похожим, однако, как показывает меньшее значение среднего переменной *Футбол - Football*, имеется несколько больший интерес к футболу (напомним, *1* показывает максимальный интерес, *4* – отсутствие интереса). Это также можно увидеть, сравнивая таблицы частот этих переменных.

Теперь посмотрим на *3М гистограмму*. Когда вы нажмете на соответствующую кнопку, будет предложено выбрать переменные для графика. Выберите, например, переменные *Футбол - Football* и *Бейсбол - Baseball* и нажмите *OK*. (Заметим, если вы выбрали более одной пары переменных, будет построен каскад 3М гистограмм). Визуальный анализ 3М гистограммы (показанной ниже) дает возможность говорить о существовании корреляции между переменными *Футбол - Football* и *Бейсбол - Baseball*. В самом деле, если вы выполните корреляционный анализ, то увидите, что корреляция Спирмена между этими переменными равна *.6785*. Это означает, что субъекты, предпочитающие смотреть футбол, также желают смотреть и бейсбол.



Печать графиков в пакетном режиме. Если в диалоговом окне *Параметры страницы/вывода* выбрана установка *Автоматически печатать все графики* (см. глава 3), система *STATISTICA* автоматически направит создаваемые графики или на печать, или в окно вывода (или сохранит в файле

Copyright © StatSoft, 1995

StatSoft

вывода, если выбрана *Печать в файл* в диалоговом окне *Печать графика*, см. том II).

Итог. Таким образом, из анализа вы узнали, что футбол и бейсбол являются самыми привлекательными видами спорта (для посетителей бара). Следующий интересный вопрос – найти пересечения между различными видами спорта. Если посетители вашего бара состоят из определенного процента спортивных “фанатов”, желающих смотреть только футбол, бейсбол или баскетбол, то, возможно, следует выбрать только одну эту игру для показа и зарезервировать другие телеэкраны, например, для тенниса или марафона. Однако прежде всего следует понять, насколько велико пересечение между видами спорта, и какие виды спорта будут привлекать аудиторию, состоящую не только из фанатиков футбола.

Пример 4: Таблицы флагов и заголовков

Обзор

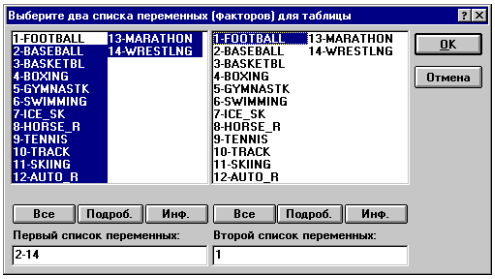
Таблицы флагов и заголовков являются экономным и наглядным способом представления нескольких двувходовых (двумерных) таблиц в одной таблице. Работая с данными, нам интересно узнать, имеют ли те же самые респонденты, которые проявили наивысший интерес к футболу, также наивысший интерес к бейсболу и баскетболу. Как отмечалось, возможно, имеет смысл показать один из этих видов спорта на одном телеэкране, а что-то существенно отличное показать на другом; такое рассредоточение программ позволит привлечь дополнительных посетителей в бар. Чтобы проверить "пересечение" между видами спорта, другими словами степень коррелированности интереса к футболу с интересом к другим видам спорта, построим таблицу флагов и заголовков.

Описание анализа

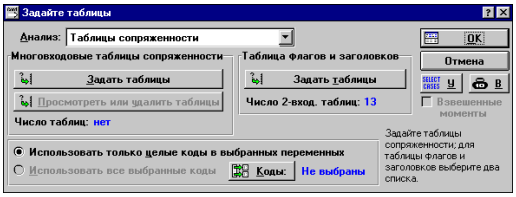
Используемый файл данных - *Sports.sta* описан в предыдущем примере (см. стр. 1494). Из стартовой панели *Основные статистики и таблицы* выберите процедуру *Таблицы и заголовки* и откройте диалоговое окно *Задайте таблицы*.

Формат таблицы. Таблица флагов и заголовков по существу содержит несколько двумерных таблиц, собранных вместе. Лучший способ понять эти таблицы – посмотреть на конкретный пример. В диалоговом окне *Задайте таблицы* нажмите кнопку *Задать таблицы* под заголовком *Таблицы флагов и заголовков.* Программа запросит ввести переменные для таблицы.

В данном примере сделаем следующий выбор: пусть переменная *1 - ФУТБОЛ – 1 - FOOTBALL* будет переменной *Столбцом* (во втором списке), а переменные *2 - БЕЙСБОЛ – 14 - БОРЬБА – 2 - BASEBALL – 14 - WRESTLING* будут переменными *Строками*.



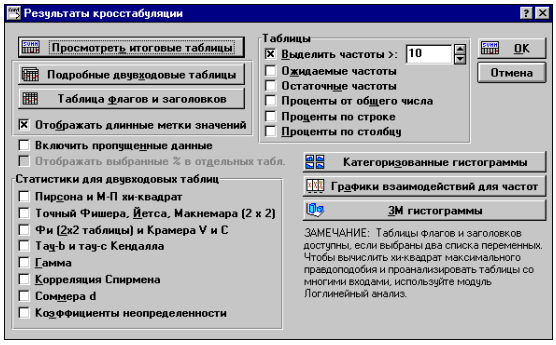
Теперь диалоговое окно *Задайте таблицы* будет выглядеть следующим образом:



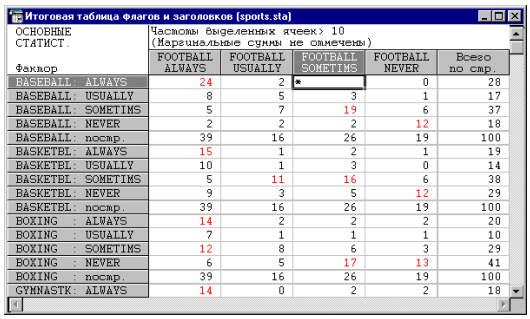
StatSoft

Copyright © StatSoft, 1995

Нажмите *OK* в этом диалоговом окне, чтобы открыть диалоговое окно *Результаты кросстабуляции*.



В этом диалоговом окне нажмите кнопку *Таблица флагов и заголовков,* чтобы отобразить таблицу результатов.

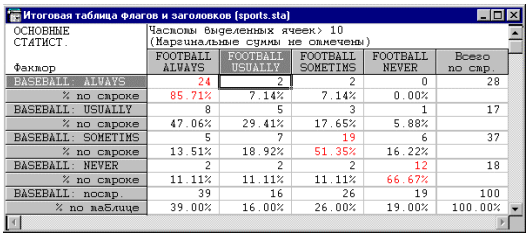


Чтение таблицы

Итак, вы можете рассматривать построенную таблицу, как объединение нескольких двувходовых таблиц. Например, в четырех начальных строках таблицы показаны частоты двумерной таблицы *ФУТБОЛ - FOOTBALL* на *БЕЙСБОЛ - BASEBALL*. Другой способ состоит в том, что вы рассматриваете значения в четырех начальных строках и четырех начальных столбцах таблицы как совместное распределение 100 респондентов в 4*4=16 ячейках, созданных пересечением интереса к футболу с интересом к бейсболу. Теперь рассмотрим различные способы представления результатов.

Частоты по строке. По умолчанию, таблица флагов и заголовков отображает частоты в строке. Таким образом, вы, например, видите, что 24 (из 100 респондентов) *всегда интересуются – always interested* футболом и *всегда интересуются – always interested* бейсболом. Посмотрите на первые три ячейки третьего столбца таблицы, вы видите, что из тех респондентов, кто *никогда не интересуется never interested* футболом, только 7 (*0+1+6*) интересуются бейсболом: *всегда – always* (*0*), *обычно - usally* (*1*) или *иногда - sometimes* (*2*). Таким образом, два вида спорта - футбол и бейсбол в наибольшей степени привлекательны для одних и тех же людей.

Проценты. Снова вернемся в диалоговое окно *Результаты кросстабуляции*. Диалоговое окно содержит настройки, позволяющие выразить результаты в процентах. Проценты могут быть вычислены относительно общего числа наблюдений в строке, относительно общего числа наблюдений в столбце или относительно общего числа наблюдений. Вы также можете включить в таблицу ожидаемые и/или остаточные частоты (разность наблюдаемых и ожидаемых частот). Выберите настройку *Проценты по строке* и нажмите снова кнопку *Таблица флагов и заголовков*.



После того, как выбрана настройка *Проценты по строке*, станет доступна настройка *Отображать выбранные % в отдельных таблицах*. Так как в одной таблице может быть слишком много информации, выбор этой настройки помещает проценты в отдельную таблицу результатов.

Copyright © StatSoft, 1995

StatSoft®

Посмотрев таблицу результатов, вы видите, что из тех респондентов, кто *всегда интересуется – always interested* бейсболом (все респонденты в первой строке), *85.71%* также *всегда интересуется – always interested* футболом. Футбол и бейсбол наиболее тесно связаны друг с другом.

Нахождение не связанного с футболом вида спорта. Посмотрим на проценты (от общего числа респондентов) чтобы найти вид спорта, максимально не связанный с футболом (иными словами, максимально далекий от футбола). Вернитесь в диалоговое окно *Результаты кросстабуляции* и выберите настройку *Проценты от общего числа* (для простоты, уберите настройку *Проценты по строке*). Снова нажмите кнопку *Таблица флагов и заголовков* и постройте таблицу с этими процентами.

Итоговая таблица флагов и заголовков (sports.sta)

ОСНОВНЫЕ СТАТИСТ.
Частоты выделенных ячеек> 10
(Маргинальные суммы не отмечены)

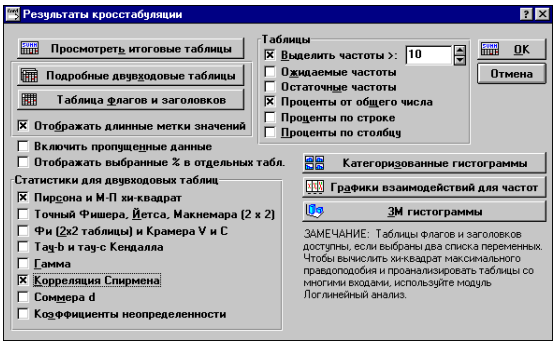
Фактор	FOOTBALL ALWAYS	FOOTBALL USUALLY	FOOTBALL SOMETIMS	FOOTBALL NEVER	Всего no cmp.
GYMNASTK: ALWAYS	14	0	2	2	18
% no таблице	14.00%	0.00%	2.00%	2.00%	18.00%
GYMNASTK: USUALLY	8	1	5	2	16
% no таблице	8.00%	1.00%	5.00%	2.00%	16.00%
GYMNASTK: SOMETIMS	12	10	10	8	40
% no таблице	12.00%	10.00%	10.00%	8.00%	40.00%
GYMNASTK: NEVER	5	5	9	7	26
% no таблице	5.00%	5.00%	9.00%	7.00%	26.00%
GYMNASTK: nocmp.	39	16	26	19	100
% no таблице	39.00%	16.00%	26.00%	19.00%	100.00%

Прокрутите таблицу результатов вниз, до *ГИМНАСТИКИ - GYMNASTK* (см. выше). Теперь входы в этой таблице – проценты от общего числа респондентов. Таким образом, 14 процентов *всех - all* респондентов сообщили, что всегда интересуются футболом и *всегда интересуются - always interested* гимнастикой. Только 7 процентов респондентов сообщили, что они *никогда не интересуются - never interested* футболом и *никогда не интересуются - never interested* гимнастикой. В то же самое время, 12 процентов (*2+2+8*) сообщили, что они никогда не интересуются футболом, но или *всегда-always* (*2%*), или *обычно - usually (2%*), или *иногда - sometimes (8%*) интересовались гимнастикой. Таким образом, гимнастика представляется хорошо рассчитанной альтернативой для второго большого телеэкрана в вашем баре и может привлечь некоторое дополнительное количество посетителей.

Статистики

В разделе *Обзор* обсуждались обычно используемые статистики для описания зависимостей между двумя переменными (см. стр. 1470). Рассмотрим некоторые из этих статистик, представленные в диалоговом окне *Результаты кросстабуляции*. Наиболее употребляемая статистика - *хи-квадрат*, поэтому выберите опцию *Пирсона и М-П (максимального правдоподобия) хи-квадрат*.

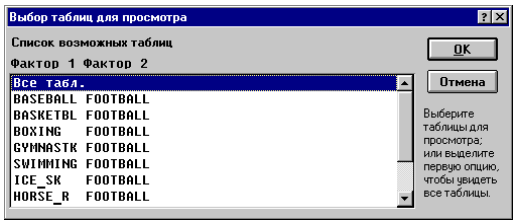
Мерой зависимости между переменными, подобно коэффициенту корреляции *r* Пирсона, является ранговая корреляция *R Спирмена* или *Rs*. Эта мера предполагает, что значения переменных содержат, по крайней мере, ранжированную информацию. Такое предположение разумно в данном примере, т.к. респондент, выразивший свой интересе к футболу словами *всегда интересуюсь - always interested,* вероятно, больший поклонник футбола, чем тот, кто просто сказал - *обычно интересуюсь - usually interested*. Таким образом, имеется естественное упорядочение значений каждой переменной. Выберите опцию *Корреляция Спирмена*. Диалоговое окно *Результаты кросстабуляции* примет следующий вид:



StatSoft®

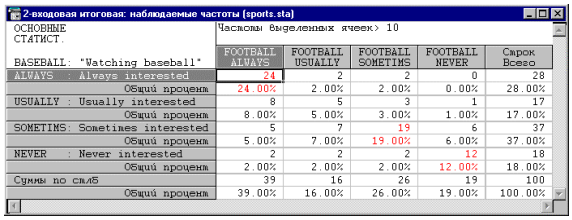
Copyright © StatSoft, 1995

Выбор таблицы. После того, как выбраны статистики, нажмите кнопку *Подробные двувходовые таблицы* для того, чтобы выбрать таблицы для анализа. На экране появится диалоговое окно *Выбор таблиц для просмотра,* в котором дан список всех двумерных таблиц:

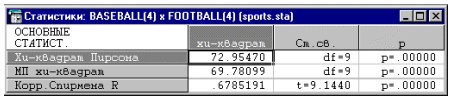


В данном окне выберите нужную таблицу. Можно воспользоваться параметром *Все таблицы,* чтобы построить каскад двувходовых таблиц. В рассматриваемом примере выберите таблицу *БЕЙСБОЛ ФУТБОЛ - BASEBALL FOOTBALL* и нажмите *OK*. Для каждой выбранной таблицы будут построены две таблицы результатов.

Первая содержит наблюдаемые частоты и все остальные характеристики, выбранные в поле *Таблицы* диалогового окна *Результаты кросстабуляции* (в частности, *Проценты от общего числа*).

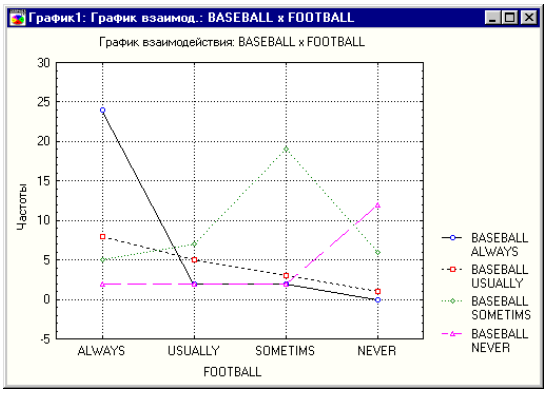


Вторая таблица содержит результаты *хи-квадрат* и корреляции *Спирмена*.

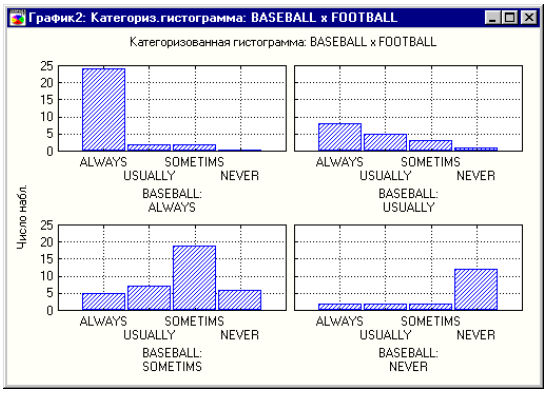


Значение статистики *Хи-квадрат* для этой таблицы равно *72.95*, что является высоко значимым (обратитесь к главе 8, *Элементарные понятия статистики*, для разъяснения понятия значимости). *R Спирмена* равна *0.6785*, что показывает существенную корреляцию между интересом к футболу и бейсболу.

В дополнение к этим методам, вы можете построить графики, если нажмете кнопку *Графики взаимодействий для частот* диалогового окна *Результаты кросстабуляции*. Откроется диалоговое окно *Выбор таблиц для просмотра*. Выберите таблицу *БЕЙСБОЛ ФУТБОЛ - BASEBALL FOOTBALL* и нажмите *OK,* чтобы построить график взаимодействий.



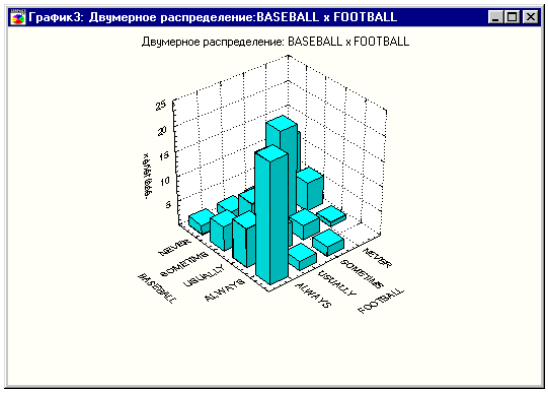
Вы также можете использовать кнопку *Категоризованные гистограммы*



или кнопку *3М гистограммы*

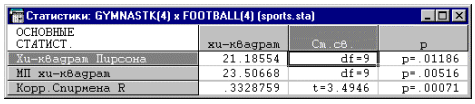
Copyright © StatSoft, 1995

StatSoft®



(из диалогового окна *Результаты кросстабуляции*) чтобы визуально исследовать частоты в выбранных двумерных таблицах.

Просмотр других таблиц. Вернитесь в диалоговое окно *Результаты кросстабуляции* и нажмите снова кнопку *Подробные двухходовые таблицы*. Выберите теперь таблицу *ГИМНАСТИКА ФУТБОЛ - GYMNASTK FOOTBALL* и нажмите *OK*. (Для этой таблицы будут вычислены те же самые статистики, т.к. не было сделано никаких новых настроек.)


Статистики: GYMNASTK(4) x FOOTBALL(4) (sports.sta)

ОСНОВНЫЕ СТАТИСТ.	хи-квадрат	Св_сВ	p
Хи-квадрат Пирсона	21.18554	df=9	p=.01186
МП хи-квадрат	23.50668	df=9	p=.00516
Корр.Спирмена R	.3328759	t=3.4946	p=.00071

Как мы видим, имеется определенная зависимость между интересом к футболу и гимнастики. Однако вряд ли ее можно рассматривать как сильную (*R Спирмена* равна *0.33*). Фактически, даже если переменная *ГИМНАСТИКА - GYMNASTK* значимо связана с футболом, можно предположить, что многие респонденты (выбранные из числа посетителей спортивных баров) на самом деле интересуются всеми видами спорта.

Итог

Основные результаты этого анализа следующие: (1) большинство респондентов в опросе интересуются *всеми - all* основными видами спорта - футболом, бейсболом, баскетболом, (2) показ гимнастики может привлечь некоторое дополнительное число респондентов из числа тех, кто отчетливо не обозначил свой интерес к футболу.

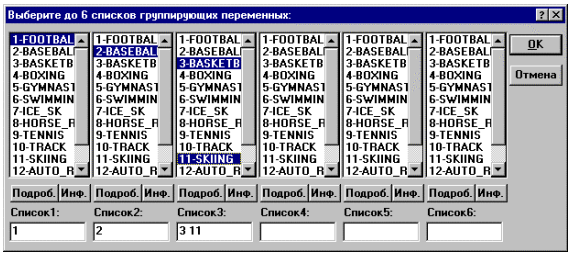
Пример 5: Таблицы сопряженности

Обзор

Для углубленного анализа результатов опроса (см. предыдущий пример), рассмотрим некоторые таблицы более высокого порядка. В частности, определим проценты респондентов, являющихся “спортивными фанатами”. Иными словами, найдем число тех респондентов, кто *всегда интересуется - always interested* и футболом, и бейсболом, и баскетболом. Для сравнения также будет вычислена таблица *ФУТБОЛ БЕЙСБОЛ ЛЫЖИ - FOOTBALL BASEBALL SKIING*.

Задание анализа

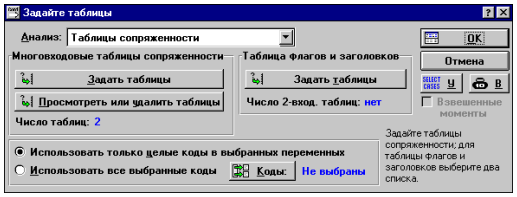
В стартовой панели модуля *Основные статистики и таблицы* выберите процедуру *Таблицы и заголовки*. Для определения таблицы нажмите на кнопку *Задать таблицы* в разделе *Многовходовые таблицы сопряженности* диалогового окна *Задайте таблицы*, будет открыто стандартное окно выбора переменных.



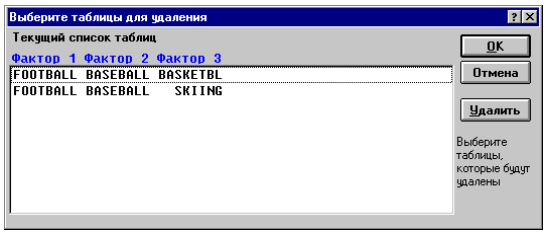
Вы можете выбрать одну и более переменных в каждом из шести списков, чтобы создать таблицы со многими входами. В данном

StatSoft®
Copyright © StatSoft, 1995

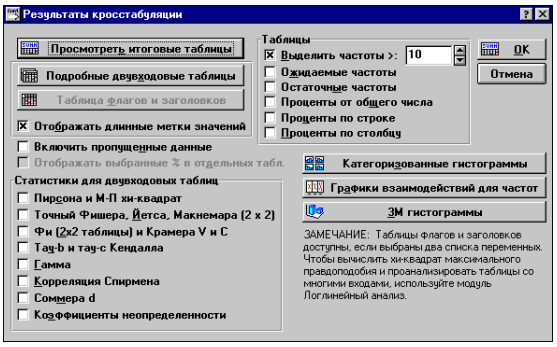
примере определим две таблицы: *ФУТБОЛ – БЕЙСБОЛ - БАСКЕТБОЛ – FOOTBALL – BASEBALL - BASKETBL* и *ФУТБОЛ – БЕЙСБОЛ - ЛЫЖИ – FOOTBALL – BASEBALL - SKIING*. В диалоговом окне выбора переменных выберите переменную *ФУТБОЛ - FOOTBALL* в первом списке, *БЕЙСБОЛ - BASEBALL* во втором, *БАСКЕТБОЛ - BASKETBL* и *ЛЫЖИ - SKIING* в третьем. Выбранные таким образом переменные создадут таблицы, в которых уровни переменной *БАСКЕТБОЛ - BASKETBL* используются для того, чтобы "построить срезы" одной из таблиц и переменная *ЛЫЖИ – SKIING*, чтобы "построить срезы" другой таблицы. Другими словами, каждая комбинация уровней *ФУТБОЛ - БЕЙСБОЛ – FOOTBALL - BASEBALL* будет просмотрена на каждом уровне переменных *БАСКЕТБОЛ - BASKETBL* (или *ЛЫЖИ - SKIING*). Теперь диалоговое окно *Задайте таблицы* выглядит следующим образом:



Поскольку выбор уже сделан, кнопка *Просмотреть или удалить таблицы* становится активной (иными словами, она не затенена). Нажмите ее и откройте диалоговое окно *Выберите таблицы для удаления*.

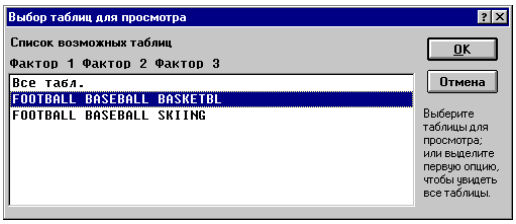


Здесь можно просмотреть выбранные таблицы и, если необходимо, удалить определенные таблицы из списка. Нажмите *OK* в диалоговом окне *Задайте таблицы*, после этого откроется диалоговое окно *Результаты кросстабуляции*.



Это то же диалоговое окно, что и в примере с таблицами флагов и заголовков, единственное отличие – не активна кнопка *Таблицы флагов и заголовков*. Выберите еще раз параметры таблицы (например, *Проценты по строке*, *Проценты от общего числа* и т.д.) и статистики (например, *Хи-квадрат*, корреляции, и т.д.), нажав либо кнопку *Просмотреть итоговые таблицы* либо кнопку *Подробные двувходовые таблицы*.

В любом случае на экране появится промежуточное диалоговое окно, в котором можно выбрать таблицу из уже выбранных. Если выбрана команда *Все таблицы*, то каскад таблиц результатов будет построен для каждой таблицы, показанной в этом диалоговом окне.



В этом примере после нажатия кнопки *Просмотреть итоговые таблицы*, выберите в диалоговом окне таблицу *ФУТБОЛ БЕЙСБОЛ БАСКЕТБОЛ - FOOTBALL BASEBALL BASKETBL*. Для *Примера 3* процедура

Copyright © StatSoft, 1995

StatSoft®

Подробные двувходовые таблицы дает следующую таблицу:

Итоговая таблица частот (sports.sta)

ОСНОВНЫЕ СТАТИСТ. — Числомы выделенных ячеек> 10 (Итоговые маргинальные не отмечены)

FOOTBALL	BASEBALL	BASKETBL ALWAYS	BASKETBL USUALLY	BASKETBL SOMETIMS	BASKETBL NEVER	Строк Всего
ALWAYS	ALWAYS	11	8	2	3	24
ALWAYS	USUALLY	2	2	3	1	8
ALWAYS	SOMETIMS	2	0	0	3	5
ALWAYS	NEVER	0	0	0	2	2
	Всего	15	10	5	9	39
USUALLY	ALWAYS	0	0	1	1	2
USUALLY	USUALLY	0	1	4	0	5
USUALLY	SOMETIMS	1	0	5	1	7
USUALLY	NEVER	0	0	1	1	2
	Всего	1	1	11	3	16
SOMETIMS	ALWAYS	0	1	1	0	2
SOMETIMS	USUALLY	0	0	2	1	3
SOMETIMS	SOMETIMS	2	2	12	3	19
SOMETIMS	NEVER	0	0	1	1	2
	Всего	2	3	16	5	26
NEVER	ALWAYS	0	0	0	0	0

Как вы видите, 11 респондентов из 100 сообщили, что они *всегда интересуются - always interested* футболом, бейсболом и баскетболом. Вернитесь в диалоговое окно *Результаты кросстабуляции*, снова нажмите кнопку *Просмотреть итоговые таблицы* и выберите таблицу *ФУТБОЛ БЕЙСБОЛ ЛЫЖИ - FOOTBALL BASEBALL SKIING*.

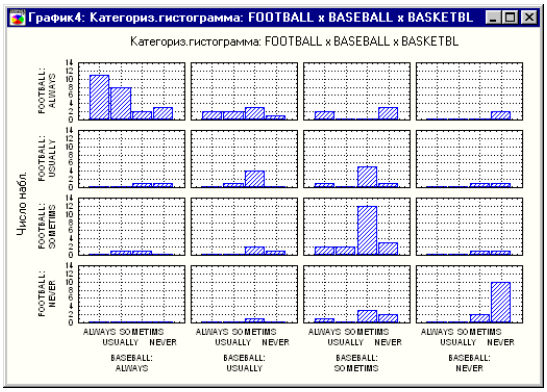
Итоговая таблица частот (sports.sta)

ОСНОВНЫЕ СТАТИСТ. — Числомы выделенных ячеек> 10 (Итоговые маргинальные не отмечены)

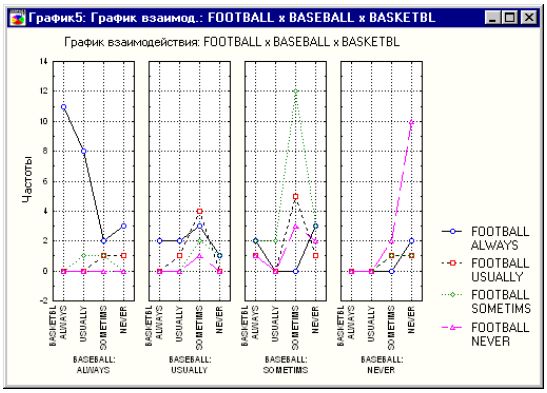
FOOTBALL	BASEBALL	SKIING ALWAYS	SKIING USUALLY	SKIING SOMETIMS	SKIING NEVER	Строк Всего
ALWAYS	ALWAYS	5	3	9	7	24
ALWAYS	USUALLY	1	0	2	5	8
ALWAYS	SOMETIMS	1	1	2	1	5
ALWAYS	NEVER	0	1	0	1	2
	Всего	7	5	13	14	39
USUALLY	ALWAYS	0	0	2	0	2
USUALLY	USUALLY	0	1	2	2	5
USUALLY	SOMETIMS	0	1	4	2	7
USUALLY	NEVER	0	0	2	0	2
	Всего	0	2	10	4	16
SOMETIMS	ALWAYS	1	0	1	0	2
SOMETIMS	USUALLY	0	0	1	2	3
SOMETIMS	SOMETIMS	3	2	8	6	19
SOMETIMS	NEVER	0	0	2	0	2
	Всего	4	2	12	8	26
NEVER	ALWAYS	0	0	0	0	0

Здесь имеется только 5 респондентов, заявивших, что они *всегда интересуются - always interested* и футболом, и бейсболом, и лыжами.

Графическое представление кросстабуляций. Результаты кросстабуляции можно представить графически с помощью кнопок *Категоризованные гистограммы*, *Графики взаимодействий для частот* и *3М гистограммы*. Как и для построения таблиц результатов, диалоговое окно *Выбор таблиц для просмотра* будет открыто, когда вы нажмете одну из графических кнопок. Для каждого случая выберите таблицу *Футбол Бейсбол Баскетбол - Football Baseball Basketbl*. *Категоризованные гистограммы* для таблицы *Футбол Бейсбол Баскетбол - Football Baseball Basketbl* показаны ниже.



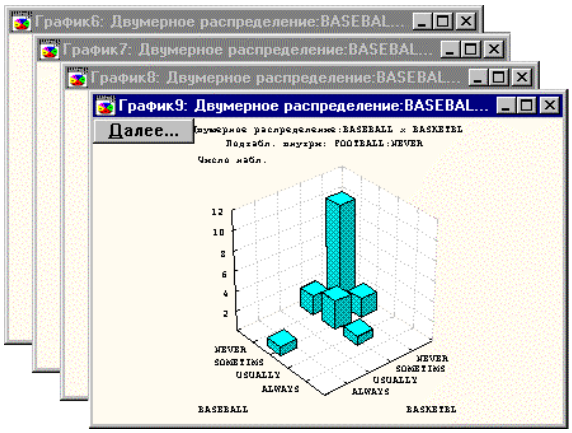
Когда вы нажмете на кнопку *Графики взаимодействий для частот*, для таблицы *ФУТБОЛ БЕЙСБОЛ БАСКЕТБОЛ - FOOTBALL BASEBALL BASKETBL* будет построен следующий график:



Если выбрана процедура *3М гистограммы*, на экране появится каскад 2М гистограмм (по одной для каждой подтаблицы *ФУТБОЛ - FOOTBALL*).

StatSoft

Copyright © StatSoft, 1995



Итог

Подводя итог, можно сказать, что 11 из 100 респондентов сообщили, что всегда интересуются футболом, бейсболом и баскетболом по ТВ. Таким образом, около 10% посетителей спортивных баров составляют спортивные фанаты, смотрящие самые популярные спортивные передачи.

Пример 6: Табулирование многомерных откликов и дихотомий

Общая идея исследования этих переменных приведена в разделе *Обзор* (стр. 1474). Если вы не освоились с переменными, имеющими многомерные отклики, и близкими к ним многомерными дихотомиями, рекомендуем прочитать раздел *Обзор*, до того как вы начнете рассматривать пример. Пример показывает, как обращаться с многомерными откликами и дихотомиями, часто возникающими в массовых опросах, а также какие возможности для анализа этих переменных имеются в модуле *Основные статистики и таблицы*.

Пример с результатами гипотетического опроса находится в каталоге *Примеры*. На нем показывается, как табулируются следующие три типа переменных:

- Простые группирующие переменные
- Переменные с многомерными откликами
- Многомерные дихотомии

Вначале покажем, как строятся простые таблицы частот для этих переменных, затем построим и исследуем таблицы сопряженности для них.

Описание файла данных

Представьте, что проводится исследование покупательских предпочтений молодых людей. Задаются следующие вопросы: (1) какую систему быстрого питания вы предпочитаете, (2) какой тип автомобиля вы предпочитаете, (3) какой местный ресторан вы посещали в течение последних двух недель. Дополнительно записывается пол респондента. Эти предпочтения записаны в файл *Fastfood.sta*. Переменные файла описаны ниже.

Пол (простая группирующая переменная). Пол респондента записывается в группирующую переменную *Пол - Gender* (*Мужчина - Male*, *Женщина - Female*).

Лучшая “быстрая” еда (переменная с многомерными откликами). Вопросник, используемый в данном исследовании, предлагает респондентам выбрать любимое “быстрое”блюдо (до трех блюд) из следующего списка:

(1) *Гамбургер - Hamburger*

(2) *Сэндвич - Sandwiches*

(3) *Цыпленок - Chicken*

(4) *Пицца - Pizza*

(5) *Мексиканские блюда - Mexican fast-food*

(6) *Китайские блюда- Chinese fast-food*

Copyright © StatSoft, 1995

(7) *Еда из морепродуктов - Seafood*

(8) *Другие национальные блюда - other ethnic or regionally popular fast-food.*

Выбор каждого респондента был введен в файл как переменная с многомерными откликами. Это означает, что первый пункт ответа записывается в столбец *Еда_1 - Food_1* (первое предпочтение), второй пункт (если он имеется) - в переменную *Еда_2 - Food_2* и третий - в переменную *Еда_3 - Food_3*.

В анализе переменная *Еда_1 - Food_1* может рассматриваться как простая группирующая переменная. Далее можно задать вопрос: Какое число респондентов (или их доля) назвали определенный тип системы быстрого питания своим *любимым - favorite*? Однако интерес может представлять также то, сколько респондентов выбрали определенную систему быстрого питания как *одну* из любимых. Такой вопрос приводит нас к тому, чтобы рассмотреть переменные *Еда_1 - Еда_3 - Food_1 - Food_3* как одну переменную с многомерным откликом. Такие переменные будем называть также многомерными.

Любимый автомобиль (переменная с многомерными откликами). В этом опросе субъектов просят назвать три самых любимых типа автомашины (фактор денег, стоимость машины, не учитывается, просто спрашивается о некотором идеальном автомобиле). Эти ответы (определенные марки и модели) были закодированы следующим образом:

(1) *Отечественный спортивный автомобиль - Domestic sports car*

(2) *Отечественный седан (закрытый автомобиль) - Domestic sedan*

(3) *Иностранная спортивная машина - Foreign sports car*

(4) *Иностранный седан - Foreign sedan.*

Данная переменная рассматривается как переменная с многомерными откликами, подобно переменной *любимая система быстрого питания - favorite fast-food*. Это означает, что ответы респондентов были введены в переменные *Машина_1 - Машина_3 - Car_1- Car_3*. Заметьте, что в данном случае субъекты могли повторить один и тот же ответ три раза (например, они могли назвать одну и ту же машину трижды). В случае с *быстрым питанием - fast-food* идентичные ответы не допускаются (игнорируются).

Например, если респондент назвал три любимых блюда *Гамбургер - Hamburger, Гамбургер - Hamburger*, и *Гамбургер - Hamburger*, тогда *Гамбургер - Hamburger* в обработку вводился только однажды (в переменную *Еда_1 - Food_1*), а соответствующие ячейки переменных *Еда_2 - Food_2* и *Еда_3 - Food_3* рассматривались как пустые.

Недавние посетители ресторанов (многомерная дихотомия). Посетителей ресторана попросили назвать, какие из четырех ресторанов они посещали за последние две недели. Полученные данные были введены в файл так, что для каждого ресторана имелась своя переменная. Всего использовано четыре переменных *Хозяин_1 - Хозяин _4 - Burger_1 - Burger_4* для следующих ресторанов:

(1) *Бутерброд Мейстер - Burger Meister*

(2) *Лучшие бутерброды у Билла - Bill's Best Burgers*

(3) *Гамбургер "Блаженство" - Hamburger Heaven*

(4) *Большой бутерброд - Bigger Burger.*

Если респондент сообщил, что в течение двух недель обедал в одном или нескольких ресторанах, то в соответствующий столбец (столбцы) ставилась единичка; если нет, столбец оставался пустым. Таким образом, переменная представляет собой многомерную дихотомию

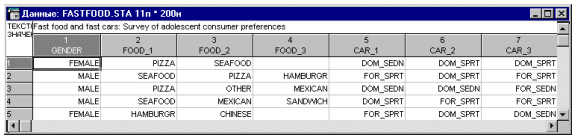
StatSoft®

Copyright © StatSoft, 1995

(со значениями *Да* или пропуск), которую желательно табулировать, т.е. указать число (или долю) респондентов, обедавших в каждом из четырех ресторанов.

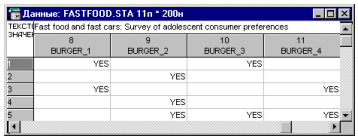
Заметьте, можно было бы рассмотреть эту переменную как переменную с многомерными откликами. Однако для этого нужно создать не менее четырех переменных (например, *Еда_1 - Еда_4 - Eat_1 - Eat_4*) и затем ввести названия ресторанов (например, *Бургер_1 - Burger_1*, *Бургер_2 - Burger_2 ...*) как значения этих переменных в столбцы таблицы (аналогично переменным *любимая машина - favorite car* и *любимая система быстрого питания - favorite fast-food*, см. выше).

Распечатка файла данных. Ниже показано несколько первых наблюдений файла данных *Fastfood.sta*.


Данные: FASTFOOD.STA 11v * 200n

Fast food and fast cars: Survey of adolescent consumer preferences

	1 GENDER	2 FOOD_1	3 FOOD_2	4 FOOD_3	5 CAR_1	6 CAR_2	7 CAR_3
1	FEMALE	PIZZA	SEAFOOD		DOM_SEDN	DOM_SPRT	DOM_SPRT
2	MALE	SEAFOOD	PIZZA	HAMBURGR	FOR_SPRT	DOM_SPRT	FOR_SPRT
3	MALE	PIZZA	OTHER	MEXICAN	DOM_SEDN	DOM_SEDN	FOR_SEDN
4	MALE	SEAFOOD	MEXICAN	SANDWICH	DOM_SPRT	FOR_SPRT	FOR_SPRT
5	FEMALE	HAMBURGR	CHINESE		FOR_SPRT	DOM_SPRT	DOM_SEDN


Данные: FASTFOOD.STA 11v * 200n

Fast food and fast cars: Survey of adolescent consumer preferences

	8 BURGER_1	9 BURGER_2	10 BURGER_3	11 BURGER_4
1	YES		YES	
2		YES		
3	YES			YES
4		YES		
5		YES	YES	YES

Для того чтобы показать, каким образом каждый опрашиваемый введен в файл, посмотрите на первое наблюдение. Первый респондент – женщина, поэтому в переменную *Пол - Gender* введено значение *Женщина - Female*. Самое любимое быстро приготовленное блюдо - *Пицца - Pizza* (введено в переменную *Еда_1 - Food_1*), второе, по предпочтению, блюдо - *Еда из морепродуктов - Seafood* (введено в переменную *Еда_2 - Food_2*), третий вид еды не указан, поэтому в переменной *Еда_3 - Food_3* стоит пропуск.

Далее этот респондент выбрал следующие 3 типа автомобилей: (1) *домашний седан - domestic sedan*, (2) *домашний спортивный автомобиль - domestic sports car*, (3) снова *домашний спортивный автомобиль - domestic sports* - переменные *Car_1, Car_2, Car_3, - Car_1, Car_2, Car_3* соответственно. Наконец, он ответил, что последние две недели обедал в двух ресторанах *Burger_1* (*Burger Meister*) и *Burger_3* (*Hamburger Heaven*), таким образом, *Да - Yes* было записано в ячейках соответствующих переменных, значения двух других переменных *Burger* остались пустыми.

Всего в исследовании опрошено 200 респондентов.

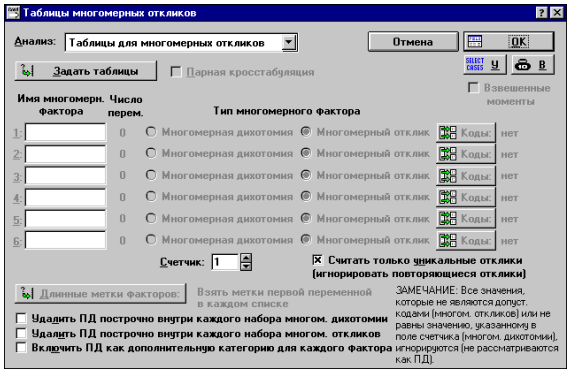
Задание простых группирующих переменных и переменных с многомерными откликами

Начнем с вычисления таблиц частот для простой группирующей переменной *Пол - Gender* и переменных с многомерными откликами. Так как имеются пропущенные значения во всех переменных *Burger_1 - Burger_4*, таблица для них будет определена позже.

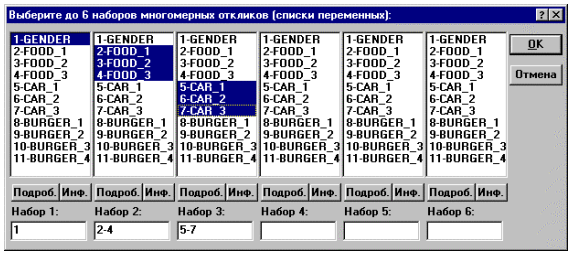
По умолчанию, наблюдения со всеми пропусками в переменных *Burger* исключаются из анализа, и частоты будут вычисляться лишь для респондентов, обедавших, по крайней мере, в одном из четырех ресторанов. Другой способ обработки пропусков состоит в том, чтобы сделать отметку в поле *Включить ПД как дополнительную категорию для каждого фактора*. После открытия модуля *Основные статистики и таблицы*, выберите *Таблицы и заголовки* в стартовой панели (см. стр. 1521). В появившемся окне *Задайте таблицы* выберите *Таблицы для многомерных откликов* в списке *Анализ*, при этом откроется диалоговое окно *Таблицы многомерных откликов*. В этом окне можно определить три типа группирующих переменных: простые группирующие переменные (*Пол - Gender* в нашем примере), переменные с многомерными откликами (*Еда_1*

Copyright © StatSoft, 1995

- Food_1 - Еда_3 - Food_3 или *Машина_1 - Car_1- Машина_3 - Car_3*), и многомерные дихотомии (*Burger_1 - 4*).

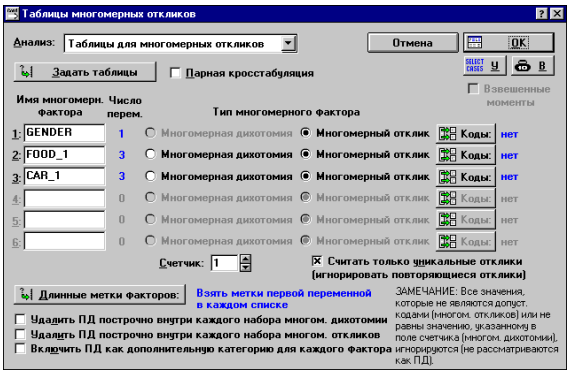


Нажмите кнопку *Задать таблицы* для того, чтобы определить переменные в диалоговом окне:



В окне можно выбрать до 6 многомерных факторов (простых группирующих переменных, многомерных откликов или дихотомий) для одной таблицы. В первой колонке выберем только переменную *Пол - Gender*; - программа автоматически рассматривает единственную выбранную переменную как простую группирующую (простая группирующая является частным случаем переменной с многомерными откликами, для нее число откликов 1). Во второй колонке выберите переменные *Еда_1 -Еда_3 - Food_1 - Food_3*, в третьей *Машина_1-Машина_3 - Car_1 - Car_3.* Вначале посмотрим на простые таблицы частот для всех выбранных факторов (таблица частот для *Burger_1 - Burger_4* будет исследована позже). Нажмите *OK,* чтобы завершить выбор.

Теперь в окне *Таблицы многомерных откликов* вы видите выбранные переменные.

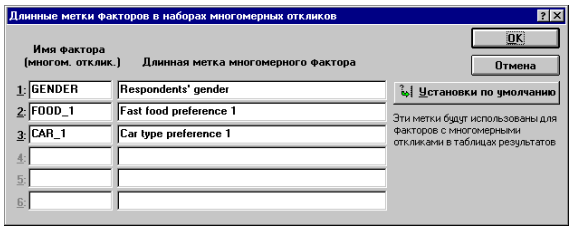


Соглашения об обозначениях. В самой левой колонке этого диалогового окна указаны имена факторов. Термин *фактор* используется для общего обозначения всех типов переменных (например, такая переменная, как *любимая еда - food preference* в действительности состоит из нескольких переменных). Мы употребляем термин многомерный фактор и для простых переменных, и для переменных с многомерными откликами, и для многомерных дихотомий. Заметим, что по умолчанию фактору присваивается имя (как длинное, так и короткое) первой переменной в соответствующем списке. Текстовые значения также будут взяты из этой переменной (см. объяснение численных/текстовых значений системы *STATISTICA*, в *Менеджере тестовых значений*, глава 3; см. также *Управление данными*, глава 7).

Если нужно изменить имя фактора, нажмите кнопку *Длинные метки факторов* и откройте диалоговое окно *Длинные метки факторов в наборах многомерных откликов*. Здесь можно ввести нужные имена.

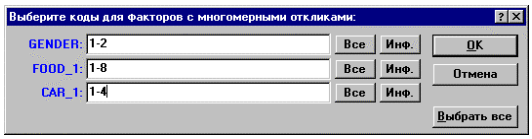
StatSoft®

Copyright © StatSoft, 1995



В этом примере слегка изменим имена факторов с многомерными откликами, так чтобы они более точно отвечали начальному смыслу: *Еда: 3 любимых системы быстрого питания - Food: Favorite 3 fast-foods*, и *Машина*: *Любимые типы машин - Car*: *Favorite types of car.*

Определение факторов. Расположенная рядом с каждым фактором опция позволяет определить его тип. Первая переменная - *Пол - Gender* – простая группирующая переменная. Для второго и третьего факторов выберите опцию *Многомерный отклик*. Далее выберите коды для определения различных категорий. Выберите коды, чтобы идентифицировать пол респондента *Мужчина - Male* и *Женщина - Female* (переменная *Пол - Gender)*, а также различные типы “быстрой” еды в переменных *Еда_1-Еда_3 - Food_1 - Food_3* и различные типы автомобилей в *Машина_1-Машина_3 - Car_1 - Car_3*. Если вы не зададите коды явно, (просто нажмете *OK*), программа возьмет их из первой переменной в каждом факторе. Данный способ обычно позволяет определить все коды, однако может случиться так, что определенный код не присутствует в первой переменной, а присутствует только во второй или в третьей. Тогда способ по умолчанию не годится (ряд значений окажутся неучтенными). Лучше задать все используемые коды точно. После нажатия одной из кнопок *Коды*, расположенной рядом с каждым фактором, можно ввести коды для фактора.



Другие опции. До проведения анализа рассмотрим некоторые другие опции диалогового окна *Таблицы многомерных откликов*. Опция *Парная кросстабуляция* заставляет программу “сравнивать” факторы с равным числом уровней и рассматривать их как парные переменные. Подробно парная кросстабуляция описана в разделе *Обзор* (см. стр. 1476).

Опции, расположенные ниже кнопки *Длинные метки факторов* в окне *Таблицы многомерных откликов* задают способ обработки пропущенных значений (пустых ячеек в файле) . Например, по умолчанию, когда обрабатывается фактор со значениями типа "многомерные дихотомии", программа учитывает только значения, указанные в поле *Счетчик* и игнорирует все остальные значения переменных. Пропущенные данные можно исключить *построчным* способом. В этом способе из анализа исключаются все респонденты, не давшие ответа хотя бы на один пункт какого-либо вопроса. Вторая установка – *Удалить ПД построчно внутри каждого набора многомерных откликов* – работает аналогично (пропуски исключаются *построчно* – внутри каждой многомерной переменной). Последняя (третья) установка - *Включить ПД как дополнительную категорию для каждого фактора* позволяет создать отдельную категорию (группу) для пропущенных данных.

Наконец, рассмотрим поле *Считать только уникальные отклики.* Оно задает способ рассмотрения многомерных идентичных или совпадающих откликов. В данном примере многомерный фактор *Машина - Car* состоит из трех переменных с кодами, определяющими типы автомобиля. Респондент мог выбрать три иностранных автомобиля как наиболее желательные, которые будут иметь один и тот же код (*FOR_SPRT*) для каждой из трех переменных *Машина_1 - Машина_3 - Car_1 - Car_3*. Иными словами, дают один и тот же ответ на все три пункта. Процедура определяет, будут ли эти

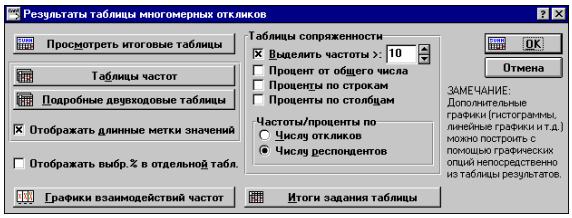
Copyright © StatSoft, 1995

одинаковые ответы учтены три раза (если не выбрана настройка *Считать только уникальные отклики*) или только один раз (если опция выбрана).

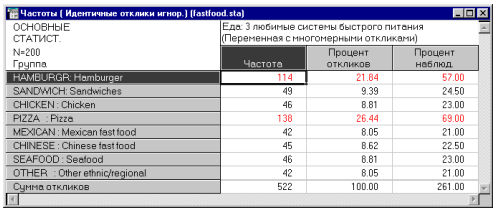
В данном примере не так интересно знать, были ли все три выбранных машины определенного типа (в связи с чем чрезмерно увеличивается число идентичных откликов); поэтому имеет смысл выбрать опцию для того, чтобы определить число респондентов, предпочитающих, например, домашнюю закрытую машину. Заметим, что переменные составляющие фактор *Еда – Food,* содержат только взаимно эксклюзивные (непересекающиеся) категории, т.к. респондентам не разрешалось давать идентичные ответы, (например, *Гамбургер - Hamburger, Гамбургер - Hamburger,* и *Гамбургер - Hamburger*). Их просили сделать выбор из 8 типов быстрой еды без повторения. Поэтому для фактора *Еда - Food* данная опция не имеет значения.

Просмотр таблиц частот для факторов с многомерными откликами

Нажмите *OK* в диалоговом окне *Таблицы многомерных откликов*, чтобы начать анализ и открыть окно *Результаты таблицы многомерных откликов*.



Вначале просмотрим простой вывод *Таблицы частот* (таблицы сопряженности, которые могут быть выбраны в этом диалоговом окне, будут рассмотрены ниже, см. стр. 1515). Введите в редактируемое поле *Выделить частоты* число 100 (что приведет к тому, что все частоты больше 100 будут выделены в таблице результатов). Затем нажмите кнопку *Таблицы частот*. Таблица частот для переменной *Пол - Gender* интерпретируется обычным образом и на ней останавливаться не будем. Таблицы частот для других двух факторов показаны ниже.


Частоты (Идентичные отклики игнор.) (fastfood.sta)

ОСНОВНЫЕ СТАТИСТ.: N=200, Группа

Еда: 3 любимые системы быстрого питания (Переменная с многомерными откликами)

	Частота	Процент откликов	Процент наблюд.
HAMBURGR: Hamburger	114	21.84	57.00
SANDWICH: Sandwiches	49	9.39	24.50
CHICKEN : Chicken	46	8.81	23.00
PIZZA : Pizza	138	26.44	69.00
MEXICAN : Mexican fast food	42	8.05	21.00
CHINESE : Chinese fast food	45	8.62	22.50
SEAFOOD : Seafood	46	8.81	23.00
OTHER : Other ethnic/regional	42	8.05	21.00
Сумма откликов	522	100.00	261.00

Частоты (Идентичные отклики игнор.) (fastfood.sta)

ОСНОВНЫЕ СТАТИСТ.: N=200, Группа

Машина: Любимые типы машин (Переменная с многомерными откликами)

	Частота	Процент откликов	Процент наблюд.
DOM_SPRT: Domestic sports car	123	29.29	61.50
DOM_SEDN: Domestic sedan	76	18.10	38.00
FOR_SPRT: Foreign sports car	157	37.38	78.50
FOR_SEDN: Foreign sedan	64	15.24	32.00
Сумма откликов	420	100.00	210.00

Прежде всего, взгляните на таблицу для фактора *Еда - Food.* Всего в исследовании было опрошено 200 респондентов (число опрошенных *N=200* показано в верхнем левом углу таблицы). Столбец *Частота* показывает число респондентов, назвавших данный способ питания как *один* из любимых. Напомним, что учитываются только уникальные ответы (см. выше), таким образом, ответ каждого респондента может быть посчитан только один раз в этом столбце. Отсюда вы можете придти к заключению, что *Пицца - Pizza* была самой популярной системой быстрого питания, указанной либо в первой, либо во второй, либо в третьей позиции 138 респондентами, *Гамбургер - hamburger* был вторым по популярности (114). Все типы систем быстрого питания отметили только 40 - 50 респондентов.

Во втором столбце таблицы результатов вычислены относительные частоты, соответствующие числам первого столбца. Можно сказать, например, что 26.44% (100*138/522) всех указанных в ответах предпочтений составляет *Пицца - Pizza.* В отличие от этой колонки, третья колонка

StatSoft®
Copyright © StatSoft, 1995

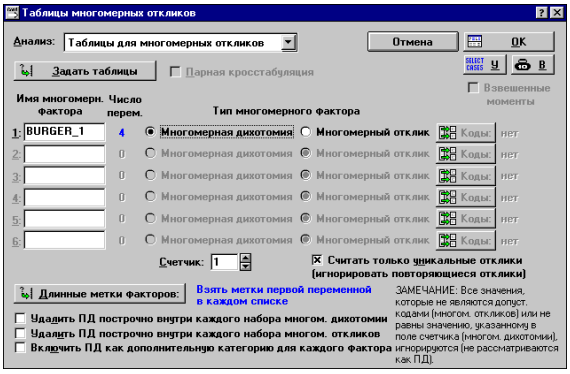
таблицы показывает проценты респондентов, отметивших соответствующий тип еды как первый, второй или третий. *Пиццу - Pizza* как лучшую систему быстрого питания выбрали 69% (100*138/200) всех респондентов.

Аналогично рассматривается таблица частот для фактора *Машина - Car*. Иностранные спортивные машины отмечены 157 респондентами на одной из трех позиций (учитываются только различные ответы); отечественные спортивные машины отмечены 123 респондентами. Вторая колонка показывает 37.38% ответов для иностранных спортивных машин; эти числа не так легко проинтерпретировать, т.к. подсчитывались только различные ответы (несколько одинаковых ответов рассматривались как один). Таким образом, если респондент указал в анкете три иностранные спортивные машины, то этот ответ учтен только один раз. Числа в третьей колонке (*Процент набл.*) более информативны; из них, например, видно, что 78.5% всех респондентов назвали иностранные спортивные машины в числе трех самых любимых.

Задание многомерной дихотомии

Возвратитесь в диалоговое окно *Таблицы многомерных откликов* (нажмите *Отмена* в окне *Результаты*), чтобы задать многомерную дихотомию в обследовании посетителей ресторанов. Нажмите кнопку *Задать таблицы*, отмените предыдущий выбор и выберите *Burger_1 - Burger_4* как переменные первого множества.

Далее установите опцию *Многомерная дихотомия* рядом с первым фактором в диалоговом окне *Таблицы многомерных откликов*. Как и ранее, можете использовать опцию *Длинные метки факторов* для того чтобы ввести подходящее имя фактора. Например, можно назвать этот фактор *Patron: Recently patronized restaurants – Хозяин: Недавно посещенные рестораны*.



Вам также необходимо задать код, который использовался в факторе - многомерной дихотомии *Patron* для того, чтобы определить обедал или нет респондент в соответствующем ресторане в течение двух недель перед опросом. Задайте нужный код в поле *Счетчик* ниже списка факторов. Так как код, равный *1* (числовой эквивалент значения *Да - Yes*; см. *Управление данными*, глава 7) использовался для того, чтобы определить какой ресторан посещался респондентом, то можно просто принять код, предложенный по умолчанию.

Напомним, каким образом многомерные дихотомии интерпретируются программой. Переменные, из которых построен фактор, рассматриваются как его уровни, затем подсчитывается число уровней со значениями, равными значению, указанному в счетчике. Все значения, не равные этому значению игнорируются. Вы можете строить более “сложные” схемы кодирования (а не просто 1-0

Copyright © StatSoft, 1995

StatSoft

как в этом примере), задавая подходящие значения в поле *Счетчик*.

Например, можно использовать отдельный код (отличный от 1) для обозначения ответа: “даже никогда не думал там обедать.” Вы могли бы ввести код *2* в переменные *Burger_1 - Burger_4* для обозначения таких резко отрицательных ответов в отношении определенных ресторанов, задать этот код в поле *Счетчик* и табулировать эти ответы. Таким образом, задавая различные значения для кодов многомерной дихотомии можно идентифицировать взаимоисключающие ответы.

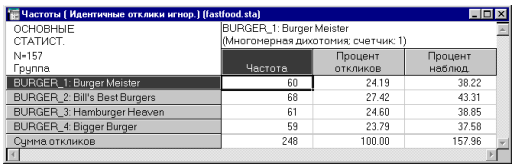
Пропущенные данные. Всегда найдутся респонденты, не обедавшие ни в одном из названных ресторанов. В исходной таблице с данными эти наблюдения будут пустыми (пропущенные значения) для всех четырех переменных *Burger_1 - Burger_4*. По умолчанию, такие респонденты исключаются из анализа. Однако не обедавшие в названных ресторанах также несут определенную информацию.

Вы можете воспользоваться опцией *Включить ПД как дополнительную категорию для каждого фактора*. В этом случае результирующая таблица частот будет иметь дополнительную категорию *Пропущенные*, в которой будет записано число респондентов, не обедавших ни в одном из названных ресторанов.

В данном примере желательно рассматривать респондентов, посетивших хотя бы один ресторан, для того, чтобы увидеть как эти рестораны (“гамбургерные рестораны”) распределены среди рассматриваемых ресторанов (см. ниже). Нажмите *OK*, чтобы открыть диалоговое окно *Результаты таблицы многомерных откликов*.

Просмотр таблицы частот для многомерной дихотомии

Из диалогового окна *Результаты* снова выберите процедуру *Таблицы частот*. Интерпретация чисел, представленных в этой таблице, аналогична таблицам для многомерных откликов.


Частоты (Идентичные отклики игнор) (fastfood.sta)

ОСНОВНЫЕ СТАТИСТ. N=157 Группа	BURGER_1: Burger Meister (Многомерная дихотомия; счетчик: 1)		
	Частота	Процент откликов	Процент наблюд.
BURGER_1: Burger Meister	60	24.19	38.22
BURGER_2: Bill's Best Burgers	68	27.42	43.31
BURGER_3: Hamburger Heaven	61	24.60	38.85
BURGER_4: Bigger Burger	59	23.79	37.58
Сумма откликов	248	100.00	157.96

Всего 157 респондентов обедали в одном из четырех ресторанов (*n=157*); 60 респондентов обедали в *Burger Meister,* 68 – в *Bill’s Best Burgers* и т.д. Значения во второй колонке (*Процент откликов*) выражают эти числа в процентах от общего числа респондентов, обедавших хотя бы в одном ресторане (т.е. от 157 респондентов).

Предполагается, что четыре (воображаемых) ресторана делят рынок быстрого питания в городе, и что 157 респондентов (из 200) в большей или меньшей степени представляют мнение общего рынка. Поэтому значения во второй колонке таблицы показывают долю рынка, которым владеет каждый ресторан.

Например, из всех мест (где подаются гамбургеры), которые посещались респондентами в течение двух недель до опроса, *Burger Meister* посещали 24.19%, *Bill’s Best Burger* - 27.42% и т.д. Третья колонка (*Процент набл.*) содержит процент респондентов, обедавших последние две недели в соответствующих ресторанах.

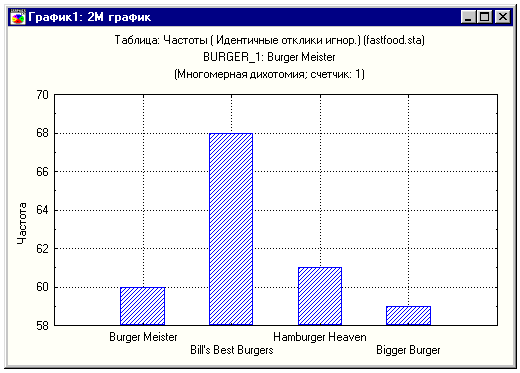
Напомним, что проценты вычислены для *N* = 157, т.е. относительно числа респондентов, обедавших, по крайней мере, в одном из четырех ресторанов. Поэтому вы можете сказать, что 38.22% респондентов, обедавших в каком-то одном из четырех гамбургеровых ресторанах,

StatSoft®

Copyright © StatSoft, 1995

обедали также в *Burger Meister* - 43.31% обедали в *Bill's Best Burger* и т.д.

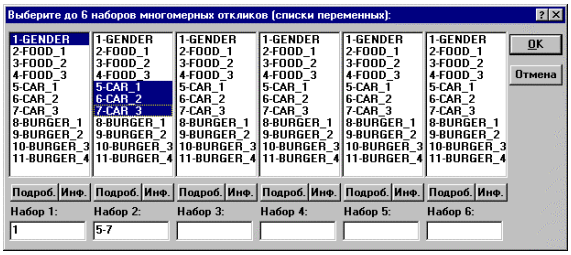
Заметим, что легко построить линейные графики или гистограммы частот и процентов с помощью процедур меню *Пользовательские графики*.



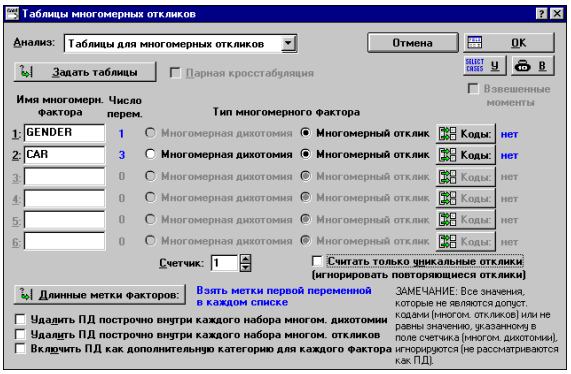
За дальнейшей информацией о *Пользовательских графиках*, обратитесь к тому II (*Пользовательские 2М графики*).

Кросстабуляция многомерных откликов и дихотомий

Рассмотрим теперь некоторые таблицы сопряженности для переменных с многомерными откликами и многомерных дихотомий. Нажмите *Отмена* в диалоговом окне *Результаты* для того, чтобы вернуться в диалоговое окно *Таблицы многомерных откликов*. Прежде всего, посмотрим на таблицу сопряженности *Пол - Gender* и *Машина - Car*. Иными словами исследуем интерес к различным типам машин у *Мужчин - Males* и *Женщин - Females*. Нажмите кнопку *Задать таблицы*, и в открывшемся диалоговом окне выберите *Пол - Gender* как единственную переменную в первом множестве, переменные *Машина_1 - Машина_3 - Car_1 - Car_3* как переменные во втором множестве.



Нажмите *OK* и вернитесь в диалоговое окно *Таблицы многомерных откликов*. Задайте далее коды для фактора *Машина - Car,* чтобы идентифицировать 4 различных типа автомобилей. Возможно, вы захотите изменить описание фактора, тогда воспользуйтесь кнопкой *Длинные метки факторов.*



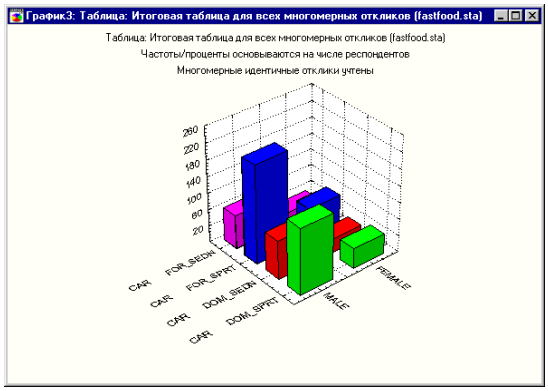
Для этой таблицы отмените опцию *Считать только уникальные отклики*. Напомним, назначение этой опции – исключить одинаковые ответы (одинаковые ответы одного и того же респондента на разные пункты считаются как один ответ). В данном примере, напротив, вы можете захотеть включить такие ответы в таблицу. Получившаяся таблица сопряженности будет показывать общее число различных типов машин, определенных респондентом как первая, либо как вторая, либо как третья, разбитых на классы значениями переменной *Пол - Gender.* Нажмите *OK* и откройте диалоговое окно *Результаты таблицы многомерных откликов*.

Copyright © StatSoft, 1995

StatSoft®

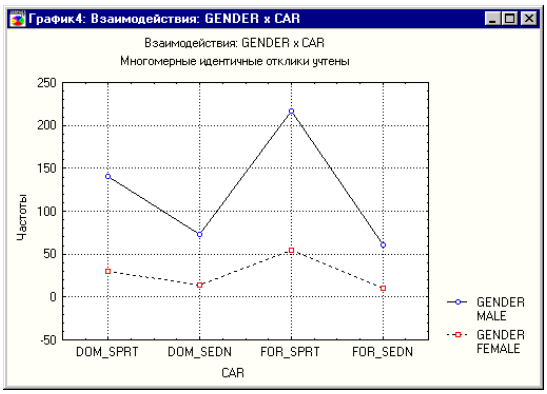
Не будем использовать опции *Процент,* а просто запросим *Просмотреть итоговые таблицы.* Тогда будет построена следующая таблица.

По умолчанию *Быстрым статистическим графиком* для этой таблицы является *3М гистограмма*. Нажмите правую кнопку мыши и выберите в меню опцию *3М гистограмма*.



Рассмотрев приведенную выше таблицу, можно придти к выводу, что и мужчины, и женщины отмечали спортивные машины чаще, чем седаны. Разницу в общем числе машин, отмеченных мужчинами и женщинами можно объяснить тем, что число мужчин и женщин в выборке существенно различается (если вы посмотрите на таблицу частот переменной *Пол - Gender,* то увидите, что в выборке присутствует только 36 женщин).

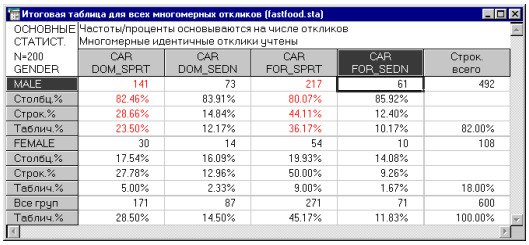
Вместо 3М гистограммы можно использовать линейный график. Вернитесь в диалоговое окно результатов и выберите опцию *Графики взаимодействий частот*.



Здесь разница в предпочтении спортивных машин более отчетлива у мужчин, чем у женщин (линия, соответствующая женщинам, более сглаженная, чем линия мужчин).

Относительные частоты: проценты по числу ответов в сравнении с процентами по числу респондентов

Так как каждый субъект дал три ответа, желательно вначале посмотреть на проценты, вычисленные относительно количества ответов (общего количества ответов, количества по строкам и по столбцам). Поэтому выберите опцию по *Числу откликов*, затем отметьте все опции *Проценты* в окне *Результаты таблицы многомерных откликов* и снова нажмите кнопку *Просмотреть итоговые таблицы,* чтобы построить следующую таблицу:



Каждый респондент в опросе назвал три машины, таким образом, общее число ответов

StatSoft®

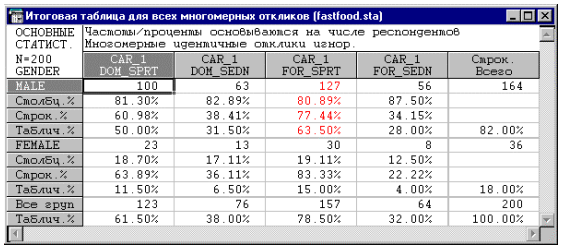
Copyright © StatSoft, 1995

равно 600. *Таблица* процентов показывает отношение частот именно к этому числу.

Например, число домашних спортивных машин, названных мужчинами, равно 141, что составляет 23.5% (100*141/600) всех ответов. Проценты в каждой строке вычислены относительно общего числа ответов в соответствующей строке. Например, 141 отечественная спортивная машина, отмеченная мужчинами, составляют 28.66% (100*141/492) всех машин, названных мужчинами. Соответственно, проценты по столбцам вычислены относительно общего количества ответов в этих столбцах: Из всех названных домашних спортивных машин 82.46% (100*141/171) были названы мужчинами.

Так как опция *Считать только уникальные отклики* отменена, то нет смысла считать проценты относительно общего числа респондентов. Напомним, что в показанной выше двумерной таблице некоторые субъекты учитываются в одной ячейке более одного раза, например, те, кто выбрал иностранную спортивную машину более одного раза.

Если вы выберете опцию *Считать только уникальные отклики* в диалоговом окне *Таблицы многомерных откликов,* а затем опцию по *Числу респондентов*, то получите итоговую таблицу следующего вида:


Итоговая таблица для всех многомерных откликов (fastfood.sta)

ОСНОВНЫЕ СТАТИСТ.
N=200
GENDER

Частоты/проценты основываются на числе респондентов
Многомерные идентичные отклики игнор.

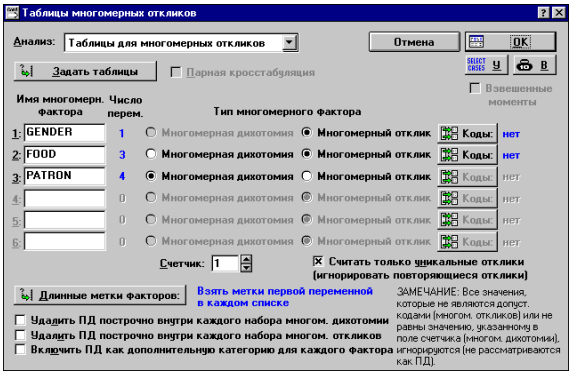
	CAR_1 DOM_SPRT	CAR_1 DOM_SEDN	CAR_1 FOR_SPRT	CAR_1 FOR_SEDN	Строк. Всего
MALE	100	63	127	56	164
Столбц.%	81.30%	82.89%	80.89%	87.50%	
Строк.%	60.98%	38.41%	77.44%	34.15%	
Таблич.%	50.00%	31.50%	63.50%	28.00%	82.00%
FEMALE	23	13	30	8	36
Столбц.%	18.70%	17.11%	19.11%	12.50%	
Строк.%	63.89%	36.11%	83.33%	22.22%	
Таблич.%	11.50%	6.50%	15.00%	4.00%	18.00%
Все груп	123	76	157	64	200
Таблич.%	61.50%	38.00%	78.50%	32.00%	100.00%

Интерпретация процентов в этой таблице отлична от предыдущей. Например, 100 мужчин указали домашние спортивные машины или как первые, или вторые, или третьи номера в своих ответах. Респонденты, представленные в этой ячейке, составляют 50% (100*100/200) от всех респондентов; 81.3% (100*100/123) от респондентов (мужчин и женщин), которые выбрали отечественную спортивную машину как первую, либо вторую, либо третью и составляют 60.98% (100*100/164) всех мужчин, участвующих в опросе.

Просмотр трехвходовой таблицы

Наконец, мы можем определить, имеются ли различия в предпочтениях ресторанов у мужчин и женщин, если они выбрали какой-то определенный любимый тип системы быстрого питания. Другими словами, теперь сопрягаются 3 переменные (фактора) *Пол - Gender, Еда - Food, Хозяин - Patron*.

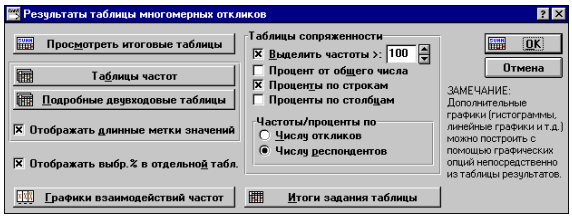
Вернемся в диалоговое окно *Таблицы многомерных откликов* и определим три фактора как выше, т.е. определим *Пол- Gender* как простую группирующую переменную, *Еда- Food* как переменную с многомерными откликами, *Хозяин - Patron* как многомерную дихотомию.



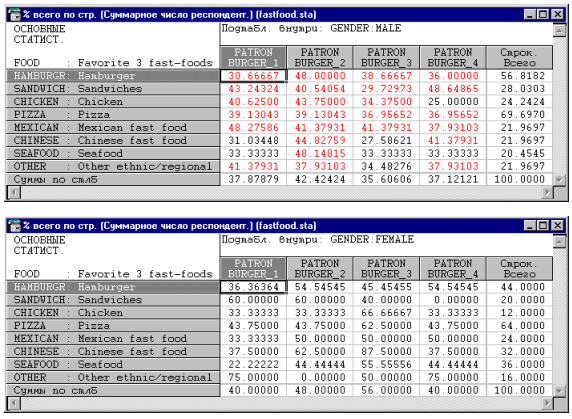
Нажмите *OK* в этом диалоговом окне и выберите в окне *Результаты* параметры: (1) *Проценты по строкам*, (2) *Частоты/проценты по числу респондентов* и (3) *Отобразить выбранные % в отдельной таблице*.

Copyright © StatSoft, 1995

StatSoft



Такой выбор позволяет просмотреть трехвходовую таблицу, которая одновременно является двувходовой для мужчин и для женщин. Кроме того, отдельно будут показаны таблицы с процентами. Нажмите на кнопку *Подробные двувходовые таблицы,* чтобы посмотреть таблицы процентов.



% всего по стр. (Суммарное число респондент.) (fastfood.sta)

ОСНОВНЫЕ СТАТИСТ. | Подгабл. внутри: GENDER:MALE

FOOD : Favorite 3 fast-foods	PATRON BURGER_1	PATRON BURGER_2	PATRON BURGER_3	PATRON BURGER_4	Строк. Всего
HAMBURGR: Hamburger	30.66667	48.00000	38.66667	36.00000	56.8182
SANDWICH: Sandwiches	43.24324	40.54054	29.72973	48.64865	28.0303
CHICKEN : Chicken	40.62500	43.75000	34.37500	25.00000	24.2424
PIZZA : Pizza	39.13043	39.13043	36.95652	36.95652	69.6970
MEXICAN : Mexican fast food	48.27586	41.37931	41.37931	37.93103	21.9697
CHINESE : Chinese fast food	31.03448	44.82759	27.58621	41.37931	21.9697
SEAFOOD : Seafood	33.33333	48.14815	33.33333	33.33333	20.4545
OTHER : Other ethnic/regional	41.37931	37.93103	34.48276	37.93103	21.9697
Суммы по стлб	37.87879	42.42424	35.60606	37.12121	100.0000

% всего по стр. (Суммарное число респондент.) (fastfood.sta)

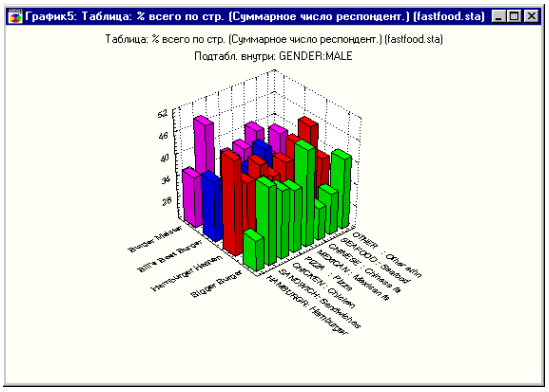
ОСНОВНЫЕ СТАТИСТ. | Подгабл. внутри: GENDER:FEMALE

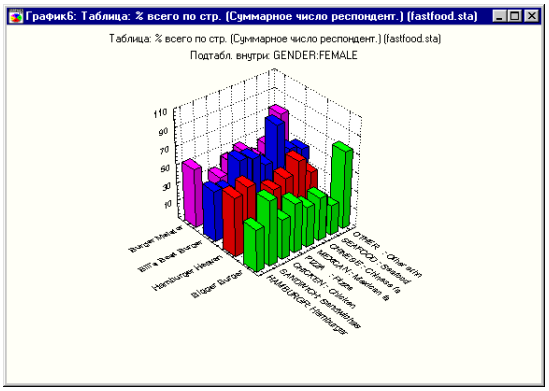
FOOD : Favorite 3 fast-foods	PATRON BURGER_1	PATRON BURGER_2	PATRON BURGER_3	PATRON BURGER_4	Строк. Всего
HAMBURGR: Hamburger	36.36364	54.54545	45.45455	54.54545	44.0000
SANDWICH: Sandwiches	60.00000	60.00000	40.00000	0.00000	20.0000
CHICKEN : Chicken	33.33333	33.33333	66.66667	33.33333	12.0000
PIZZA : Pizza	43.75000	43.75000	62.50000	43.75000	64.0000
MEXICAN : Mexican fast food	33.33333	50.00000	50.00000	50.00000	24.0000
CHINESE : Chinese fast food	37.50000	62.50000	87.50000	37.50000	32.0000
SEAFOOD : Seafood	22.22222	44.44444	55.55556	44.44444	36.0000
OTHER : Other ethnic/regional	75.00000	0.00000	50.00000	75.00000	16.0000
Суммы по стлб	40.00000	48.00000	56.00000	40.00000	100.0000

Из двувходовой таблицы *Еда - Food - Хозяин - Patron* видно, что 30.67% всех мужчин - тех, кто отметил *Гамбургер - hamburger* как одно из любимых блюд, обедали в *Burger Meister*; 48.00% обедали в *Bill's Best Burgers*, и т.д.

После просмотра процентов в этой таблице видно, что мужчины, независимо от их предпочтения в еде, с большей вероятностью обедали в *Bill's Best Burgers* (за исключением строки с *Мексиканской - Mexican* едой). В подтаблице для женщин нет такой ясной зависимости (однако, здесь результаты основываются на небольшом числе наблюдений – только 36 женщин были опрошены).

Двумерные таблицы процентов также могут быть представлены в виде 3-х мерных гистограмм, доступных из меню *Быстрые статистические графики* (щелкните на таблице результатов правой кнопкой мыши и выберите опцию *3М гистограммы* из выпадающего меню).





Заключительное замечание

Если вы недостаточно освоились с дихотомиями и переменными, имеющими многомерные отклики (факторы), возможно, первым разумным шагом является рассмотрение таблицы частот или таблицы сопряженности этих переменных. Как отмечено в разделе *Обзор* (стр. 1474), лучший способ понять устройство таких таблиц – провести кросстабуляцию простых данных и проследить, каким образом учитывается каждое наблюдение в таблице частот.

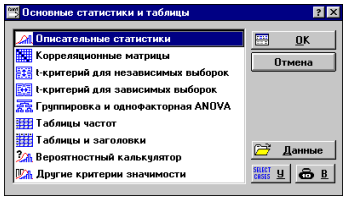
StatSoft®

Copyright © StatSoft, 1995

ДИАЛОГОВЫЕ ОКНА, ОПЦИИ, СТАТИСТИКИ

Стартовая панель

Когда вы откроете модуль *Основные статистики и таблицы*, появится стартовая панель.



Здесь можно выбрать различные аналитические процедуры.

Описательные статистики

Эти процедуры используются для того, чтобы вычислить описательные статистики, простые таблицы частот и итоговые графики. Рассматриваемые далее статистики доступны в диалоговом окне *Описательные статистики* (стр. 1522) и в диалоговом окне *Статистики* (стр. 1531).

Корреляционные матрицы

Используйте эту процедуру для того, чтобы вычислить коэффициенты корреляции Пирсона (*r Пирсона*). *STATISTICA* может вычислять как квадратные, так и прямоугольные корреляционные матрицы, используя *попарное* или *построчное* удаление пропущенных данных. Различные сопутствующие статистики могут быть также вычислены (включая уровни значимости, средние, стандартные отклонения, регрессионные коэффициенты и т.д.). Непараметрические альтернативы коэффициента корреляции Пирсона доступны при анализе таблиц сопряженности (см. диалоговое окно *Результаты кросстабуляции,* стр. 1559) а также в модуле *Непараметрическая статистика*.

t-критерий для независимых выборок

Используйте эту процедуру, чтобы вычислить *t*-критерии *для независимых выборок или групп* (например, для сравнения ответов мужчин и женщин). Предполагается, что файл организован таким образом, что каждое наблюдение представляет одного индивидуума, отмеченного группирующей переменной. Система *STATISTICA* также может вычислять *t*-критерии, если файл организован таким образом, что каждая переменная содержит данные для одной группы. Различные графики позволяют понимать результаты (см. описание *t*-критериев для независимых выборок, стр. 1537). Значимость различия между групповым средним и средним популяции может быть также проверена в процедуре *Другие критерии значимости*; здесь могут также тестироваться различия между долями. Непараметрические альтернативы *t*-критерию для независимых выборок доступны в модуле *Непараметрическая статистика* (см. главу 11).

t-критерий для зависимых выборок

Используйте эту процедуру, чтобы вычислить *t-критерии для зависимых или коррелированных выборок* (например, чтобы сравнить ответы *одного и того же* респондента до и после выборов). Можно вычислить как обычные *t*-критерии для двух выборок, так и матрицы *t*-критериев. Различные графики позволяют понять результаты (см. описание *t*-критериев для зависимых выборок, стр. 1537). Непараметрические альтернативы *t*-критерию для независимых выборок доступны в модуле *Непараметрическая статистика* (см. главу 11).

Copyright © StatSoft, 1995

StatSoft®

Группировка: внутригрупповые описательные статистики и внутригрупповые корреляции

Используйте эту процедуру, чтобы вычислить различные описательные статистики (средние, стандартные отклонения и т.д.) и корреляции внутри групп (например, внутри групп Пол и Возраст). Эта процедура позволяет выполнить полный однофакторный дисперсионный анализ *ANOVA* и проверить гипотезы об однородности дисперсий. Продолжением этого анализа являются процедуры множественного сравнения, которые позволяют проверить, какие именно средние различаются между собой (см. стр. 1542, 1551).

Таблицы частот

Используйте процедуру, чтобы вычислить и просмотреть распределение переменных. Таблицы частот можно вычислить для группирующих переменных, для переменных, принимающих целые значения и для непрерывных переменных. Можно также определить категории (классы) с помощью сложных логических условий. Тесты нормальности здесь также доступны.

Дополнительно методы подгонки распределения к данным (в том числе распределений, отличных от нормального) имеются в модуле *Непараметрическая статистика* и модуле *Анализ выживаемости*.

Таблицы и заголовки

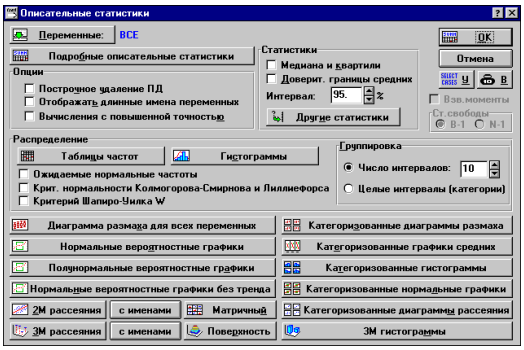
Используйте эту процедуру для вычисления многовходовых таблиц сопряженности. Система *STATISTICA* вычислит различные статистики для этих таблиц, а также построит разнообразные двух- и трехмерные итоговые графики (см. диалоговое окно *Результаты кросстабуляции*, стр. 1559). Разнообразные средства позволяют провести анализ не только простых группирующих переменных, но также переменных с многомерными откликами и многомерных дихотомий (см. стр. 1474 и 1563).

Другие критерии значимости

Процедура вычисляет тесты для сравнения средних, корреляции, процентов и т.д. (см. стр. 1567). Система *STATISTICA* также включает *Вероятностный калькулятор* с широким набором непрерывных функций распределения. *Вероятностный калькулятор* позволяет вычислить уровень значимости для коэффициента корреляции Пирсона и преобразовать его в *z*-значения Фишера.

Описательные статистики

Обсуждение описательных статистик приведено в обзорном разделе этой главы (стр. 1447). Когда вы выберете *Описательные статистики* в стартовой панели (см. выше), откроется диалоговое окно *Описательные статистики*.



Это диалоговое окно содержит опции задания переменных (и опции группировки) и просмотра результатов.

Переменные

Опция открывает диалоговое окно выбора переменных для анализа.

StatSoft®

Copyright © StatSoft, 1995

Подробные описательные статистики

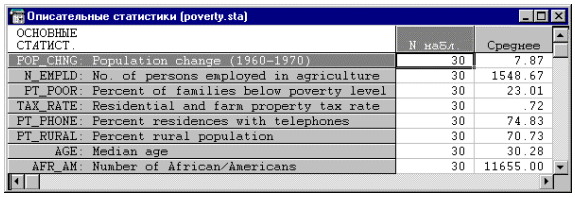
Опция строит таблицы результатов с описательными статистиками выбранных переменных.

Дополнительные статистики запрашиваются в диалоговом окне *Статистики* с помощью кнопки *Дополнительные статистики*.

Опции

Построчное удаление ПД. Если выбрано построчное удаление пропущенных данных, то система *STATISTICA* игнорирует все наблюдения, имеющие пропуски в какой-либо переменной в заданном списке. Если опция не выбрана, то производится попарное удаление пропусков.

Показать длинные имена переменных. Если эта опция отмечена, в таблице результатов показываются длинные имена переменных (если они имеются) вместе с короткими именами.



Вычисления с повышенной точностью. Система *STATISTICA* позволяет проводить вычисления с повышенной точностью с тем, чтобы минимизировать ошибки округления. Опция полезна, если переменная имеет малую *относительную* вариацию (вариация, деленная на среднее), например, если все значения переменной находятся между 10,000.000001 и 10,000.000002. Заметим, что имеется различие между малой вариацией и малой *относительной* вариацией (например, *относительная* вариация значений от 0000000000000001 до .0000000000000002 не мала). Для очень больших файлов данных со многими тысячами наблюдений, вычисления будут несколько медленнее, если выбрана эта опция.

Статистики

Заметим, что дополнительные описательные статистики (*мода*, *гармоническое среднее*, *геометрическое среднее*, *среднее отклонение* и др.) доступны в модуле *Непараметрическая статистика*.

Медиана и квартили. Если опция выбрана, то в таблице, которая строится с помощью кнопки *Подробные описательные статистики* будет показана *медиана* (50% процентиль) и *квартили* (25% и 75% процентили) для всех переменных. Если значения переменной нельзя разбить точно пополам и на четверти, используются различные методы оценивания медианы и квартилей. Определенный метод можно выбрать в диалоговом окне *STATISTICA Параметры по умолчанию: общие настройки* (см. описание этого окна в главе 3, окно доступно с помощью опции меню *Общие настройки*).

Доверительные границы для среднего. Идея *доверительных границ* для среднего обсуждается в разделе *Вводный обзор*, стр. 1447. Эта опция позволяет вычислить границы доверительного интервала с выбранным уровнем доверия *альфа.*

Дополнительные статистики. Кнопка открывает диалоговое окно (см. стр. 1531) выбора дополнительных описательных статистик для таблицы *Подробные описательные статистики.* Доступны статистики: число наблюдений n, среднее, сумма, медиана, стандартное отклонение, дисперсия, стандартная ошибка среднего, 95% доверительные границы для среднего, минимум, максимум, нижние и верхние квартили, размах,

Copyright © StatSoft, 1995

StatSoft®

квартильный размах, асимметрия, эксцесс, стандартные ошибки асимметрии и эксцесса (дополнительные статистики имеются в модуле *Непараметрическая статистика*).

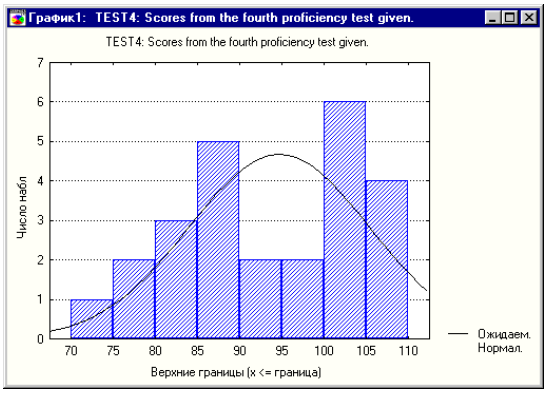
Распределение

Таблицы частот. Опция строит каскад таблиц результатов с частотами распределения для выбранных переменных (одна таблица - одна переменная). Способ группировки зависит от выбора в поле *Группировка* (справа): Диапазон изменения переменной может быть разделен на определенное *Число интервалов* (используйте этот способ, если переменные непрерывны).

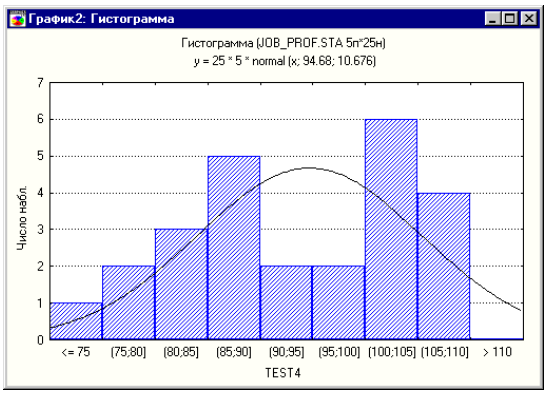
Другой способ – использовать *Целые интервалы (категории).* Если эта опция выбрана, то: (1) - все нецелые значения переменной игнорируются, (2) - не проводится проверка нормальности (см. ниже).

Гистограммы. Опция строит каскад гистограмм, аналогично процедуре *Таблицы частот*. Если выбрана опция *Ожидаемые нормальные частоты* (см. ниже), то на гистограмму будет наложена плотность нормального распределения с оцененными параметрами.

Нормальная аппроксимация. Эта опция доступна лишь тогда, когда отмечена опция *Группировка*. Если такой выбор сделан, то в построенной таблице результатов будут содержаться оцененные нормальные частоты для каждой категории (группы).



Заметим, опция построения гистограмм, доступная из меню *Быстрые статистические графики* (вызываемого правой кнопкой мыши, см. том II) дает несколько другую компоновку графика, описание функции подгонки в заголовке (см. ниже) содержит информацию о среднем и стандартном отклонении переменной (94.68 и 10.676).



Различные функции распределения можно подогнать к данным, используя модули *Непараметрическая статистика и распределения* (см. глава 7), *Анализ процессов* (см. том IV), а также опцию *Статистические 2М графики - Гистограммы* (см. том II); распределения для данных типа длительности жизни имеются в модуле *Анализ выживаемости.*

Критерий нормальности Колмогорова-Смирнова. Эта опция доступна, если выбрана

StatSoft®
Copyright © StatSoft, 1995

опция *Группировка*. Тогда последовательность таблиц частот будет включать результаты одновыборочного критерия нормальности Колмогорова-Смирнова. Если *D* статистика значима, то гипотеза о нормальности распределения значений переменной отвергается. Вычисляются два значения вероятности для каждого критерия: первое, табулировано Massey (1951), относится к случаю, когда параметры распределения – среднее и стандартное отклонение известны заранее и не оцениваются. Однако обычно эти параметры оцениваются по наблюдениям. Тогда проверяется сложная гипотеза (с неизвестными заранее параметрами распределения, и вычисляются вероятности *Лиллиефорса* (Lilliefors, 1967). В последние годы также рассматривается W-критерий Шапиро-Уилка (см. ниже), как подходящий для проверки гипотезы нормальности (см. Shapiro, Wilk, and Chen, 1968).

W критерий Шапиро-Уилка. Опция доступна только при выборе опции *Группировка*. В этом случае последовательность таблиц частот будет включать результаты одновыборочного теста нормальности W Шапиро-Уилка. Если *W*-статистика значима, то гипотеза нормальности отвергается. Заметим, что в *STATISTICA* используется улучшенная версия алгоритма (Royston (1982)), которая применима к выборкам, содержащим до 2000 наблюдений; если число наблюдений больше 2000, этот критерий неприменим.

Группировка

Эти процедуры дают эффект только для таблиц частот и гистограмм, построенных с помощью процедур из раздела *Распределение* (слева). Опция *Описательные статистики* имеет только два режима группировки значений выбранных переменных для таблицы частот.

Число интервалов. Если эта опция выбрана, то в таблицах частот или гистограммах диапазон изменений переменных будет разбит на требуемое число интервалов. Опция используется, когда табулируются непрерывные переменные. Критерии нормальности доступны только для этой опции. Заметим, что число групп иногда может отличаться от заданного, т.к. *STATISTICA* строит “приближенные” интервалы - границы интервала и ширина округляются и имеют последнюю десятичную цифру 1, 2, или 5 (например, 10.5, 11.0, 11.5, и т.д.). Такие интервалы легче интерпретировать, чем интервалы, содержащие много десятичных знаков (например, 10.12423, 10.13533, и т.д.). Полный контроль над методом группировки переменных доступен с помощью процедуры *Таблицы частот* (см. стр. 1521).

Целые интервалы. Выберите этот метод группировки, если табулированные переменные можно интерпретировать как состоящие из целых группы, или содержащие только целые значения. В этом методе все нецелые значения игнорируются.

Диаграмма размаха для всех переменных

Процедура построит диаграмму размаха для выбранных переменных.

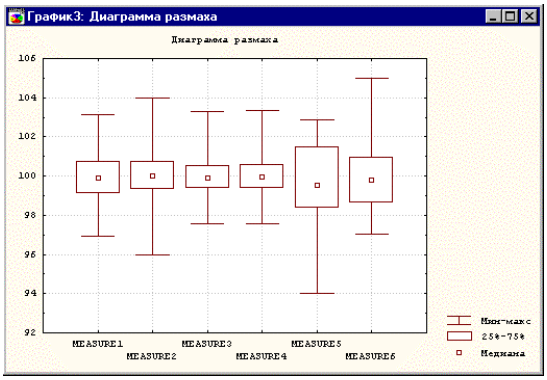


График включает три компоненты:

Copyright © StatSoft, 1995

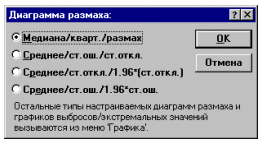
StatSoft®

(1) центральную точку, показывающую главную тенденцию или положение (*медиана* на графике вверху);

(2) прямоугольник, показывающий разброс значений относительно главной тенденции (*25% -75% процентили*);

(3) отрезки вокруг прямоугольника, показывающие диапазон значений переменной (*Мин-макс*).

После нажатия кнопки *Диаграмма размаха* откроется диалоговое окно:

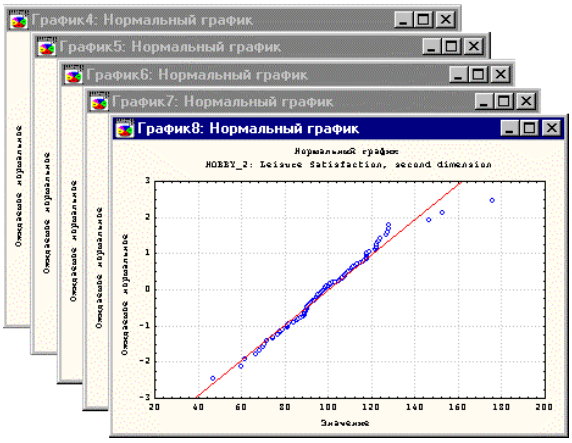


Здесь можно выбрать: (1) Медиану (точка центра выборки), квартили (прямоугольник), размах (отрезки); (2) Средние, стандартные ошибки средних, стандартные отклонения; (3) Средние, стандартные отклонения, стандартные отклонения, умноженные на 1.96 (95% доверительный интервал для среднего от нормального распределения); (4) Средние, стандартные ошибки среднего, стандартные ошибки среднего, умноженные на 1.96.

Медиана и квартили вычисляются согласно установкам, сделанным в поле *Процентили* в диалоговом окне *STATISTICA Параметры по умолчанию: общие настройки*. Вычислительные детали приведены в главе 3 (см. также Приложение I).

Нормальные вероятностные графики

Процедура строит последовательность нормальных вероятностных графиков – один график для одной выбранной переменной.



Стандартный нормальный вероятностный график строится следующим образом. Вначале отклонения от соответствующих средних (*остатков*) ранжируются. По этим рангам вычисляются *z* значения (стандартизованные значения нормального распределения). *z* значения отмечаются на оси *Y*. Если наблюдаемые значения (отложенные по оси *X*) нормально распределены, то все значения попадут на прямую линию. Если распределение отличается от нормального, то будет наблюдаться отклонение от прямой.

На этом графике можно отчетливо увидеть выбросы. Если подгонка неудачна и данные имеют определенную форму расположения вокруг прямой (например, в виде буквы *S*), можно попытаться преобразовать переменные различным способом (например, прологарифмировать, см. *Изменение (текущих) спецификаций таблицы исходных данных* (глава 3, или *STATISTICA BASIC*, том V).

Вероятностные графики для распределений, отличных от нормальных. Вероятностные графики можно построить для различных непрерывных распределений, отличных от нормального (Вейбулла, логнормального, бета, экспоненциального и др.) с помощью процедуры *Графики вероятность - вероятность* в меню

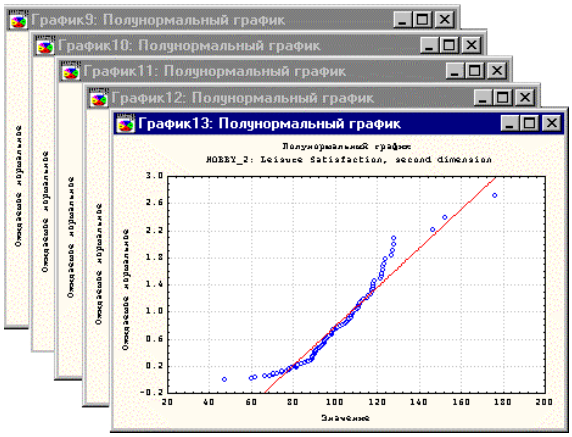
StatSoft®

Copyright © StatSoft, 1995

Статистические 2М графики. Детали см. в описании *Статистические 2М графики,* том II.

Полунормальные вероятностные графики

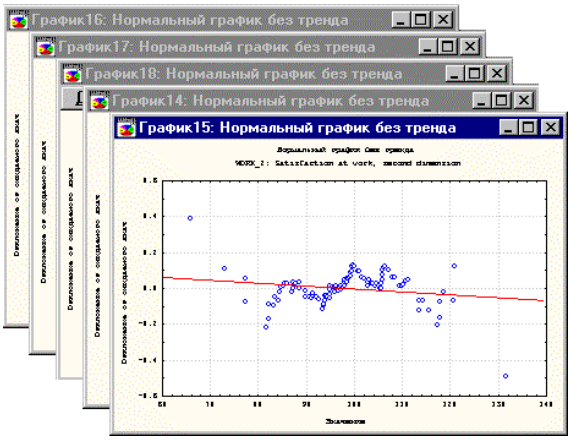
Опция строит последовательность полунормальных вероятностных графиков – по одному графику для каждой выбранной переменной.



Единственное отличие этих графиков от нормальных вероятностных графиков в том, что рассматривается только правая часть нормальной плотности. Только положительные нормальные значения откладываются на оси *Y*. Этот график используется, когда хотят игнорировать знак остатков (т.е. рассматривают только их абсолютные значения).

Нормальные вероятностные графики без тренда

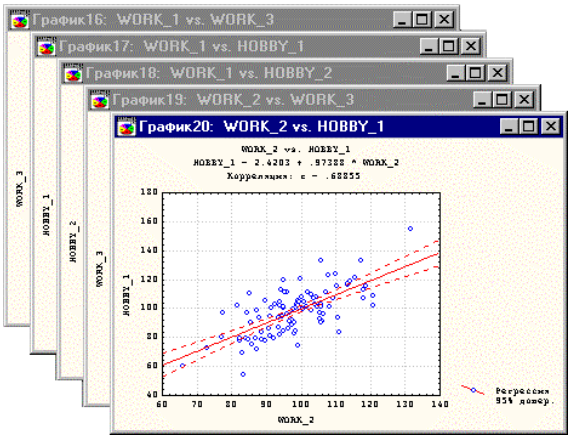
Опция строит последовательность нормальных вероятностных графиков с исключенным трендом - по одному графику для каждой выбранной переменной.



Единственное отличие этих графиков от нормальных – исключение линейного тренда из данных.

2М диаграмма рассеяния (с именами)

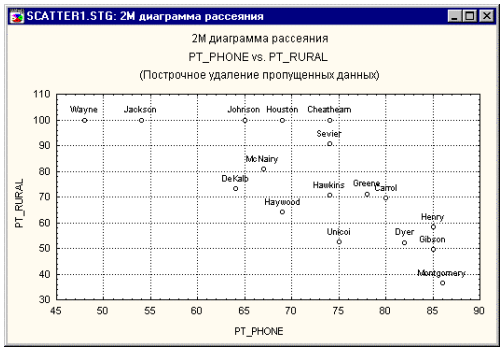
Опция строит последовательность диаграмм рассеяния с именами или номерами наблюдений – по одному графику для каждой выбранной переменной.



Если выбрана опция *2М диаграмма рассеяния* (см. график выше), то на графике будет изображена подогнанная регрессионная прямая и гиперболические доверительные границы прогноза.

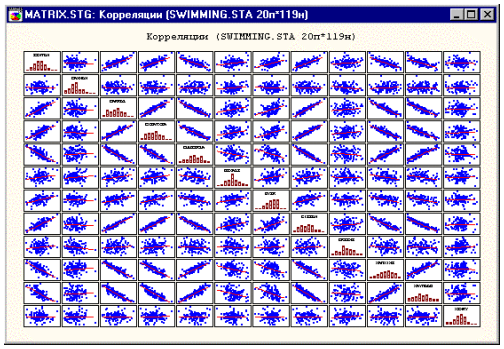
Copyright © StatSoft, 1995

StatSoft®



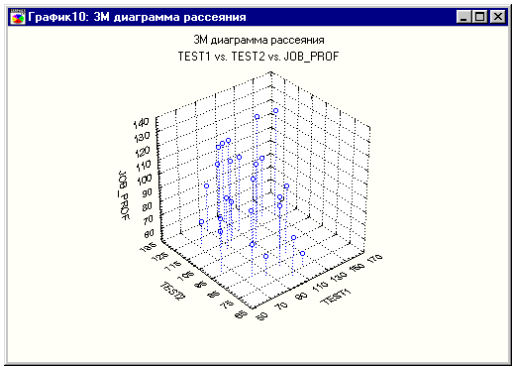
Матричный

Опция строит матрицу диаграмм рассеяния для выбранных переменных.

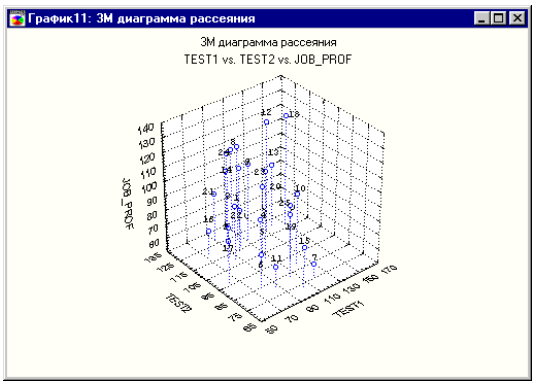


3M диаграмма рассеяния (с именами)

Опция строит 3-х мерную диаграмму рассеяния для выбранных переменных.

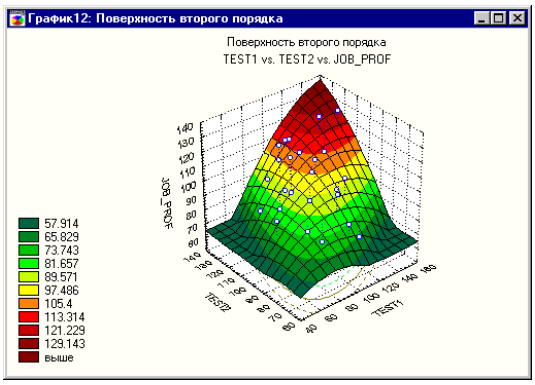


Если использована кнопка *с именами*, точки на диаграмме рассеяния будут помечены номерами или именами соответствующих наблюдений (если наблюдения имеют имена).



Поверхность

Эта опция строит 3-х мерную диаграмму рассеяния для выбранной тройки переменных вместе с подогнанной поверхностью второго порядка.

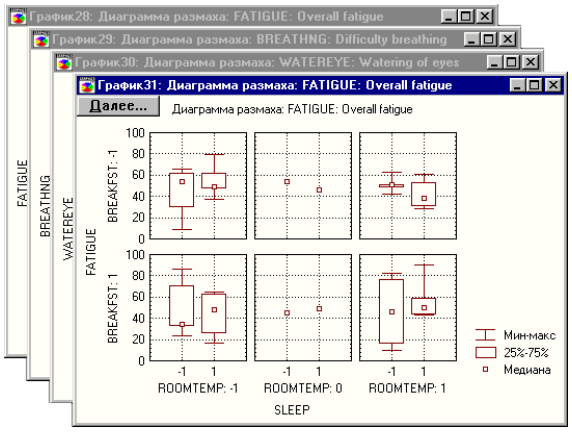


Конкретные наблюдения показаны на графике отрезками, соединенными с подогнанной поверхностью. Этот метод разведочного анализа иногда бывает полезен для выделения “скрытых” аспектов зависимостей между несколькими переменными (как подгонять другие типы поверхностей, см. подробнее, например, том II).

StatSoft®
Copyright © StatSoft, 1995

Категоризованные диаграммы размаха

Эта опция строит каскад категоризованных диаграмм размаха для выбранных переменных, по одной диаграмме для каждой выбранной переменной.

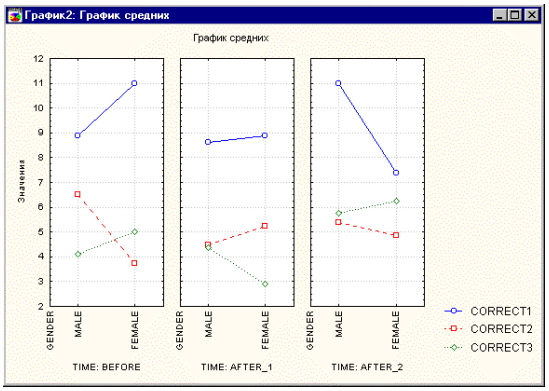


После нажатия кнопки, программа попросит ввести до трех группирующих переменных, (например *Пол - Gender*, *Национальность - Ethnic Origin*). Для каждой выбранной (с помощью кнопки *Переменные*) переменной *STATISTICA* построит диаграмму размаха раздельно по группам (например, раздельно для мужчин и женщин каждой национальности). Заметим, что анализ с более сложной *Группировкой* анализируемых групп возможен при выборе процедуры *Группировка и однофакторная ANOVA* панели запуска (см. стр. 1521).

Возможно построение категоризованных диаграмм размаха с помощью процедуры *Статистические графики*, доступной из выпадающего меню *Графика* панели инструментов таблицы исходных данных или таблицы результатов в любой момент проведения анализа. Эта опция также предлагает многочисленные дополнительные возможности аналитического и графического вывода результатов.

Категоризованные графики (взаимодействий) средних

Опция строит стандартные графики взаимодействий средних для одной, двух или трех группирующих переменных (например, *Пол - Gender*, *Национальность - Ethnic Origin*).



После выбора этой опции программа просит пользователя задать группирующие переменные. Это позволяет отобразить на графике взаимодействия средние для наблюдений, разделенных на 1-3 категории. (Отметим, что каскад графиков взаимодействий, показывающих точные взаимодействия, может быть построен с помощью процедуры *Группировка: Внутригрупповые описательные статистики*, см. стр. 1542.)

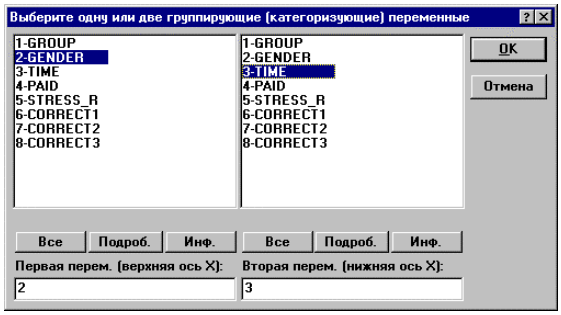
Если с помощью окна *Переменные* была выбрана всего одна (“зависимая”) переменная, пользователь может дополнительно задать до трех группирующих переменных. Если в окне *Переменные* было задано более одной переменной, то для группировки наблюдений может быть выбрано до двух группирующих переменных; причем различные (зависимые) переменные будут представлены на графиках различными типами линий (или разными цветами).

Сопоставление факторов осям на графике взаимодействий. Это диалоговое окно позволяет соотносить факторы

Copyright © StatSoft, 1995

StatSoft

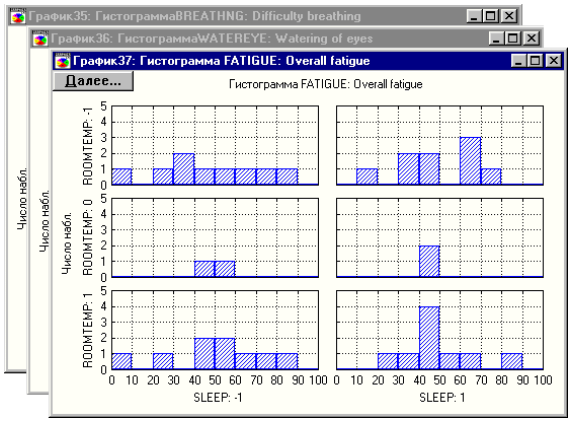
(группирующие переменные) с типом линий, а также выбирать факторы, соответствующие верхней и нижней полуосям оси *X*.



Если на графике изображается более трех факторов, график строится “порциями” между уровнями других (не выбранных) факторов.

Категоризованные гистограммы

Эта опция строит каскад категоризованных гистограмм для выбранных переменных, по одному на каждую выбранную переменную.

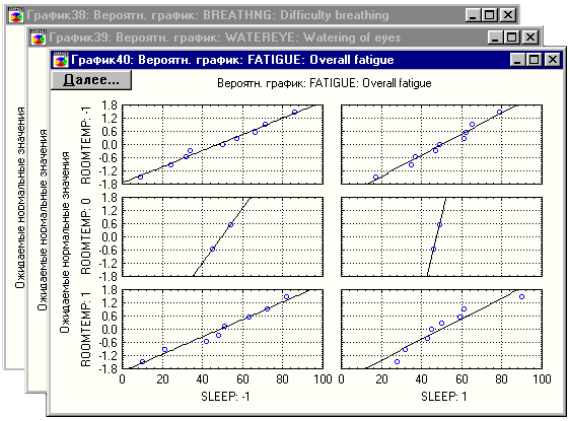


После нажатия этой кнопки (или кнопки *Категоризованные нормальные графики*, см. ниже), программа попросит вас выбрать до двух группирующих переменных (например, *Пол - Gender*, *Национальность - Ethnic Origin*). Для каждой выбранной (с помощью кнопки *Переменные*) переменной, *STATISTICA* строит гистограммы, раздельно по каждой категории (например, раздельно для мужчин и женщин каждой национальности). Отметим, что более сложная *Группировка* и графики доступны при помощи процедуры *Группировка: Внутригрупповые описательные статистики* из стартовой панели (стр. 1521).

Категоризованные графики также можно построить с помощью процедуры *Статистические графики*, доступной из выпадающего меню *Графика* панели инструментов таблицы исходных данных или таблицы результатов в любой момент проведения анализа. Эта опция также предлагает многочисленные дополнительные возможности аналитического и графического вывода результатов.

Категоризованные нормальные графики

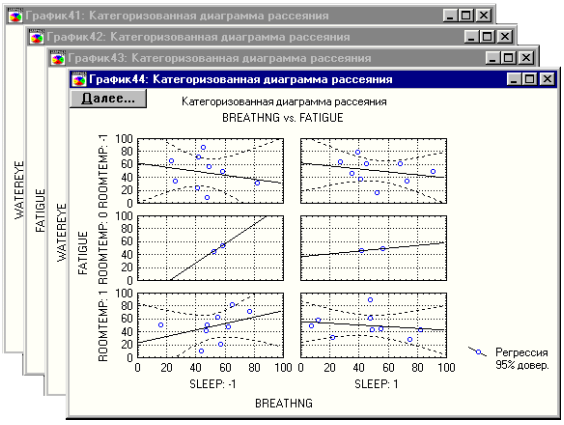
Эта опция строит каскад категоризованных нормальных вероятностных графиков для выбранных переменных, по одному на каждую выбранную переменную.



StatSoft®
Copyright © StatSoft, 1995

Категоризованные диаграммы рассеяния

Эта опция строит каскад категоризованных диаграмм рассеяния для выбранных переменных, по одному на каждую выбранную переменную.

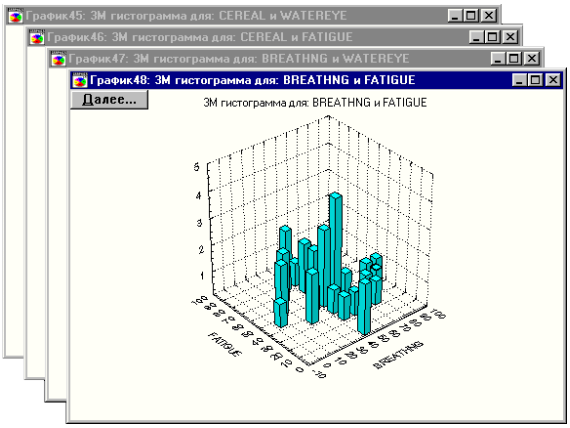


После нажатия соответствующей кнопки программа попросит пользователя выбрать два набора переменных (из выбранных ранее с помощью кнопки *Переменные*). Диаграммы рассеяния будут построены для каждой пары переменных, включающей переменные из разных списков. Затем программа попросит задать до двух группирующих переменных (например, *Пол -Gender*, *Национальность - Ethnic Origin*). После этого *STATISTICA* построит каскад диаграмм рассеяния раздельно по каждой категории (например, раздельно для мужчин и женщин каждой национальности). Отметим, что более сложная *Группировка* и графики доступны при помощи процедуры *Группировка: Внутригрупповые описательные статистики* стартовой панели (стр. 1521).

Категоризованные диаграммы рассеяния также можно построить с помощью процедуры *Статистические графики*, доступной из выпадающего меню *Графика* панели инструментов таблицы исходных данных или таблицы результатов в любой момент проведения анализа. Эта опция также предлагает многочисленные дополнительные возможности аналитического и графического вывода результатов.

3М гистограммы

Эта опция строит каскад 3М гистограмм для выбранных переменных, по одному на каждую выбранную переменную.



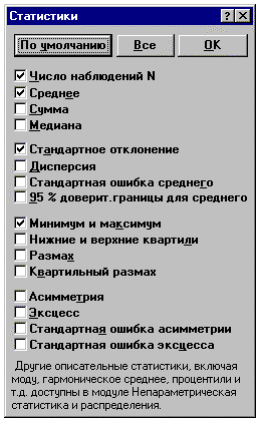
После нажатия этой кнопки программа попросит пользователя выбрать два набора переменных (из выбранных ранее с помощью кнопки *Переменные*). 3М гистограммы будут построены для каждой пары переменных, включающей переменные из разных списков.

Диалоговое окно: Статистики

Эта настройка определяет, какие статистики будут вычислены, когда пользователь выберет *Подробные описательные статистики* из диалогового окна *Описательные статистики* (см. стр. 1522).

Copyright © StatSoft, 1995

StatSoft



Вычисление медианы и квартилей (25%, 50%, 75% процентилей). Если значения переменной нельзя разбить точно пополам и на четверти, используются различные методы оценивания медианы и квартилей. Определенный метод можно выбрать в диалоговом окне *STATISTICA Параметры по умолчанию: общие настройки* (см. описание этого окна см. в главе 3, окно также доступно с помощью опции *Общие настройки*).

По умолчанию

Этот параметр задает вычисление статистик: *Число наблюдений n*, *Среднее*, *Стандартное отклонение*, *Минимум & Максимум*.

Все

Будут вычислены все статистики.

Число наблюдений n

Вычисляется число наблюдений для каждой выбранной переменной.

Среднее

Вычисляет средние выбранных переменных.

Сумма

Вычисляет сумму.

Медиана

Вычисляет медиану. Медиана (50% процентиль) - это значение, которое делит распределение всех значений пополам; 50% всех значений меньше медианы, 50% - больше медианы.

Стандартное отклонение

Стандартное отклонение вычисляется как корень квадратный из суммы квадратов отклонений (от среднего), деленный на *n-1*.

Дисперсия

Вычисляется дисперсия переменной как суммы квадратов отклонений (от среднего) деленная на *n-1*.

Стандартная ошибка среднего

Стандартная ошибка среднего вычисляется как стандартное отклонение, деленное на корень квадратный из *n*.

95% доверительные границы для среднего

Доверительные границы вычисляются с помощью соответствующих значений *t* критерия. Обратитесь к разделу *Вводный обзор* (стр. 1447). Уровень доверия можно изменить с помощью параметра *Интервал* в диалоговом окне *Описательные статистики* (см. стр. 1523).

Минимум и максимум

Вычисляет максимум и минимум.

Нижняя и верхняя квартили

Вычисляется 25% и 75% процентиль (нижняя и верхняя квартиль*)*. Определенный метод вычисления можно выбрать в диалоговом окне

StatSoft®

Copyright © StatSoft, 1995

STATISTICA Параметры по умолчанию: общие настройки (см. описание этого окна главе 3, окно также доступно с помощью опции *Общие настройки*).

Размах

Вычисляется размах: максимум минус минимум.

Квартильный размах

Вычисляется квартильный размах, как верхняя квартиль минус нижняя квартиль.

Асимметрия

Асимметрия - мера симметричности распределения. Если распределение симметричное, асимметрия равна 0.

Эксцесс

Эксцесс - мера “остроты” пика распределения. Если распределение - стандартное нормальное, эксцесс равен 0.

Стандартная ошибка асимметрии

Вычисляется стандартная ошибка асимметрии.

Стандартная ошибка эксцесса

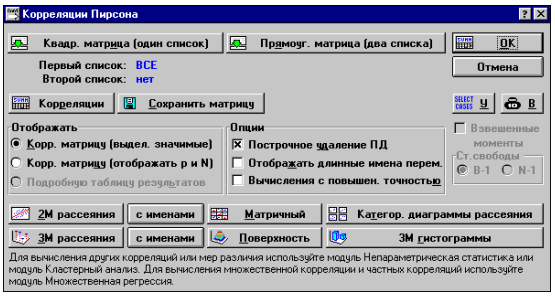
Вычисляется стандартная ошибка эксцесса.

Другие статистики

Другие описательные статистики (включая моду, гармоническое среднее, геометрическое среднее, и другие) доступны в модуле *Непараметрическая статистика* (см. главу 11).

Корреляционные матрицы

По поводу корреляционных матриц обратитесь к главе *Вводный обзор*, стр. 1449.



Квадратная матрица (один список переменных)

Опция открывает диалоговое окно, где можно выбрать единственный список переменных. Используется для вычисления квадратных корреляционных матриц. Эти матрицы можно сохранить (с помощью кнопки *Сохранить матрицу*, см. ниже) в специальном формате *STATISTICA* “матричный формат исходных данных” (см. *Управление данными*, глава 7).

Прямоугольная матрица (два списка)

Опция открывает диалоговое окно, где можно выбрать два списка переменных. Используется для вычисления прямоугольных корреляционных матриц. Эти матрицы также можно сохранить (с помощью кнопки *Сохранить матрицу*, см. ниже) в специальном формате *STATISTICA* “матричный формат исходных данных” (см. *Управление данными*, глава 7).

Корреляции

Вычисляется таблица результатов с коэффициентами корреляции между выбранными переменными. Формат результатов

Copyright © StatSoft, 1995

StatSoft

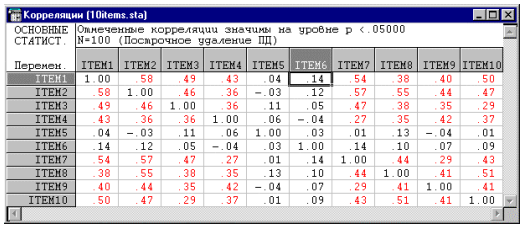
зависит от выбора установок *Экран* и *Сервис* (см. ниже).

Сохранить матрицу

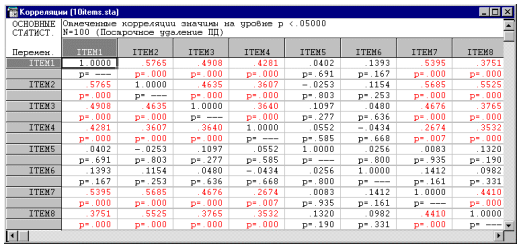
Опция сохраняет *Квадратные матрицы* корреляций (см. выше) в файле. Информация о матричных форматах *STATISTICA* дана в главе 7, *Управление данными*.

Отображать

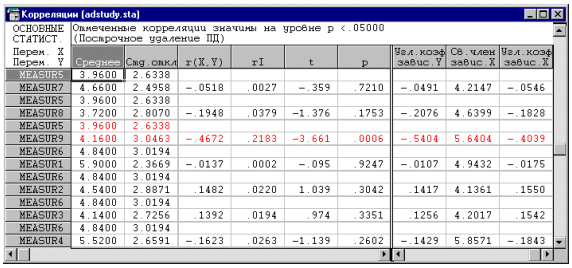
Корр. матрицу (выделить значимые). Если эта опция установлена, то после нажатия кнопки *Корреляции* (см. выше) будет построена матрица с коэффициентами, выделенными на уровне *p*.



Корр. матрицу (отображать *p* и *n*). Если опция выбрана, то, нажав кнопку *Корреляции* (см. выше), вы построите матрицу, которая вместе с коэффициентами корреляции содержит также *p*-уровни и число наблюдений n.



Подробную таблицу результатов. Если опция выбрана, то, нажав кнопку *Корреляции* (см. выше), вы построите подробную таблицу результатов (не в матричном формате). Таблица содержит не только коэффициенты корреляции, но также средние (после парного удаления пропусков), стандартные отклонения (после попарного удаления пропусков), *r-квадрат*, *t* значение, *p* значение, объем выборки, коэффициенты уравнения регрессии и свободный член в уравнении регрессии .



Эта опция доступна, если выбрано 20 или менее переменных (в противном случае она неактивна). Опция используется для исследования только некоторых определенных корреляций (не для разведочного анализа).

Другие опции

Вычисления с повышенной точностью. *STATISTICA* может провести вычисления с повышенной точностью, когда необходима высокая точность для “трудных” данных. Используйте опцию, когда переменные имеют малую относительную вариацию (например, отношение стандартного отклонения к среднему меньше .0000000000001).

Построчное удаление пропущенных данных. Если выбрано построчное удаление пропущенных данных, то *STATISTICA* игнорирует все наблюдения, имеющие пропуски, для какой-либо переменной в списке. Если опция не выбрана, то производится попарное удаление пропусков.

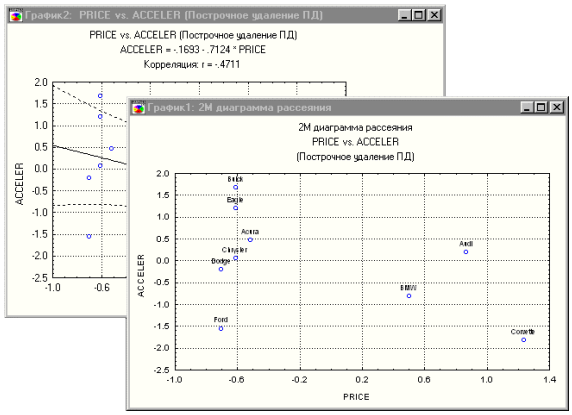
Отображать длинные имена переменных. Опция позволяет посмотреть длинные имена переменных, если они имеются, в таблице результатов.

StatSoft®

Copyright © StatSoft, 1995

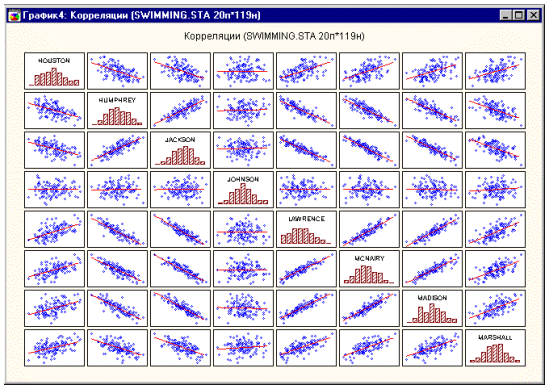
2M диаграмма рассеяния (с именами)

Опция строит последовательность диаграмм рассеяния с именами или номерами наблюдений – один график для каждой выбранной переменной. На каждом графике будет изображена подогнанная регрессионная прямая и гиперболические доверительные границы прогноза.



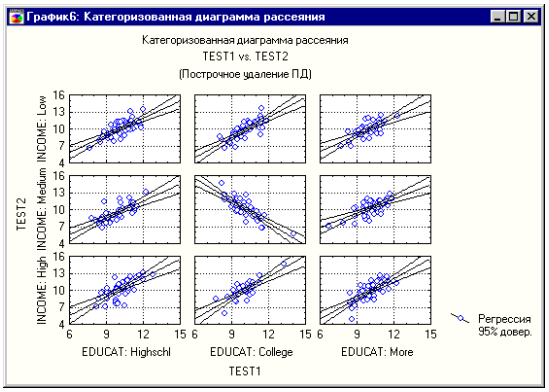
Матричный

Эта опция строит матрицу диаграмм рассеяния для выбранных переменных.



Категоризованные диаграммы рассеяния

Эта опция строит каскад категоризованных диаграмм рассеяния для выбранных переменных, по одной на каждую пару (т.е. на каждую корреляцию).



Если ранее был выбран только один набор переменных - квадратная матрица (один список переменных), то после нажатия кнопки программа запросит выбрать два набора переменных, из тех, что были выбраны ранее. Диаграммы рассеяния будут построены для каждой пары переменных. Затем будет предложено задать до двух группирующих переменных (например, *Пол - Gender*, *Национальность - Ethnic Origin*).

После этого *STATISTICA* построит каскад диаграмм рассеяния раздельно по каждой группе (например, раздельно для мужчин и женщин каждой национальности). На каждом графике будет изображена подогнанная к данным регрессионная прямая и гиперболические доверительные границы прогноза. Отметим, что более сложная *Группировка* и графики доступны с помощью процедуры *Группировка: внутригрупповые описательные статистики* (стр. 1542).

Категоризованные диаграммы рассеяния также можно построить с помощью процедуры *Статистические графики*, доступной из

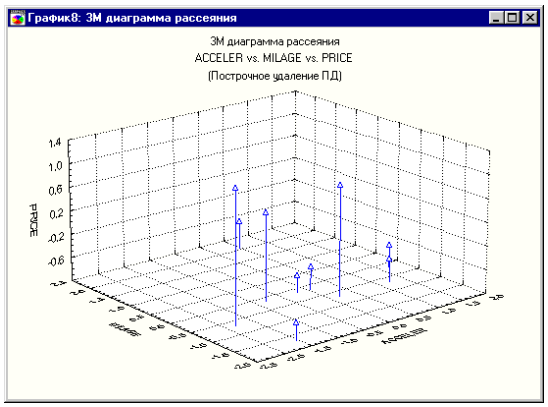
Copyright © StatSoft, 1995

StatSoft®

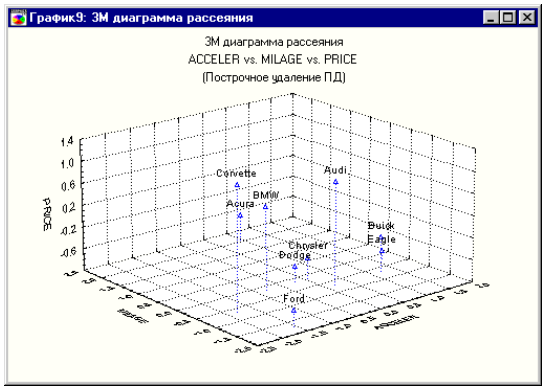
выпадающего меню *Графика* панели инструментов таблицы исходных данных или таблицы результатов в любой момент проведения анализа. Эта опция предлагает многочисленные дополнительные возможности аналитического и графического вывода результатов.

3М диаграмма рассеяния (с именами)

Эта опция строит 3-х мерную диаграмму рассеяния для выбранной тройки переменных.

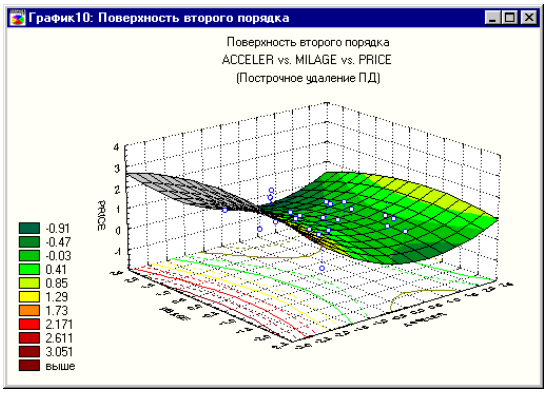


Если использована кнопка *с именами*, то точки на диаграмме рассеяния будут отмечены номерами или именами соответствующих наблюдений (если наблюдения имеют имена).



Поверхность

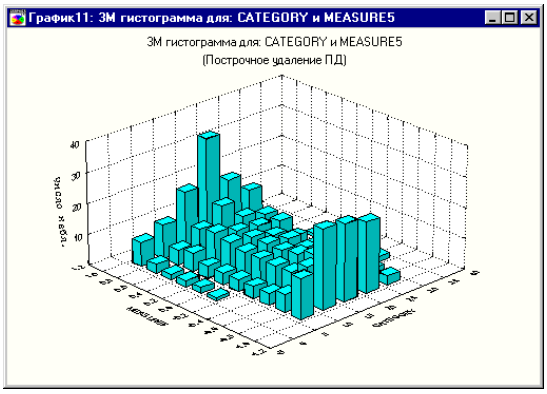
Эта опция строит 3-х мерную диаграмму рассеяния для выбранной тройки переменных с подогнанной поверхностью второго порядка.



Применение графиков подогнанных поверхностей при проведении разведочного анализа данных обсуждается в томе II, *Статистические 3М графики*.

3М гистограммы

Эта опция строит каскад 3-х мерных гистограмм для выбранных переменных, по одной на каждую пару.



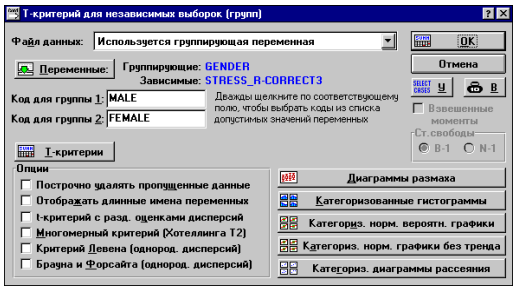
Если ранее был выбран один набор переменных (*Квадратная матрица (один список переменных)*), программа попросит пользователя

StatSoft®
Copyright © StatSoft, 1995

выбрать две группы переменных из выбранных ранее для анализа, а после этого построит 3-х мерные гистограммы для каждой пары переменных, включающей переменные из разных списков.

t-критерий для независимых выборок

Общий обзор *t*-критериев для независимых выборок приведен в разделе *Вводный обзор* (стр. 1458).



Файл данных

Имеется два варианта задания исходных данных для применения t-критерия для независимых выборок: *Используется группирующая переменная* и *Каждая переменная содержит данные для одной группы*. По умолчанию, произведена установка, в которой каждая строка в файле представляет одно наблюдение (например, индивидуума, респондента), а каждый столбец – переменную или измерение (например, ответ на позицию анкеты). В этой установке группирующая переменная (например, *Пол - Gender*) должна быть включена в анализ, чтобы определить, к какой группе относится каждое наблюдение.

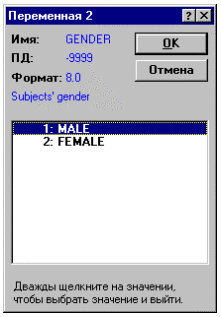
Принятая по умолчанию установка в окне предполагает, что данные упорядочены именно таким образом. Однако для простого анализа иногда бывает легче ввести данные для каждой группы отдельно, в столбцы файла (например, данные для мужчин в первую колонку, для женщин – во вторую). Если данные организованы таким образом, сделайте установку *Каждая переменная содержит данные для одной группы* (см. стр. 1539).

Переменные

При нажатии кнопки открывается окно выбора двух переменных. Выберите одну группирующую переменную (например, Пол) и список зависимых переменных. *STATISTICA* вычислит *t-критерий* для всех зависимых переменных и сравнит две группы, определяемые групповыми кодами (см. ниже).

Код для группы 1; код для группы 2

Задайте два кода, чтобы идентифицировать две группы, которые сравниваются между собой. Если вы не уверены в кодах, дважды щелкните мышью на поле ввода, чтобы увидеть окно (см. *Окно значений переменных*, глава 3), в котором показаны коды группирующей переменной в текущем файле данных.

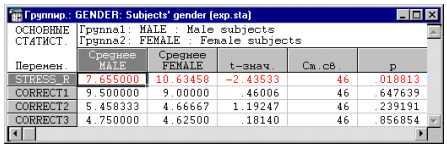


Здесь можно выбрать код, дважды щелкнув по нему мышью.

T-критерии

После нажатия этой кнопки программа вычислит значения *t*- критериев для независимых выборок и представит результаты в таблице.

Copyright © StatSoft, 1995



Представление результатов зависит от выбора параметров в поле *Опции* диалогового окна (см. ниже).

Опции

Выбор опций определяет вывод и формат результатов *t*-критерия для независимых выборок.

Построчно удалять пропущенные данные. Если выбрано построчное удаление пропущенных данных, то *STATISTICA* игнорирует все наблюдения, имеющие пропуски для какой-либо переменной в списке. Если опция не выбрана, то производится попарное удаление пропусков.

Отображать длинные имена переменных. Если опция выбрана, в таблице результатов показываются длинные имена переменных вместе с короткими.

***t*-критерий с разделенными оценками дисперсий.** Для вычисления *t*-критерия для независимых выборок, *STATISTICA* должна оценить дисперсию разности для соответствующих пар переменных (групп). По умолчанию, эта дисперсия вычисляется усреднением внутригрупповых дисперсий. Если дисперсии в двух группах сильно отличны, и число наблюдений также существенно отличается в каждой группе, то *t-критерий*, вычисленный таким образом, может неточно отразить статистическую значимость различий между группами. В этом случае следует использовать *t-критерий* с раздельными оценками дисперсий и приближенным числом степеней свободы (см. Blalock, 1972).

Ммногомерный критерий (Хотеллинга T^2). Хотеллинга T^2-критерий – это многомерный критерий для различий между средними в двух группах. Критерий является обобщением t-критерия Стьюдента на многомерные наблюдения. [Для одномерных наблюдений статистика критерия Хотеллинга совпадает со статистикой критерия Стьюдента.] Этот тест будет вычислен лишь при условии, что выбрана одна или более независимых переменных. Так как критерий основан на внутригрупповых матрицах дисперсий/ковариаций для зависимых переменных, пропущенные данные будут удалены автоматически построчным методом из вычислений.

Критерий Левена (однородности дисперсий). Обычный *t-критерий* для независимых выборок основан на предположении равенства (однородности) дисперсии в двух группах. Мощным статистическим критерием проверки этого предположения является критерий Левена. Если критерий Левена статистически значим, гипотеза об однородности дисперсии отвергается. Однако следует помнить, что *t-критерий* для независимых выборок является робастным (устойчивым) при числе наблюдений в каждой группе больше 30; таким образом, значимое значение критерия Левена не следует рассматривать как решительный довод против использования *t*-критерия.

Диаграммы размаха

Опция строит каскад диаграмм размаха для зависимых переменных, один график для каждой переменной (более подробно см. стр. 1525).

Категоризованные гистограммы

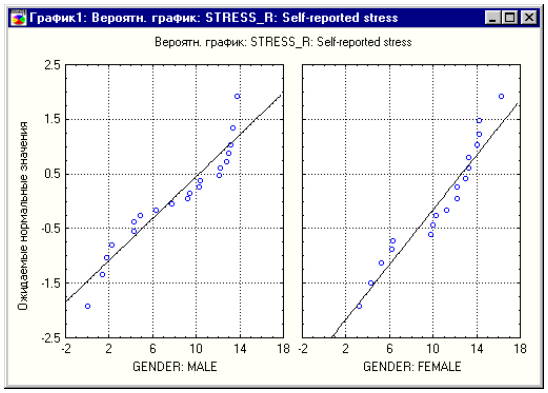
Опция (доступная также в окне *Описательные статистики*, см. стр. 1522) строит каскад категоризованных гистограмм (одна для каждой зависимой переменной) для распределения соответствующей переменной в двух группах.

StatSoft®

Copyright © StatSoft, 1995

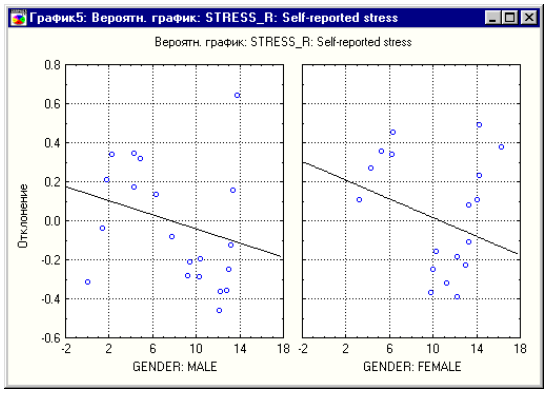
Категоризованные нормальные вероятностные графики

Эта опция строит каскад нормальных вероятностных графиков для зависимых переменных, разделенных по категориям.



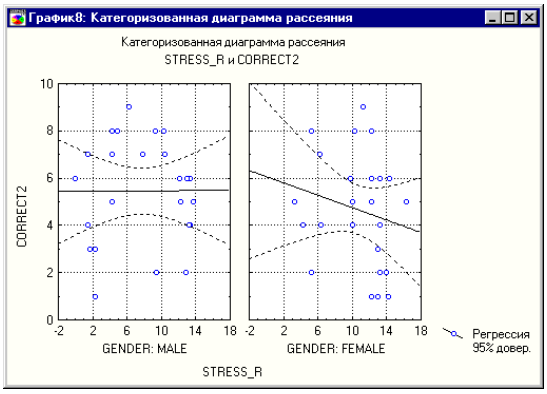
Категоризованные нормальные вероятностные графики без тренда

Эта опция строит каскад нормальных вероятностных графиков без тренда для зависимых переменных, разделенных на группы.



Категоризованные диаграммы рассеяния

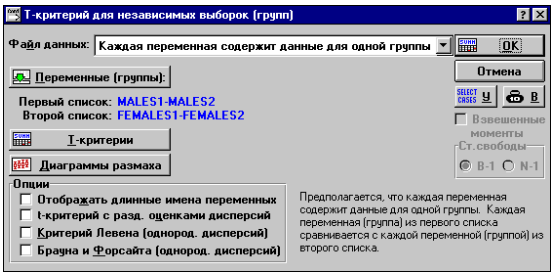
Эта опция (также доступная из окна *Описательные статистики*, см. стр. 1522) строит каскад категоризованных диаграмм рассеяния для выбранных переменных, по одной на каждую пару.



После нажатия этой кнопки программа запросит выбрать два набора переменных (из списка зависимых переменных). Диаграммы рассеяния будут построены для каждой пары переменных.

t-критерий для независимых выборок – сравнение столбцов (переменных)

Общий обзор *t*-критериев для независимых выборок приведен в разделе *Вводный обзор* (стр. 1458) Это диалоговое окно доступно когда сделана установка *Каждая переменная содержит данные для одной группы* для файла исходных данных (см. ниже) в диалоговом окне *T-критерий для независимых выборок* (стр. 1537).



Copyright © StatSoft, 1995

StatSoft

Как отмечалось ранее, этот порядок данных не типичен для приложений.

Файл данных

Имеется два варианта файла с исходными данными: *Используется группирующая переменная* и *Каждая переменная содержит данные для одной группы*. По умолчанию, произведена установка, в которой, каждая строка в файле данных представляет одно наблюдение (например, индивидуума, респондента), а каждый столбец – переменную или измерение (например, ответ на позицию опросника). В этой установке группирующая переменная (например, *Пол - Gender*) должна быть включена в анализ, чтобы определить к какой группе относится каждое наблюдение.

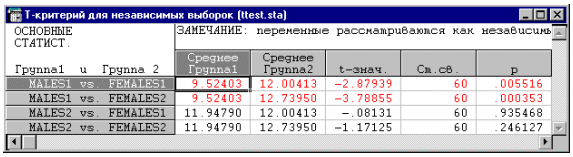
Принятая по умолчанию установка в окне предполагает, что данные упорядочены именно таким образом. Однако для простого анализа данных, иногда бывает легче ввести данные для каждой группы в отдельные колонки файла (например, данные для мужчин - в первую колонку, данные для женщин – во вторую). Если данные организованы таким образом, сделайте установку *Каждая переменная содержит данные для одной группы* (см. стр. 1539).

Переменные (группы)

Опция открывает окно выбора двух переменных. Каждая переменная (группа) выбранная в первом списке, сравнивается с переменной (группой) из второго списка.

T-критерии

Нажмите на кнопку *t*- критерии для независимых выборок, программа произведет вычисления и представит результаты в таблице.

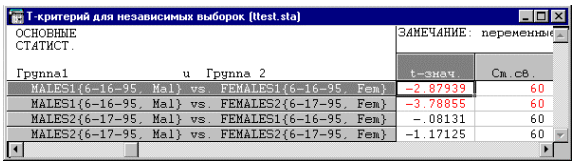


Представление результатов зависит от выбора в разделе *Опции* диалогового окна (см. ниже).

Опции

Выбор опций определяет детали и формат результатов *t*-критерия для независимых выборок.

Отображать длинные имена переменных. Если опция выбрана, в таблице результатов показываются длинные имена переменных вместе с короткими.



***t*-критерий с разделенными оценками дисперсий.** При вычислении *t*-критерия для независимых выборок *STATISTICA* оценивает дисперсию разности для соответствующих пар переменных (групп). По умолчанию, эта дисперсия вычисляется усреднением внутригрупповых дисперсий. Если дисперсии в двух группах сильно отличаются, и число наблюдений также существенно отлично, то *t-критерий*, вычисленный таким образом, может неточно отразить статистическую значимость различий между группами. В этом случае следует использовать *t*-критерий с раздельными оценками дисперсий и приближенным числом степеней свободы (см. Blalock, 1972).

Критерий Левена (однородности дисперсий). Обычный *t-критерий* для независимых выборок основан на предположении равенства (однородности) дисперсии в двух группах. Мощным статистическим критерием проверки этого

StatSoft®

Copyright © StatSoft, 1995

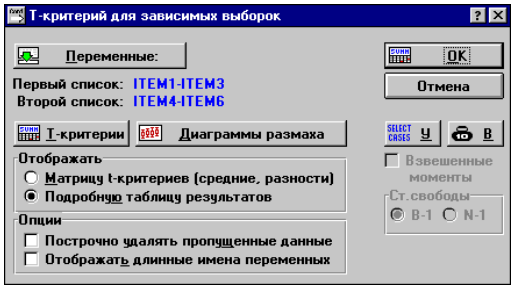
предположения является критерий Левена. Если критерий Левена статистически значим, то гипотеза об однородности дисперсии должна быть отвергнута. Однако нужно помнить, что *t-критерий* для независимых выборок является робастным (устойчивым) при числе наблюдений в каждой группе больше 30; таким образом, значимое значение критерия Левена не обязательно следует рассматривать как довод против применения *t*-критерия.

Диаграммы размаха

Опция строит каскад диаграмм размаха для зависимых переменных; один график для каждой переменной (более подробно см. стр. 1525).

t-критерий для зависимых выборок

Общий обзор *t*-критериев для зависимых выборок приведен в разделе *Вводный обзор* (стр. 1458).

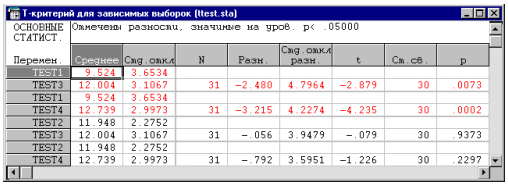


Переменные

Опция открывает окно выбора двух списков переменных. Каждая переменная из первого списка сравнивается с каждой переменной из второго списка.

T-критерии

После нажатия кнопки, *t*-критерии для зависимых выборок будут вычислены и показаны в таблице результатов.



Диаграммы размаха

Опция строит каскад диаграмм размаха для зависимых переменных; один график для каждой переменной (более подробно см. стр. 1525).

Опции

Отображать матрицу *t*-критериев. Если опция установлена, то, после нажатия кнопки *T*-критерии, будет построен каскад таблиц результатов, содержащих разности средних для каждой пары переменных (из двух списков), *t*-значения для соответствующих разностей и уровни значимости (*p-уровень*) для соответствующих *t*-значений.

Отображать подробную таблицу результатов. Дополнительно для каждого *t*-критерия вычисляются стандартные отклонения, число наблюдений *n* и т.д.

Построчно удалять пропущенные данные. Если выбрано построчное удаление пропущенных данных, то *STATISTICA* игнорирует все наблюдения, имеющие пропуски, для какой-либо переменной в списке. Если опция не выбрана, то производится попарное удаление пропусков.

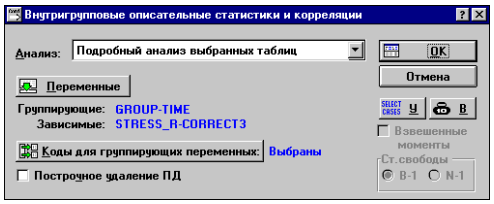
Отображать длинные имена переменных. Если опция выбрана, в таблице показываются длинные имена переменных.

Внутригрупповые описательные статистики и корреляции (группировка)

Общий обзор процедуры классификации дан в разделе *Вводный обзор* (стр. 1458).

Copyright © StatSoft, 1995

StatSoft®



Анализ

Имеются две установки, обсуждаемые ниже.

Подробный анализ выбранных таблиц. В этом случае пользователь может определить одно множество группирующих переменных и один лист зависимых переменных.

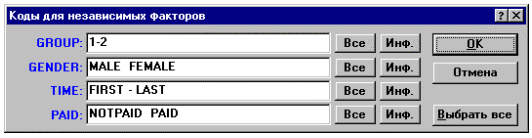
Пакетная обработка (и печать) списков таблиц. В данном случае можно задать до 6 списков группирующих переменных и список зависимых переменных (см. стр. 1548). *STATISTICA* вычислит таблицы для всех комбинаций переменных из первого списка группирующих переменных. Опция позволяет вычислить большое число таблиц в *пакетном режиме*.

Переменные

Опция открывает окно выбора двух списков переменных, в которых можно выбрать до 6 группирующих переменных и список зависимых переменных.

Коды для группирующих переменных

Открывает стандартное диалоговое окно выбора кодов для группирующих переменных (для того, чтобы определить к какой группе принадлежит каждое наблюдение).

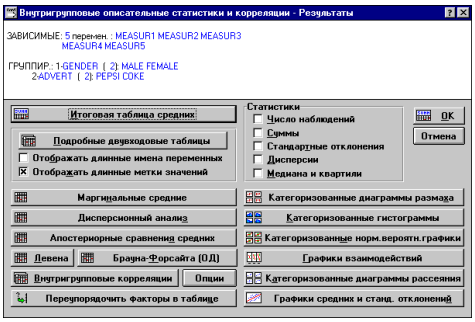


Если нажать *OK* без выбора какого-либо кода, *STATISTICA* автоматически определит все целые коды для группирующих переменных в файле данных.

Построчное удаление пропущенных данных. Если выбрано построчное удаление пропущенных данных, то *STATISTICA* игнорирует все наблюдения, имеющие пропуски для какой-либо переменной в списке. Если опция не выбрана, то пропуски удаляются попарно.

Внутригрупповые описательные статистики и корреляции - Результаты

Диалоговое окно *Внутригрупповые описательные статистики и корреляции - Результаты* позволяет вычислить различные внутригрупповые описательные статистики и построить разнообразные графики.



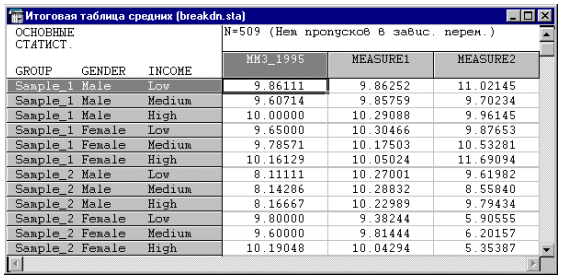
Опции *Быстрые основные статистики*. Вы можете вычислить различные описательные статистики (включая корреляции, медианы, квартили и т.д.) для сгруппированных данных в каждом модуле *STATISTICA* с помощью процедур *Быстрые основные статистики* (см. глава 10).

StatSoft®

Copyright © StatSoft, 1995

Итоговая таблица средних

После нажатия этой кнопки, вы увидите описательные статистики для всех групп.



Статистики определяются выбором, сделанным в поле *Статистики* текущего окна (см. ниже).

Статистики

Определяются статистики, вычисляемые в таблице результатов, когда пользователь выберет процедуру *Итоговая таблица средних*, или *Подробные двувходовые таблицы*, или *Маргинальные средние* (см. ниже).

Число наблюдений. Вычисляет число наблюдений в каждой группе.

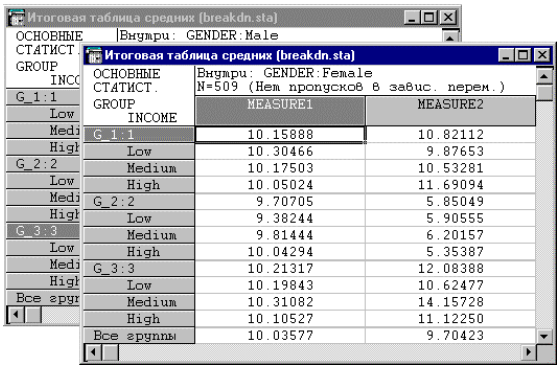
Суммы. Вычисляются суммы зависимых переменных в каждой группе.

Стандартные отклонения. Вычисляются стандартные отклонения зависимых переменных внутри каждой группы. Заметим, что стандартное отклонение, вычисляемое как корень квадратный из суммы квадратов отклонений (от среднего), деленной на *n-1*; - является оценкой *сигма* популяции.

Дисперсия. Вычисляются оценки дисперсии или выборочные дисперсии зависимых переменных для каждой группы. Выборочная дисперсия равна сумме квадратов отклонений (от среднего значения), деленной на *n-1*.

Подробные двувходовые таблицы

После того, как вы нажмете эту кнопку, каскад двувходовых таблиц появится на экране.



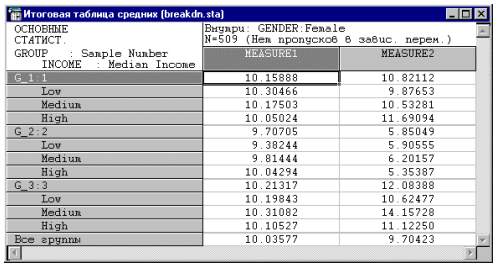
Если выбрано 2 или менее группирующих переменных, то будет построена только одна таблица. Если выбрано более двух группирующих переменных, таблицы будут построены для последних двух группирующих переменных, на соответствующих уровнях оставшихся группирующих переменных. Воспользуйтесь опцией *Переупорядочить факторы в таблице* (см. ниже), чтобы изменить порядок группирующих переменных и порядок построения таблиц. Статистики, включенные в таблицы вывода, определяются установками в поле *Статистики* (см. выше). Дополнительные опции позволяют показать на экране длинные имена переменных и длинные значения меток (см. ниже). В двувходовых таблицах вычисляются не только итоговые статистики внутри каждой группы, но также приводятся маргинальные итоговые статистики, т.е. статистики для каждого фактора в отдельности (другой фактор игнорируется). Определенные пользователем маргинальные статистики доступны с помощью опции *Маргинальные средние* (см. ниже).

Отображать длинные имена переменных. Если опция выбрана, в таблице результатов

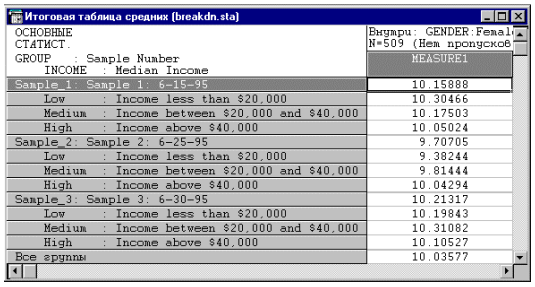
Copyright © StatSoft, 1995

StatSoft®

показываются длинные имена переменных вместе с короткими.

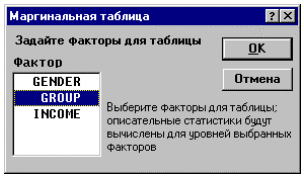


Отображать длинные значения меток. Показывает длинные значения меток для уровней группирующих переменных в первой колонке таблицы данных.

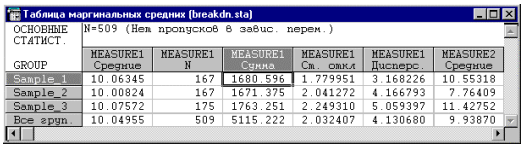


Маргинальные средние

Опция позволяет вычислить итоговые статистики для выбранных маргинальных таблиц, т.е. только для определенного набора группирующих переменных, игнорируя все остальные. Например, пусть имеются 3 группирующих переменных *Пол, Возраст, Национальность - Gender, Age Group, Ethnic Origin.* Если выбраны только *Пол - Gender* и *Национальность - Ethnic Origin*, то будет вычислена двувходовая итоговая таблица *Пол - Национальность*, без *Возраста - Age Group*.

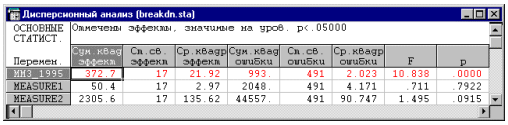


Нажмите *OK* и увидите таблицу маргинальных средних:



Дисперсионный анализ

Опция открывает таблицу результатов однофакторного дисперсионного анализа каждой зависимой переменной.



Если результат ANOVA статистически значим, можно далее воспользоваться процедурой *Апостериорные сравнения средних* (см. ниже) чтобы идентифицировать значимые разности между индивидуальными группами (средними). Используйте модуль *Дисперсионный анализ* чтобы вычислить полные таблицы дисперсионного анализа. Заметим, что однофакторная ANOVA, доступная в данной диалоговом окне, особенно удобна для быстрого однофакторного анализа многих групп.

Апостериорные сравнения средних

Опция позволяет вычислить обычные *апостериорные критерии* для попарных множественных сравнений средних (НЗР, Шеффе, Ньюмена-Кеулса, Дункана, Тьюки, Spjotvoll & Stoline). После нажатия кнопки, выберите зависимые переменные для тестирования (см. *Апостериорное сравнение средних*, стр. 1551).

Критерий Левена (однородности дисперсий). Обычный *t-критерий* для

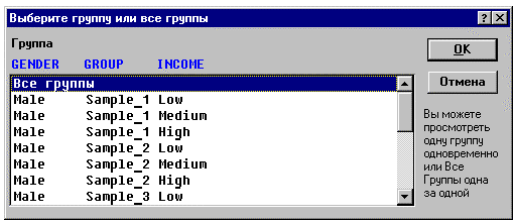
StatSoft®
Copyright © StatSoft, 1995

зависимых выборок основан на предположении равенства (однородности) дисперсии данных в двух группах. Мощным статистическим критерием проверки этого предположения является критерий Левена. Если критерий Левена статистически значим, то гипотеза об однородности дисперсии должна быть отвергнута. Однако нужно помнить, что *F* статистика (в ANOVA) является робастным критерием для различий в средних, если (1) число наблюдений *n* в группе больше 10, и (2) средние между группами не коррелируют со стандартными отклонениями. Таким образом, значимое значение критерия Левена не обязательно следует рассматривать как довод против применения *t*-критерия.

Предположение о некоррелированности средних и стандартных отклонений может быть легко проверено с помощью *Графиков средних и стандартных отклонений* из этого диалогового окна (см. ниже).

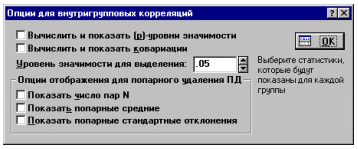
Внутригрупповые корреляции

Если выбрано более одной зависимой переменной, то с помощью опции можно вычислить внутригрупповые корреляционные матрицы этих переменных (см. стр. 1548). Вначале выбираются группы, для которых вычисляется корреляционная матрица.



Если выбрано *Все группы*, будет построен набор таблиц (показывающих корреляционные матрицы для всех групп).

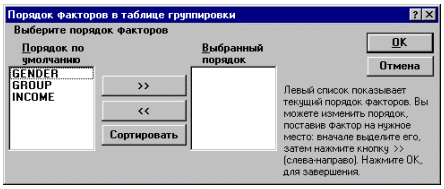
Опции. Кнопка открывает диалоговое окно с выбором параметров представления *внутригрупповых корреляций* (см. выше).



В частности, в окне можно определить *альфа* (*p*) уровень для выделения значимых корреляций, задать вычисление внутригрупповых ковариационных матриц, вычислить дополнительные статистики (например, число исключенных пар *n*, средние, стандартные отклонения; см. *Внутригрупповые корреляции*, стр. 1548, где приведено описание этих статистик).

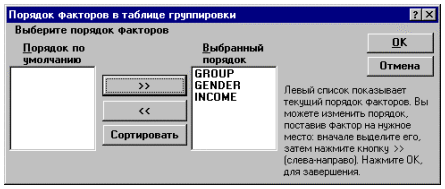
Переупорядочить факторы в таблице

Опция позволяет изменить порядок группирующих переменных. Заметим, при вычислении внутригрупповых статистик, *STATISTICA* будет проводить вычисления внутри групп, построенных заданным способом, при этом значения фактора, указанного последним будут обработаны в первую очередь, фактор, указанный предпоследним, будет обрабатываться во вторую очередь и т.д.

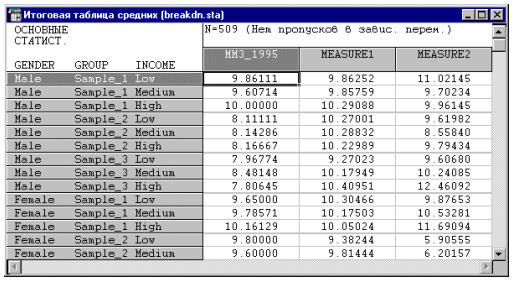


Чтобы изменить порядок группирующих переменных, переместите их на нужное место из левой колонки (*Порядок по умолчанию*) в правую колонку (*Выбранный порядок*) и нажмите *OK*.

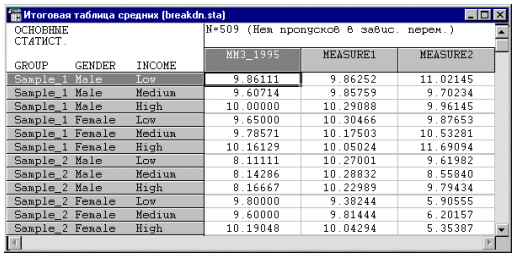
Copyright © StatSoft, 1995



До переупорядочения имели:



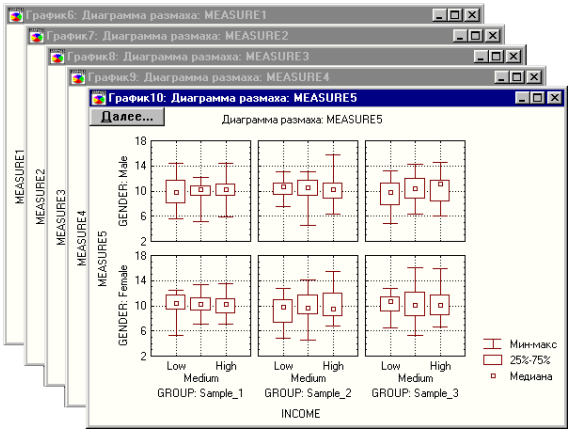
После изменения порядка имеем:



Вы можете всегда вернуться к принятому по умолчанию порядку, нажав кнопку *Сортировать* в диалоговом окне *Порядок факторов*.

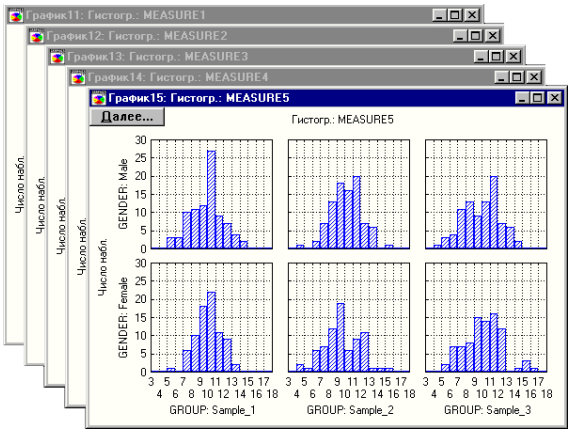
Категоризованные диаграммы размаха

Опция строит каскад категоризованных диаграмм размаха для зависимых переменных, один график для одной переменной. Для каждой переменной *STATISTICA* построит диаграмму размаха, разбитую на классы с помощью группирующих переменных (например, отдельно для мужчин и женщин в каждой национальной группе).



Категоризованные гистограммы

Опция строит каскад категоризованных гистограмм.



Для каждой выбранной переменной категоризованные гистограммы будут построены для распределений выбранных переменных внутри групп.

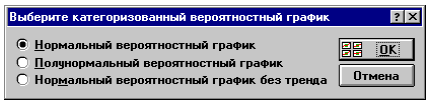
Категоризованные нормальные вероятностные графики

Опция строит каскад нормальных вероятностных графиков для каждой выбранной зависимой переменной, разбитой на классы (группы)

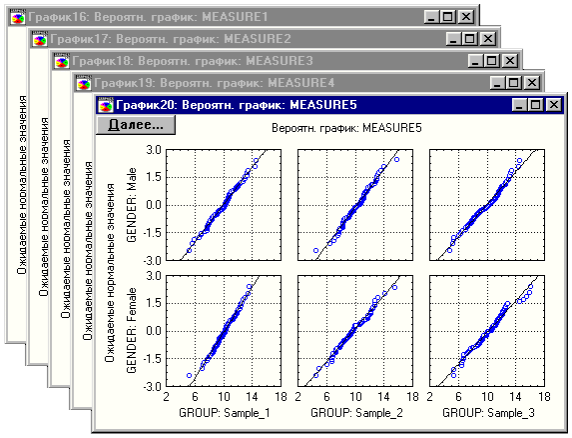
StatSoft®

Copyright © StatSoft, 1995

группирующими переменными. После нажатия кнопки появится диалоговое окно, где можно выбрать стандартные нормальные графики, полунормальные графики, нормальные графики без тренда (см. стр. 1522).

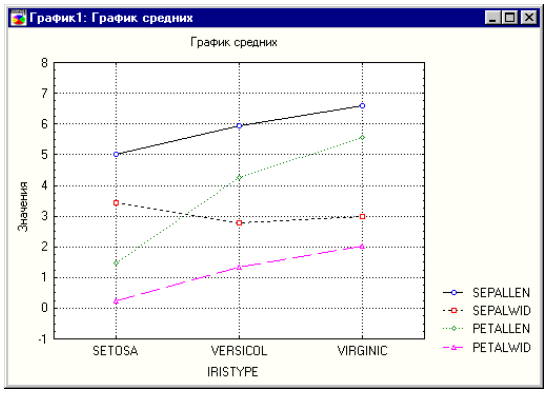


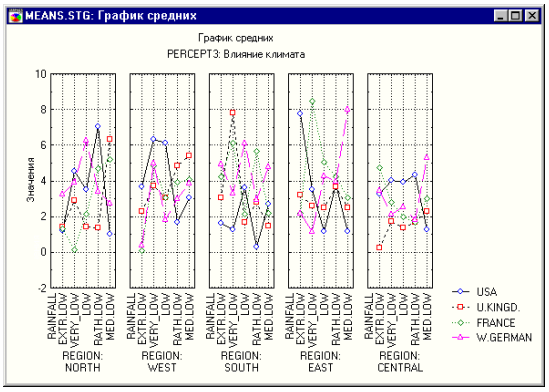
После нажатия *OK* в этом диалоговом окне система попросит выбрать зависимые переменные для графика.



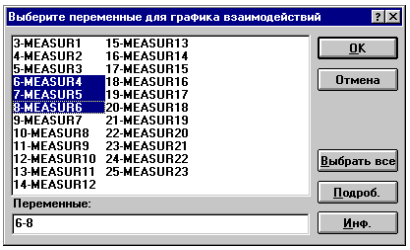
Графики взаимодействий

Опция строит график взаимодействий средних внутри групп.





В текущем анализе выбрано более одной переменной. В этом случае после нажатия кнопки *Графики взаимодействий* появится диалоговое окно *Выберите переменные для графика взаимодействий.*



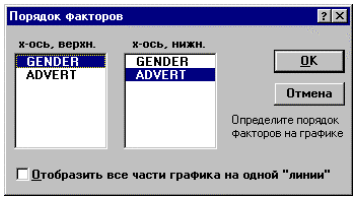
После выбора переменных нажмите *OK,* откроется диалоговое окно *Порядок факторов* (см. ниже).

В текущем анализе выбрана одна зависимая переменная. Так как вы уже выбрали зависимую переменную, то после нажатия кнопки *Графики взаимодействий* откроется диалоговое окно *Порядок факторов* (см. ниже). На графике взаимодействий группирующие переменные будут представлены различными типами линий и различным цветом.

Порядок факторов. Окно *Порядок факторов* имеет вид:

Copyright © StatSoft, 1995

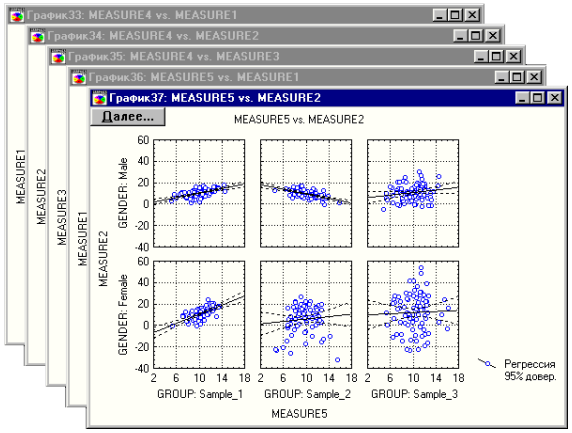
StatSoft®



В окне можно определить, каким деталям графика (нижней метке на оси *X*, верхней метке на оси *X*) соответствуют выбранные переменные.

Категоризованные диаграммы рассеяния

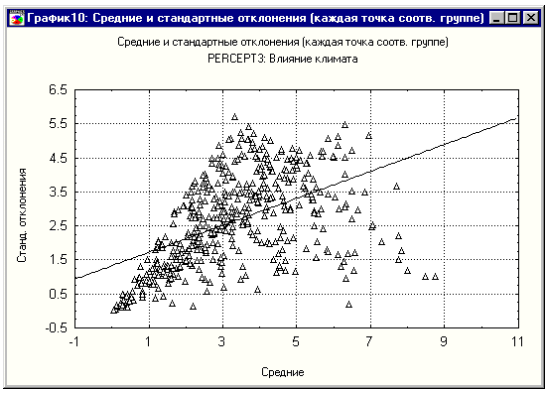
Опция строит каскад категоризованных диаграмм рассеяния.



Категоризованные диаграммы рассеяния будут построены для каждой выбранной переменной.

Графики средних и стандартных отклонений

Опция позволяет построить график средних и стандартных отклонений для выбранных переменных на пересечении групп.

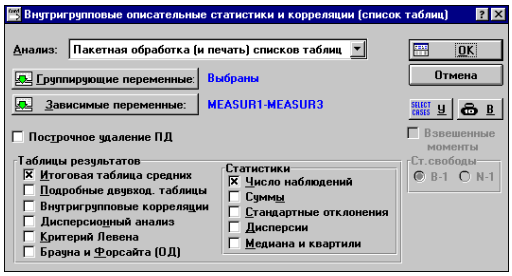


Этот график полезен для распознавания потенциальных выбросов среди средних, что может послужить источником ложного заключения о статистической значимости различий между средними. Одно из общих и наиболее серьезных нарушений предположений ANOVA состоит в том, что средние коррелированы со стандартными отклонениями на пересечении групп. Например, пусть имеется 5 групп из 10 наблюдений каждая. Если в одной группе 2 наблюдения являются резко выделяющимися (выбросы), то дисперсия в этой группе будет много больше и среднее будет сильно отличаться от общего среднего.

Внутригрупповые описательные статистики и корреляции (список таблиц)

Общая идея процедуры *Группировка* дана в *Вводном обзоре* (стр. 1462). Диалоговое окно *Внутригрупповые описательные статистики и корреляции (список таблиц)* открывается, если вы выбрали опцию *Пакетная обработка (и печать) списков таблиц*.

StatSoft®
Copyright © StatSoft, 1995



Анализ

Имеются две установки, обсуждаемые ниже.

Подробный анализ выбранных таблиц. В этом случае пользователь может определить один список группирующих переменных и один список зависимых переменных.

Пакетная обработка (и печать) списков таблиц. В данном случае можно задать до 6 списков группирующих переменных и список зависимых переменных (см. стр. 1548). *STATISTICA* вычислит таблицы для всех комбинаций переменных из первого списка группирующих переменных. Опция позволяет вычислить большое число таблиц в *пакетном режиме*.

Группирующие переменные

Открывается стандартное диалоговое окно, в котором можно выбрать до 6 списков группирующих переменных и список зависимых переменных (см. стр. 1548). *STATISTICA* вычислит таблицы для всех комбинаций переменных из первого списка группирующих переменных.

Зависимые переменные

Опция открывает стандартное окно для выбора зависимых переменных в таблицах группировки.

Построчное удаление пропущенных данных. Если выбрано построчное удаление пропущенных данных, то *STATISTICA* игнорирует все наблюдения, имеющие пропуски, для какой-либо переменной в списке. Если опция не выбрана, то производится попарное удаление пропусков.

Таблицы результатов

Если сделаны установки в данных опциях, то каскады таблиц с различными статистиками и критериями будут вычислены и показаны на экране (см. ниже).

Итоговая таблица средних

После нажатия этой кнопки, увидите описательные статистики для всех групп. Статистики, вычисленные в таблице результатов, определяются выбором, сделанным в поле *Статистики* (см. ниже).

Подробные двувходовые таблицы

Если выбрано более двух списков группирующих переменных, таблицы будут построены для последних двух группирующих переменных, на соответствующих уровнях оставшихся группирующих переменных. Воспользуйтесь опцией *Переупорядочить факторы в таблице* (см. ниже), чтобы изменить порядок группирующих переменных и порядок построения таблиц. Статистики, включенные в таблицы результатов, определяются установками в поле *Статистики*, (см. ниже).

Внутригрупповые корреляции

Опция позволяет вычислить внутригрупповые корреляции. Если выбрано более одной зависимой переменной, то с помощью опции можно вычислить внутригрупповые корреляционные матрицы этих переменных (см. стр. 1548).

Copyright © StatSoft, 1995

Дисперсионный анализ

Опция открывает таблицу результатов однофакторного дисперсионного анализа зависимых переменных.

Критерий Левена (однородности дисперсий). Обычный *t-критерий* основан на предположении равенства (однородности) дисперсии в двух группах. Мощным статистическим критерием проверки этого предположения является критерий Левена. Если критерий Левена статистически значим, то гипотеза об однородности дисперсии отвергается. Однако следует помнить, что *F* статистика является робастным (устойчивым) критерием, если выполнены следующие условия: (1) число наблюдений *n* в группе больше 10, и (2) средние между группами не коррелированы со стандартными отклонениями. Таким образом, значимое значение критерия Левена не обязательно следует рассматривать как довод против применения *t*-критерия.

Дополнительные статистики

Определяются статистики, вычисляемые в таблице результатов, когда пользователь выберет *Итоговые таблицы средних* или *Подробные двувходовые таблицы*.

Число наблюдений. Вычисляется число наблюдений в каждой группе.

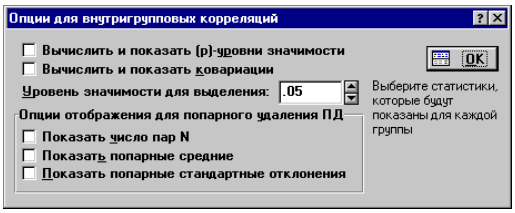
Суммы. Вычисляются суммы зависимых переменных в каждой группе.

Стандартные отклонения. Вычисляются стандартные отклонения зависимых переменных внутри каждой группы. Заметим, что стандартное отклонение, вычисляемое как корень квадратный из суммы квадратов отклонений (от среднего), деленной на *n-1*; является оценкой *сигма* популяции.

Дисперсия. Вычисляются оценки дисперсии или выборочные дисперсии зависимых переменных для каждой группы. Выборочная дисперсия равна сумме квадратов отклонений (от среднего значения), деленной на *n-1*.

Опции вычисления внутригрупповых корреляций

В этом диалоговом окне выбираются опции, позволяющие показать сопровождающие статистики для коэффициентов корреляции.



Вычислить и показать (*p*)-уровни значимости

Для каждого коэффициента корреляции вычисляется уровень значимости *p*.

Вычислить и показать ковариации

Вычисляется матрица ковариаций.

Опции отображения для попарного удаления пропущенных данных (ПД)

Опции имеют смысл при попарном способе удаления пропущенных данных.

Показать число пар N

Для каждой пары переменных показывается число наблюдений, по которым вычислен соответствующий коэффициент корреляции.

StatSoft®

Copyright © StatSoft, 1995

Показать попарные средние

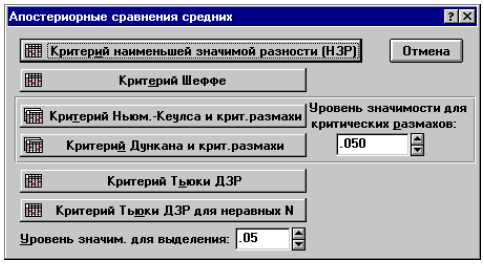
Для каждой пары переменных показываются средние, которые используются при вычислении соответствующего коэффициента корреляции.

Показать попарные стандартные отклонения

Для каждой пары переменных показываются стандартные отклонения, которые используются при вычислении соответствующего коэффициента корреляции.

Апостериорные сравнения средних

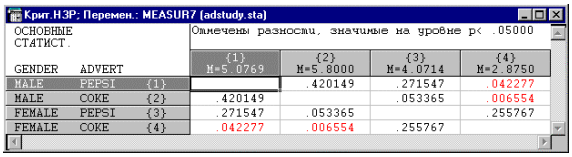
В этом диалоговом окне можно выбрать различные процедуры *апостериорного* сравнения средних.



Обсуждение *апостериорных* процедур множественного сравнения можно найти в книгах Wienar (1962), Hays (1988), Milliken, Jonson (1984). Идея этих процедур кратко состоит в следующем. Обычно после получения статистически значимого результата в дисперсионном анализе (ANOVA) [иными словами, если с помощью F-критерия отвергнута гипотеза о равенстве групповых средних], интересно знать, какие средние отличаются друг от друга, а какие нет (однородны или неоднородны рассматриваемые группы). Формально можно было бы выполнить серию обычных *t*-критериев и сравнить всевозможные пары средних. Однако такая процедура чисто случайно *увеличивает шансы* получить значимый результат (получить значимый результат там, где его нет). Представьте, вы имеете 20 выборок по 10 случайных чисел и вычислили 20 средних в каждой из них. Далее возьмите наибольшее среднее и сравните с наименьшим. Очевидно, вероятность, что эти средние различны, существенно выше, чем если бы сравнивались два произвольно выбранных средних. *t-критерий для независимых выборок* проверяет, является различие между средними значимым или нет в предположении, что сравниваются *только две выборки*. Проведение многократных (множественных) сравнений приводит к получению чисто случайных (искусственных) результатов. *Апостериорные* процедуры служат защитой от такого рода результатов. В этих процедурах учитывается, что производятся множественные (многократные) попарные сравнения групп.

Критерий наименьшей значимой разности (НЗР)

Этот тест эквивалентен *t*-критерию для независимых выборок, построенному по *k* группам, включенным в сравнение. Он обеспечивает наименьшую защиту выводов от увеличения ошибки 1-го рода *альфа*, возникающего из-за многих сравнений. Формально процедура состоит в применении *t*-критерия Стьюдента к каждой из *k(k-1)/2* паре групп. [В общем случае об этом критерии можно сказать лишь, что если истинные средние в группах равны, то примерно *альфа k(k-1)/2* пар средних будут ошибочно признаны неравными.] После выбора опции будет построена матрица *апостериорных* уровней значимости.



Copyright © StatSoft, 1995

StatSoft®

Критерий Шеффе

Опция строит *апостериорные p*-уровни для критерия Шеффе. Этот критерий более консервативный, чем критерий Ньюмена-Кеулса или Дункана (см. Wienar, 1962).

Критерий Ньюмена-Кеулса и критические размахи

Данный тест основан на статистике, построенной по *стьюдентизированным* размахам. [Стьюдентизированный размах – это размах выборки, деленный на корень квадратный из выборочной дисперсии.] Вначале групповые средние располагаются в порядке возрастания, иными словами, строится их вариационный ряд. Далее для каждой пары средних *при данном числе выборок* вычисляет вероятность получения разностей данной или большей величины при нулевой гипотезе (средние не различаются).

Вторая таблица результатов дает *критические размахи* разностей средних для соответствующего уровня ошибки 1-го рода *альфа* (по умолчанию принятому *.05*).

Критерий Дункана

Критерий основан на той же логике, что и критерий Ньюмена-Кеулса, однако является менее консервативным (см., например, Milliken and Johnson, 1984).

Критерий Тьюки достоверно значимой разности (ДЗР)

Критерий Тьюки ДЗР (см. Winer, 1971) находится между тестом Ньюмена-Кеулса и Шеффе с точки зрения консервативности. [Критерий Тьюки работает следующим образом: выборочные групповые средние упорядочиваются. Вначале рассматривается разность максимального и минимального значения (размах). Если принимается гипотеза о равенстве этого размаха нулю, то процедура множественных сравнений завершается и все группы признаются однородными. Если нет, то рассматривается разность между максимальным и следующим за минимальным групповым средним и вновь проверяется равенство этой разности нулю. В итоге или находятся групповые средние, равные максимальному, или оказывается, что таких средних нет. Затем максимальное выборочное среднее удаляется из рассмотрения и заменяется вторым по величине групповым средним и т.д. В этой процедуре также получаются доверительные интервалы для всех разностей между истинными групповыми средними. Заметим, что описанную процедуру сравнения можно распространить на линейные комбинации средних с весами, дающими в сумме нуль (так называемые контрасты), что и делается в дисперсионном анализе].

Критерий Тьюки достоверно значимой разности для выборок неравного размера (неравных N)

Этот критерий обобщает обычный критерий ДЗР Тьюки на выборки неравных размеров (см. Spjotvoll and Stoline, 1973, стр. 975).

Уровень ошибки I рода *альфа* для выделения

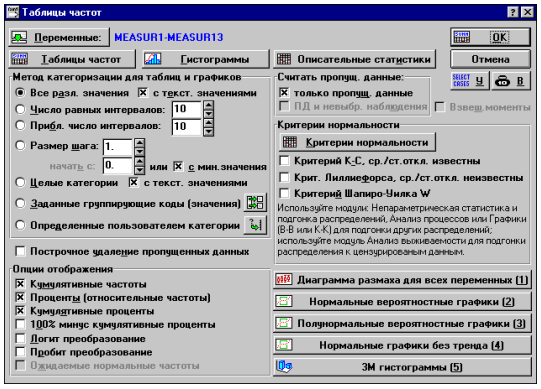
Попарные множественные сравнения проверяют разности пар средних значений и выдают матрицу результатов. Во всех таблицах средние, значимо различающиеся на данном уровне значимости *альфа* выделяются цветом. Введите нужный уровень *альфа* для выделения значимых результатов. Позднее вы можете изменить этот уровень с помощью кнопки *Опции* (см. стр. 1568).

StatSoft®

Copyright © StatSoft, 1995

Таблицы частот

Общая идея процедуры *Таблицы частот* обсуждалась в *Вводном обзоре* (стр. 1465).

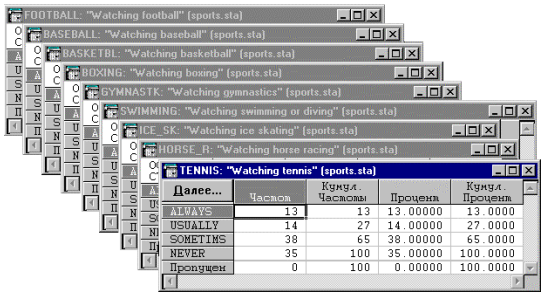


Переменные

Опция открывает окно выбора списка переменных для анализа.

Таблицы частот

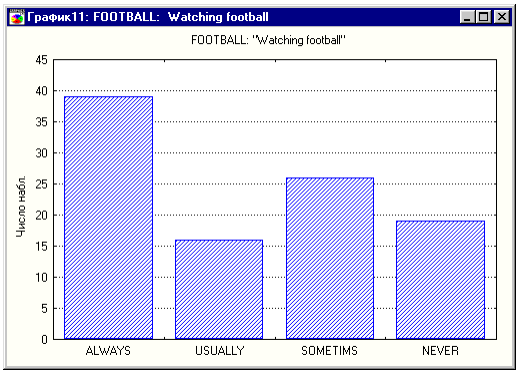
Опция строит последовательность таблиц частот для выбранных переменных.



Способ, которым переменные группируются в таблицах частот, определяется выбором одной из опций, объединенных заголовком *Метод категоризации для таблиц и графиков* (см. ниже). Если выбран какой-либо *Критерий нормальности,* то дополнительная таблица с результатами тестирования появится на экране (критерии, позволяющие проверить гипотезы согласия данных с распределениями, отличными от нормального, доступны в модулях *Непараметрическая статистика* и *Анализ выживаемости*).

Гистограммы

Опция строит последовательность гистограмм для выбранных переменных, одна гистограмма для одной переменной.



Способ, каким переменные группируются в гистограммах, аналогичен группировкам в таблицах частот (см. выше).

Описательные статистики

Опция строит таблицу результатов с описательными статистиками для выбранных переменных. Способ обработки данных при вычислении описательных статистик зависит от установки, сделанной в поле *Метод категоризации в таблицах и графиках* (см. ниже). Именно, доступны следующие опции:

(1) Если выбран *Размер шага* и определено минимальное значение, то все наблюдения со значениями меньше минимума, будут проигнорированы при вычислении;

(2) Если выбрана опция *Целые категории*, то все нецелые значения данных игнорируются;

Copyright © StatSoft, 1995

StatSoft®

(3) Если выбрана опция *Заданные группирующие коды (значения),* все значения переменных, которые не равны одному из выбранных целых кодов, будут игнорироваться;

(4) Если выбрана опция *Определенные пользователем категории,* то *Описательные статистики* недоступны (т.к. нужно определить сложные составные условия, определяющие принадлежность наблюдения к определенному классу (группе), см. ниже).

Метод категоризации для таблиц и графиков

Установки этих опций определяют: (1) как будут сгруппированы или табулированы выбранные переменные в таблицах частот и гистограммах, (2) как обрабатываются наблюдения при вычислении *Описательных статистик* (см. выше).

Все различные значения. Выберите эту опцию, если хотите, чтобы частоты строились с учетом всех различных значений анализируемых переменных.

С текстовыми значениями. Выберите эту опцию, если вы хотите, чтобы частоты строились с учетом всех различных текстовых значений выбранных переменных.

Число равных интервалов. Если эта опция выбрана, то диапазон значений каждой переменной делится на указанное число интервалов.

Приблизительное число интервалов; прибл. число (приближенное число). Если опция выбрана, то *STATISTICA* построит "приближенные" интервалы и выберет приближенный размер шага так, что последняя десятичная цифра будет 1, 2, или 5 (например, 10.5, 11.0, 11.5, и т.д.). Такие интервалы легче интерпретировать, чем интервалы, содержащие много десятичных знаков (например, 10.12423, 10.13533, и т.д.). Таким образом, строятся ясные и легко читаемые таблицы, в которых, однако, число интервалов может отличаться от заданного в опции *Число интервалов.*

Размер шага. Опция задает ширину интервала категоризации в таблицах частот (и гистограммах).

Начать с минимального значения. Если опция выбрана, то группировка начинается с минимального значения переменной (первый интервал группировки включает это значение). Если опция не выбрана, то левая граница первого интервала группировки задается пользователем в поле *начать с*.

Целые категории. Если опция выбрана, то границами интервалов категоризации в таблицах частот (и гистограммах) будут целые числа, а размер шага равен наименьшему целому значению. Все нецелые значения переменных будут проигнорированы программой.

С текстовыми значениями. Категории при выборе *Таблиц частот* и *Гистограмм* будут помечены текстовыми значениями (например, *мужчины - males*, *женщины - females*) а не целыми значениями (например, *1*, *2*), которые доступны в текущем файле данных.

Заданные группирующие коды (значения). Если опция выбрана, то таблицы частот (и гистограммы) будут построены с помощью целых кодов, определенных пользователем (с помощью расположенной рядом с опцией кнопки).

Определенные пользователем категории. Опция позволяет определить до 16 логических условий, позволяющих отнести наблюдения к определенной категории (до 16) в таблице частот. Логические условия могут быть сложными и включать различные переменные файла данных, а также наблюдения.

Для каждого наблюдения условия выбора наблюдений (отнесения к определенной категории) проверяются последовательно;

StatSoft®

Copyright © StatSoft, 1995

наблюдение относится к той категории, в какую оно попало первым (где соответствующее логическое условие приняло значение *истина - true*). Нажмите расположенную рядом кнопку, чтобы получить представление о правилах построения логических условий.

Построчное удаление пропущенных данных. Если выбрано построчное удаление пропущенных данных, то *STATISTICA* игнорирует все наблюдения, имеющие пропуски, для какой-либо переменной в списке. Если опция не выбрана, то производится попарное удаление пропусков.

Опции отображения

Кумулятивные частоты. Опция вычисляет кумулятивные или накопленные частоты. Это сумма частот по интервалам группировки.

Проценты (относительные частоты). Опция вычисляет относительные частоты (проценты).

Кумулятивные проценты. Опция вычисляет кумулятивные или накопленные проценты.

100% минус кумулятивные проценты. Опция вычисляет 100-минус-кумулятивные проценты.

Логит преобразование частот. Для частот каждой группы осуществляется преобразование логит: $Logit_i=natural\ log(p_i/(1-p_i))$, где p_i – кумулятивная частота группы *i*.

Пробит преобразование частот. Для кумулятивных частот каждой категории осуществляется преобразование пробит – вычисляются *z*-значения, связанные с вероятностью в соответствующей ячейке.

Ожидаемые нормальные частоты. Вычисляется нормальное приближение для наблюдаемых частот.

Пропущенные данные

Только пропущенные данные. Опция вычисляет группу пропущенных данных. Проценты и кумулятивные (накопленные) проценты также вычисляются для пропущенных данных.

ПД и невыбранные наблюдения. Опция вычисляет дополнительно группы невыбранных (и не пропущенных) наблюдений в таблице частот. Вычисляются проценты и кумулятивные проценты относительно общего числа наблюдений.

Критерии нормальности

Если выбраны соответствующие опции, то после таблиц частот на экране появятся таблицы с критериями нормальности для выбранных переменных (одна таблица для одного критерия). Чтобы показать только критерии, нажмите кнопку *Критерии нормальности*.

Критерии нормальности. Кнопка выведет таблицы результатов с запрошенными критериями нормальности для выбранных переменных.

Критерий Колмогорова-Смирнова, ср./ст.откл. известны. Опция вычисляет одновыборочный критерий Колмогорова-Смирнова. Если *D* статистика Колмогорова-Смирнова значима, то гипотеза нормальности распределения значений переменной отвергается. Два значения вероятности для каждого критерия вычисляются: первое табулировано Massey (1951). Это значение относится к случаю, когда параметры распределения – среднее и стандартное отклонение известны заранее и не оцениваются. Однако обычно эти параметры не известны и оцениваются по наблюдениям. В этом случае проверяется сложная гипотеза (с неизвестными параметрами распределения) и вычисляются вероятности *Лиллиефорса* (Lilliefors, 1967). В последние годы W критерий Шапиро-Уилка (см.

Copyright © StatSoft, 1995

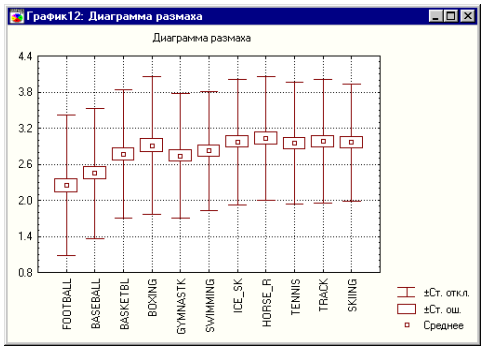
ниже) рассматривается как подходящий для проверки гипотезы нормальности (см. Shapiro, Wilk, and Chen, 1968).

Критерий Лиллиефорса, ср./ст.откл. неизвестны. Если эта опция выбрана, то программа вычислит одновыборочную *D* статистику Колмогорова-Смирнова, а также вероятности Лиллиефорса (см. предыдущий параграф; см. также Lilliefors, 1967).

W критерий Шапиро-Уилка. Опция вычисляет статистику W Шапиро-Уилка для каждой выбранной переменной. Если *W* статистика значима, то гипотеза о нормальном распределении значений переменной отвергается. В *STATISTICA* используется улучшенная версия алгоритма, предложенная Royston (1982), которая применима к выборкам, содержащим до 2000 наблюдений. Если число наблюдений больше 2000, этот критерий не применим.

Диаграмма размаха для всех переменных

Опция строит диаграммы размаха для выбранных переменных.



Нормальные вероятностные графики

Опция строит последовательность нормальных вероятностных графиков – один график для одной выбранной переменной.

Полунормальные вероятностные графики

Опция строит последовательность полунормальных вероятностных графиков – один график для каждой выбранной переменной.

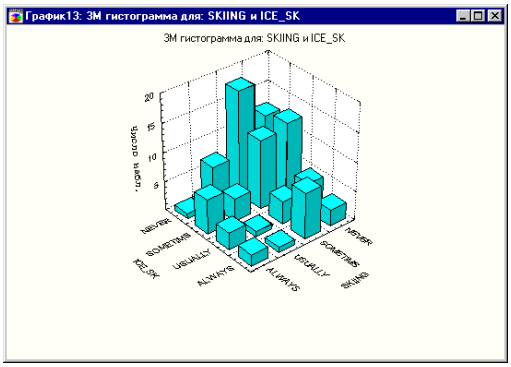
Единственное отличие этих графиков от нормальных вероятностных графиков состоит в том, что рассматривается только правая часть нормальной плотности. Иными словами, только положительные нормальные значения откладываются на оси *Y*. Этот график используется, когда хотят игнорировать знак наблюдений (т.е. рассматривают только их абсолютные значения).

Нормальные вероятностные графики без тренда

Опция строит последовательность нормальных вероятностных графиков без тренда – один график на одну выбранную переменную.

3М гистограммы

Опция строит каскад 3М гистограмм для пар выбранных переменных, один график на каждую пару.



После нажатия этой кнопки программа попросит пользователя выбрать два набора переменных (из списка выбранных ранее с помощью кнопки *Переменные*). 3М гистограммы будут построены

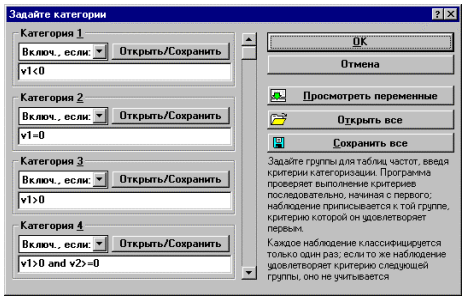
StatSoft®
Copyright © StatSoft, 1995

для каждой пары переменных, включающей переменные из разных списков.

Определенные пользователем категории

Это средство используется для того, чтобы определить сложные подмножества наблюдений с помощью логических условий. Варианты этого диалогового окна используются в различных модулях *STATISTICA* (см., например, *Управление данными*, глава 7; *Статистические графики*, том II).

В данном случае опция позволяет задать до 16 условий, которые определяют принадлежность наблюдения к определенной категории. Условия проверяются последовательно. Наблюдение относится к первой по порядку категории, в которую оно попало.



Пример

Следующие два условия выбора наблюдений приводят к образованию двух категорий (групп) в таблице частот.

Категория 1 Включить, если: **`v1>=0`**

Категория 2 Включить, если: **`v1<0`**

Согласно этим условиям, первая группа состоит из всех наблюдений, для которых значения первой переменной больше или равны 0, вторая группа состоит из наблюдений, для которых значения первой переменной меньше нуля.

Категория

Можно определить до 16 условий, которые позволяют отнести наблюдение к определенной категории. Заданные условия проверяются последовательно. Наблюдение относится к первой по порядку категории, в которую оно попало (т.е. для которой логическое условие стало истинным).

Включить/исключить если

Когда выбрано *Включить, если* и условие истинно, то наблюдение включается в соответствующую категорию. Если выбрано *Исключить, если* наблюдение включается в соответствующую категорию, если условие неистинно.

Синтаксис условий выбора наблюдений

В поле *Категория* указываются условия выбора наблюдений, соответствующие стандартным *Условиям выбора наблюдений программы STATISTICA* (подробности см. в главе 3). Далее приведены соглашения о синтаксисе условия выбора наблюдений. Те же соглашения используются во всех случаях, когда в программу *STATISTICA* вводятся пользовательские условия выбора наблюдений.

- Переменные вызываются по их номерам (например, *v1 = 1*) или именам (например, *Пол - Gender = 1*); обратите внимание, что имена переменных можно набирать в верхнем или нижнем регистре (т.е. *'ПОЛ' - 'GENDER'* эквивалентно *'пол' - 'gender'*).
- Текстовые значения переменной заключаются в апострофы (например, *v1 = 'МУЖЧИНА' - 'MALE'*); обратите внимание, что текстовые значения можно набирать в верхнем или нижнем регистре (т.е. *'Yes'* эквивалентно *'YES'*).

Copyright © StatSoft, 1995

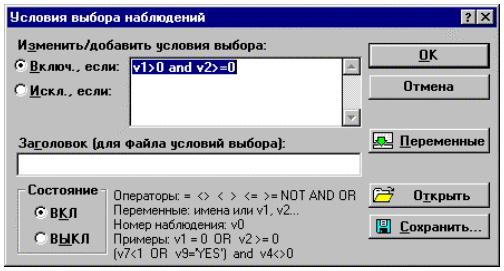
StatSoft®

- Для сложных логических условий или изменения порядка выполнения операторов используются скобки (см. пример выше).
- Логические операторы:

`=`	- равно
`#, <>, ><`	- не равно
`<`	- меньше
`>`	- больше
`<=`	- меньше либо равно
`>=`	- больше либо равно
`NOT, ~`	- логическое НЕ
`AND, &`	- логическое И
`OR, !`	- логическое ИЛИ

Открыть/Сохранить

Опция открывает стандартный файл *Условия выбора наблюдений* (который можно использовать, чтобы выбрать/фильтровать наблюдения, см. глава 3) а также сохранить условие для каждого класса (категории). Нажмите на кнопку *Открыть/Сохранить,* чтобы открыть диалоговое окно *Условия выбора наблюдений,* где можно редактировать, сохранить или загрузить условия выбора.



Условия выбора наблюдений будут сохранены в файле с расширением **.sel*.

Просмотреть переменные

Нажатие кнопки открывает диалоговое окно *Переменные*, в котором можно увидеть переменные в таблице.

Открыть все

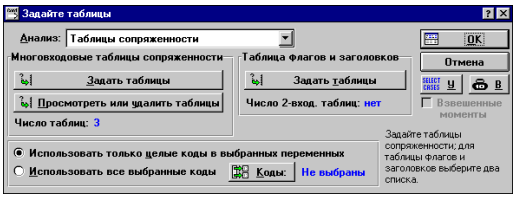
В отличие от опции *Открыть,* эта опция открывает диалоговое окно, в котором можете выбрать файл со всеми условиями выбора наблюдений.

Сохранить все

В отличие от опции *Сохранить*, эта кнопка открывает диалоговое окно *Сохранение условий выбора наблюдений*, в котором можно указать файл для сохранения условий выбора. Файл будет сохранен в текстовом формате (расширение **.txt*).

Таблицы сопряженности и таблицы флагов и заголовков

Когда выбрана процедура *Таблицы и заголовки* из стартовой панели, открывается диалоговое окно *Задайте таблицы*.



Анализ

Таблицы сопряженности. При этом значении поля *STATISTICA* будет

StatSoft®
Copyright © StatSoft, 1995

кросстабулировать стандартные группирующие переменные (например, *Пол - Gender, Национальность - Ethnic Origin;* см. также обзор *Таблицы сопряженности и таблицы флагов и заголовков*, стр. 1466).

Таблицы многомерных откликов. Если опция выбрана, пользователь может определить переменные с многомерными откликами и многомерные дихотомии (см. обзор *Многомерные отклики/дихотомии*, стр. 1474) в диалоговом окне *Таблицы многомерных откликов* (см. стр. 1563), и вычислить:

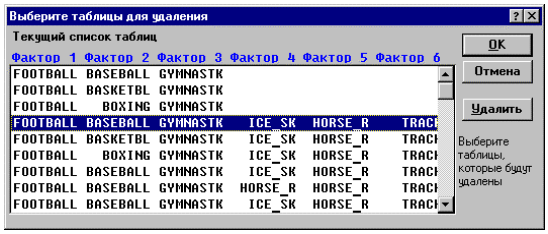
(1) Таблицы частот для выбранных переменных,

(2) Таблицы сопряженности, включающие такие переменные (одинарные группирующие переменные могут быть кросстабулированы с многомерными откликами и дихотомиями).

Многовходовые таблицы сопряженности

Задать таблицы. Используйте эту опцию, чтобы выбрать переменные для многовходовых таблиц сопряженности. После нажатия кнопки откроется стандартное окно выбора, состоящее из 6 списков переменных. Таблицы сопряженности будут вычислены для всех возможных комбинаций выбранных переменных. Можно также использовать опцию для добавления нового списка таблиц или отдельной таблицы к выбранным таблицам.

Просмотреть или удалить таблицы. Опция позволяет просмотреть и редактировать список таблиц.



Чтобы удалить таблицу, выделите ее имя и нажмите *Удалить*.

Таблицы флагов и заголовков

Таблица флагов и заголовков, по существу, - это двувходовая таблица, в которой табулируются не две отдельные переменные, а два списка группирующих переменных – один список табулируется в строках, другой – в столбцах. После нажатия кнопки *Задать таблицы* стандартное окно выбора с двумя списками переменных появится на экране.

Использовать только целые коды в выбранных переменных

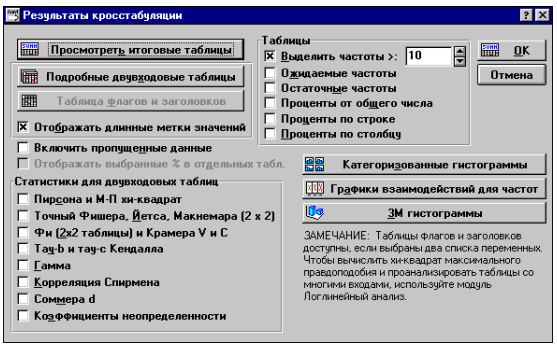
Используются все целые коды выбранных переменных для кросстабуляции.

Использовать все выбранные коды

В данной опции пользователь может определить коды для кросстабуляции.

Результаты кросстабуляции

Общая идея процедуры *Таблицы сопряженности и таблицы флагов и заголовков* обсуждалась в *Вводном обзоре* (стр. 1466).

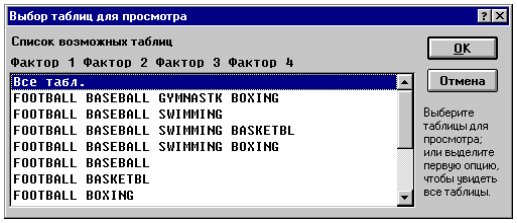


Copyright © StatSoft, 1995

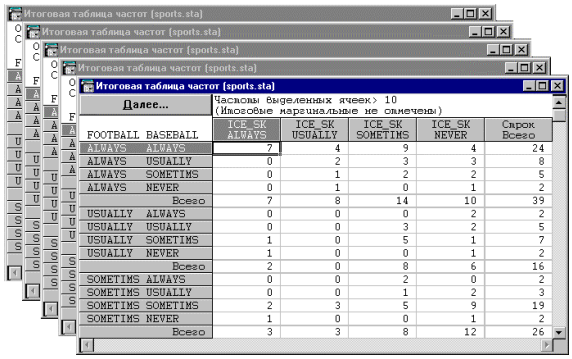
StatSoft®

Просмотреть итоговые таблицы

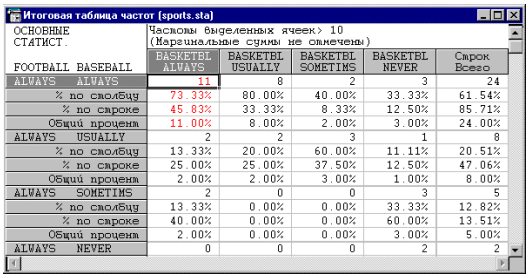
Опция строит итоговую таблицу для многовходовых таблиц сопряженности. Если определено более одной таблицы, то открывается диалоговое окно, в котором можно выбрать таблицы для просмотра. Выберите таблицу и нажмите *OK*.



Если выбраны *Все табл.*, будет построен каскад таблиц результатов или графиков.



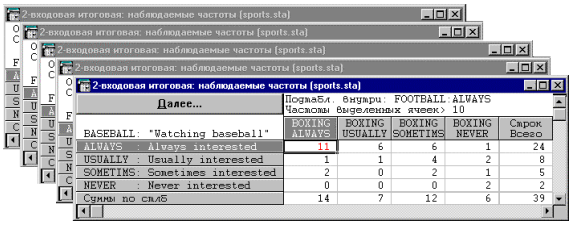
Итоговая таблица сопряженности содержит частоты исходной таблицы; последняя переменная (фактор) будет табулирован в столбцах, все оставшиеся факторы табулированы в строках. Если присутствуют больше двух факторов, то таблица может рассматриваться как совокупность нескольких двувходовых таблиц (для последних двух факторов). Таким образом, таблица до 6-го порядка может быть просмотрена в одной таблице данных. Детали вывода определяются параметрами в поле *Таблицы* (см. ниже).



Отметим, что *Ожидаемые частоты* или *Остаточные частоты* вычисляются с помощью *итеративной подгонки* (например, см. Haberman, 1972, 1974; см. также модуль *Логлинейный анализ*).

Подробные двувходовые таблицы

Опция строит таблицы результатов для двувходовых таблиц. Если определено более одной таблицы, то открывается диалоговое окно, в котором можно выбрать таблицы для отображения на экране. Для таблиц с более чем двумя переменными (факторами) строится каскад таблиц результатов для последних двух факторов внутри уровней других факторов.



Детали результатов определяются установками полей *Статистики для двувходовых таблиц* и *Таблицы* (см. ниже). Если выбраны все *Статистики для двувходовых таблиц*, то для каждой двувходовой таблицы вычисляются все запрашиваемые статистики.

Например, ниже показана таблица результатов для всех выбранных установок (заметим, что

StatSoft®
Copyright © StatSoft, 1995

некоторые статистики применимы только к таблицам 2 x 2).

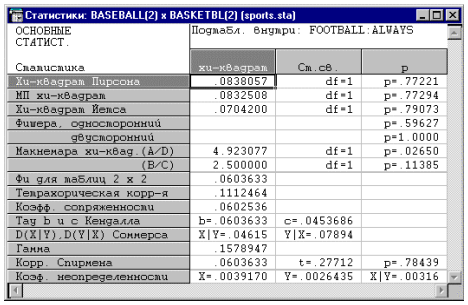
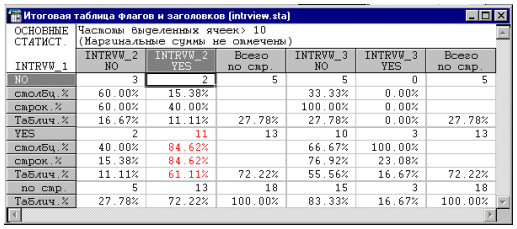


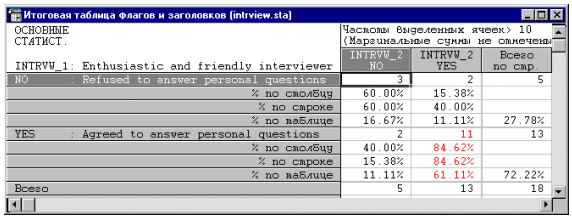
Таблица флагов и заголовков

Опция доступна, только если в предыдущем окне выбрана процедура *Таблица флагов и заголовков* (см. *Таблицы сопряженности и таблицы флагов и заголовков*, стр. 1558). В общем можно сказать, что в таблице флагов и заголовков один список будет табулирован в столбцах, другой список факторов - в строках.

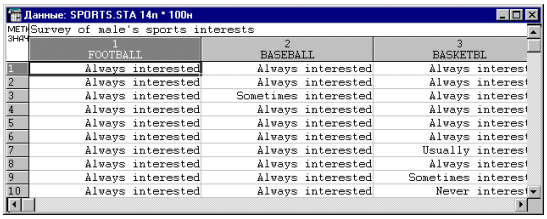


Отображать длинные метки значений

Если опция установлена, длинные метки значений будут отображены в первой колонке двувходовой таблицы результатов.



Если соответствующий фактор не имеет длинной метки, опция игнорируется. Значение метки можно просмотреть с помощью кнопки *Отображение числовых/текстовых значений* при нажатом вниз ключе CTRL.

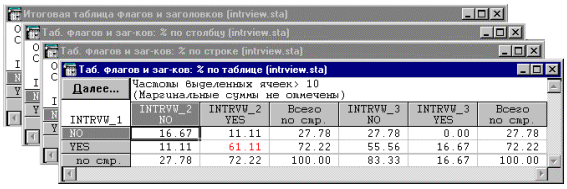


Включить пропущенные данные

Дополнительная категория (уровень) для пропущенных данных добавляется в каждую переменную (фактор) таблицы.

Отображать выбранные % в отдельных таблицах

Опция доступна, если выбрана одна из опций *Проценты...* (см. ниже). По умолчанию, если запрошено вычисление процентов, то они будут отображены в одной таблице с частотами. Если установлена данная опция, таблицы с процентами будут выведены на экран отдельно.



Статистики для двувходовых таблиц

Приводимые ниже опции определяют статистики, которые будут вычислены в подробных двувходовых таблицах. Если выбраны все статистики, отдельная таблица результатов будет показана вслед за каждой

Copyright © StatSoft, 1995

StatSoft®

двухходовой таблицей со статистиками для этой таблицы (см. выше).

Таблицы

Выделить частоты > ... Все частоты по строкам, которые превышают введенное значение, будут выделены.

Ожидаемые частоты. Для всех двухходовых таблиц вычисляются ожидаемые частоты в предположении независимости всех факторов (переменных) в таблице.

Итог. таблица: Ожидаемые частоты (sports.sta)

ОСНОВНЫЕ СТАТИСТ. Часовы выделенных ячеек> 10
хи-квадрат Пирсона: 72.9547, сс=9, p=.000000

FOOTBALL	BASEBALL ALWAYS	BASEBALL USUALLY	BASEBALL SOMETIMS	BASEBALL NEVER	Строк Всего
ALWAYS	10.92000	6.63000	14.43000	7.02000	39.0000
USUALLY	4.48000	2.72000	5.92000	2.88000	16.0000
SOMETIMS	7.28000	4.42000	9.62000	4.68000	26.0000
NEVER	5.32000	3.23000	7.03000	3.42000	19.0000
Все груп	28.00000	17.00000	37.00000	18.00000	100.0000

Таким образом, для *Подробных двухходовых таблиц* ожидаемые частоты вычисляются на основе маргинальных частот двух факторов; для *Итоговых таблиц* с более чем двумя факторами, ожидаемые частоты вычисляются на основе маргинальных частот всех факторов (предполагается, что между факторами нет взаимодействий).

Остаточные частоты. Если опция выбрана, то для всех двухходовых таблиц и итоговой таблицы вычисляются остаточные частоты: наблюдаемые частоты минус ожидаемые частоты.

Итог. таблица: Наблюдаемые частоты минус ожидаемые

ОСНОВНЫЕ СТАТИСТ. Часовы выделенных ячеек> 10
хи-квадрат Пирсона: 72.9547, сс=9, p=.000000

FOOTBALL	BASEBALL ALWAYS	BASEBALL USUALLY	BASEBALL SOMETIMS	BASEBALL NEVER	Строк Всего
ALWAYS	13.08000	1.37000	-9.43000	-5.02000	0.000000
USUALLY	-2.48000	2.28000	1.08000	-.88000	-.000000
SOMETIMS	-5.28000	-1.42000	9.38000	-2.68000	.000000
NEVER	-5.32000	-2.23000	-1.03000	8.58000	-.000000
Все груп	-.00000	0.00000	.00000	.00000	-.000000

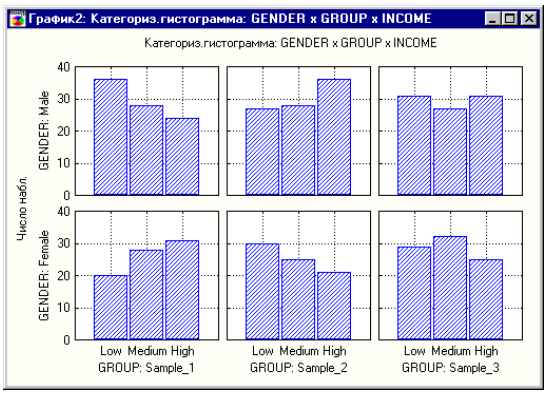
Проценты от общего числа. Опция вычисляет проценты для каждой ячейки в *Итоговых таблицах*, *Подробных двухходовых таблицах* и в *таблицах флагов и заголовков.*

Проценты по строке. Опция вычисляет проценты относительно общего количества наблюдений в сотоветствующей строке текущей таблицы, для каждой ячейки в *Итоговых таблицах*, *Подробных двухходовых таблицах* и в *Таблицах флагов и заголовков*.

Проценты по столбцу. Опция вычисляет проценты относительно общего количества наблюдений в соответствующем столбце текущей таблицы для каждой ячейки в *Итоговых таблицах*, *Подробных двухходовых таблицах* и в *Таблицах флагов и заголовков.*

Категоризованные гистограммы

После выбора этой опции, если происходит вычисление более одной таблицы, пользователь может выбрать, какие именно таблицы отобразить на графике. После этого *STATISTICA* произведет расчет и построение категоризованных гистограмм для выбранных таблиц.



Отметим, что каждый график может содержать информацию максимум о трех переменных (факторах); т.е. одна категоризованная гистограмма может отобразить до трех таблиц: Последний (измененный) фактор будет представлен в столбце графика; следующий перед ним будет представлен различными гистограммами, “табулированными” горизонтально; третий фактор будет представлен гистограммами “табулированными”

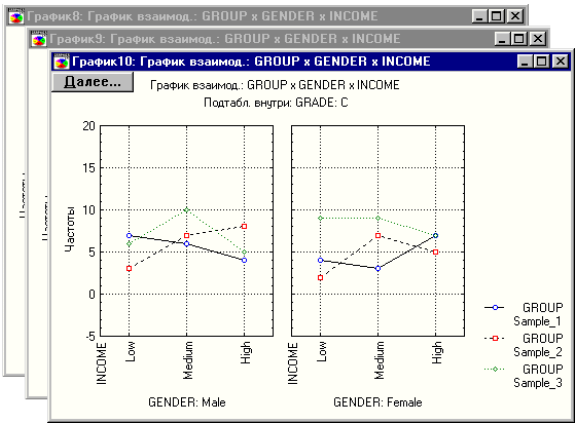
StatSoft®

Copyright © StatSoft, 1995

вертикально. Для таблиц с более чем тремя факторами будут построены каскады категоризованных гистограмм.

Графики взаимодействий частот

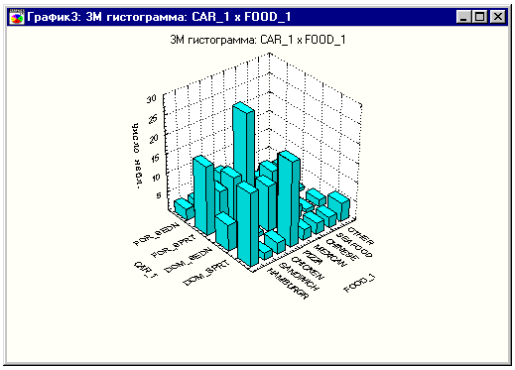
Другой способ представления распределения частот между тремя переменными (факторами) может быть получен с помощью линейного графика (*графика взаимодействий*).



Заметим, если в таблице имеется больше трех факторов, то программа построит последовательность графиков взаимодействия на каждой комбинации уровней оставшихся факторов. После нажатия кнопки пользователь может выбрать таблицу (или несколько таблиц) для графика (графиков). *STATISTICA* построит график (графики) взаимодействия для выбранных таблиц.

3М гистограммы

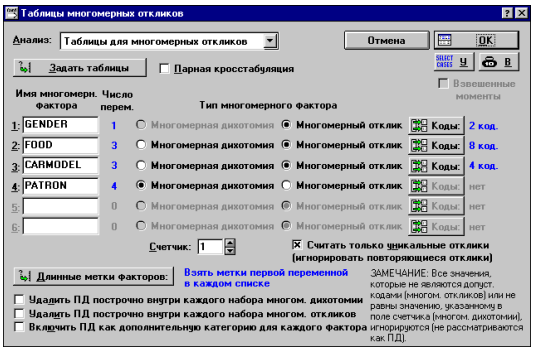
Опция строит 3М гистограмму для выбранной таблицы.



После нажатия кнопки пользователь может выбрать таблицу (или несколько таблиц) для графика (графиков). Каждая 3М гистограмма представляет совместное распределение частот двух выбранных переменных (факторов).

Задание многомерных откликов/дихотомий

После выбора в стартовой панели процедуры *Таблицы и заголовки* и выбора анализа *Таблиц для многомерных откликов* (см. ниже), вы увидите диалоговое окно *Таблицы многомерных откликов*.



Анализ

Выберите одну из двух установок.

Таблицы сопряженности. При выборе данной опции *STATISTICA* кросстабулирует

Copyright © StatSoft, 1995

StatSoft®

стандартные группирующие переменные (например, *Пол - Gender, Национальность - Ethnic Origin;* см. также обзор *Таблицы сопряженности и таблицы флагов и заголовков,* стр. 1466).

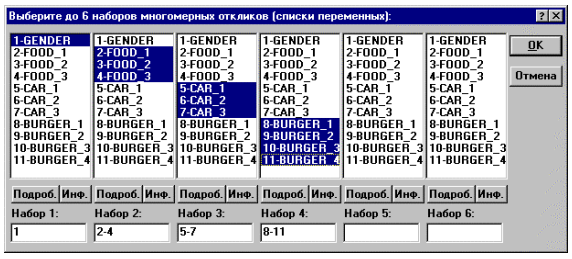
Таблицы для многомерных откликов. При выборе данной опции пользователь может задать другие переменные вместе с многомерными откликами и многомерные дихотомии (см. обзор *Многомерные отклики/дихотомии*, стр. 1474) и вычислить:

(1) Таблицы частот для таких переменных, или

(2) Вычислить таблицы сопряженности, включающие другие переменные (например, одинарные группирующие переменные могут быть кросстабулированы вместе с многомерными откликами и дихотомиями).

Задать таблицы

Опция открывает стандартное окно с шестью списками переменных.



Каждый список рассматривается как одна переменная (фактор) с многомерными откликами или многомерная дихотомия. Если в списке указана только одна переменная, то она рассматривается как простая группирующая переменная в кросстабуляции.

Парная кросстабуляция

Идея парной кросстабуляции факторов с многомерными откликами описана в обзоре *Парная кросстабуляция переменных с многомерными откликами* (стр. 1476). Опция применима, если выбран один или более многомерных факторов и, по крайней мере, два из них содержат одинаковое число переменных. В этом случае *STATISTICA* выполнит следующую обработку:

- Вычислит число исходных переменных в каждом факторе;
- Выполнит парную кросстабуляцию первых двух или более переменных с многомерными откликами.

Например, пусть имеется 3 многомерных фактора: первые 2 состоят из двух исходных переменных, третий – из трех. Тогда только первые два фактора обрабатываются.

Имя многомерного фактора

По умолчанию, программа возьмет имена факторов из имени первой переменой в соответствующем списке. Чтобы изменить имя фактора, просто введите новое имя в редактируемое поле.

Тип многомерного фактора

Многомерная дихотомия. Опция позволяет рассматривать соответствующий фактор как многомерную дихотомию.

Счетчик. *STATISTICA* вычислит, как часто данное значение присутствует на различных уровнях переменных, включенных в многомерные дихотомии.

Многомерный отклик. Опция позволяет рассматривать фактор как многомерный отклик. Если *Коды* не выбраны, *STATISTICA* автоматически табулирует все целые значения первой переменной в соответствующем списке. Если в списке только одна переменная, то при кросстабуляции она рассматривается как простая группирующая переменная.

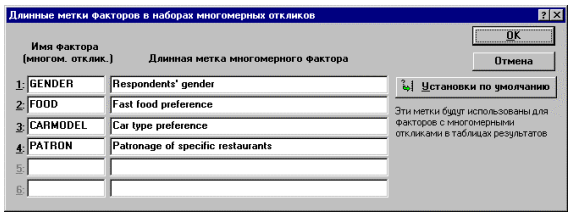
StatSoft®

Copyright © StatSoft, 1995

Считать только уникальные отклики (игнорировать повторяющиеся отклики). Если опция выбрана, одинаковые ответы респондентов будут игнорироваться. Например, если многомерный фактор *Предпочитаемый вид спорта* содержит две исходных переменных (первое предпочтение, второе предпочтение), и респондент назвал один и тот же вид спорта в обеих позициях, *STATISTICA* учитывает только один ответ. (В других программах такой ответ будет учтен дважды; не выбирайте данную опцию, тогда получите такую же таблицу частот или таблицу сопряженности.)

Длинные метки многомерных факторов

Опция вводит длинные метки для многомерных факторов.



По умолчанию *STATISTICA* присваивает многомерному фактору короткую и длинную метку первой переменной в списке (если такие метки имеются). Вы можете изменить метки, введя их в соответствующие поля. Чтобы использовать установки по умолчанию, нажмите кнопку *Установки по умолчанию*.

Удалить пропущенные данные построчно внутри каждого набора многомерной дихотомии

Если опция выбрана, то наблюдение объявляется пропущенным (“нет ответа”), если имеются пропуски хотя бы в одной переменной, составляющей многомерный фактор.

Удалить пропущенные данные построчно внутри каждого набора многомерных откликов

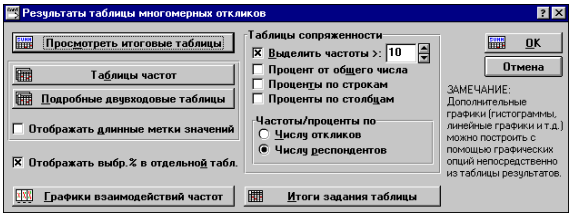
Если опция выбрана, то наблюдение объявляется пропущенным (“нет ответа”), если имеются пропущенные данные хотя бы в одной исходной переменной, составляющей многомерный фактор.

Включить пропущенные данные как дополнительную категорию для каждого фактора

По умолчанию, пропущенные данные игнорируются в анализе. Если опция выбрана, *STATISTICA* введет для пропущенных данных дополнительную категорию.

Просмотр таблиц с многомерными откликами и дихотомиями

Общая идея процедуры *Таблицы сопряженности и таблицы флагов и заголовков* обсуждалась во *Вводном обзоре* (стр. 1474).



Просмотреть итоговые таблицы

Итоговая таблица сопряженности содержит частоты исходной таблицы; последняя переменная (фактор) будет табулирована по столбцам, все оставшиеся факторы табулируются по строкам. Если присутствуют больше двух факторов, то таблица может рассматриваться как объединение нескольких двувходовых таблиц (для последних двух

Copyright © StatSoft, 1995

StatSoft®

факторов). Таким образом, в одной таблице результатов можно просмотреть таблицы до 6-го порядка. Детали вывода определяются установками в поле *Таблицы сопряженности*.

Таблицы частот

Опция строит последовательность таблиц частот, - одна таблица для одного фактора. На каждом уровне фактора вычисляются:

(1) число откликов;

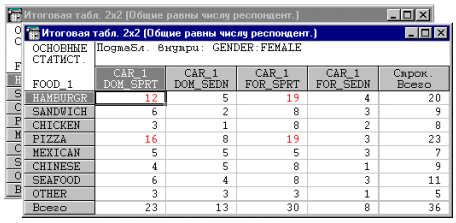
(2) процент ответов (относительно общего числа откликов);

(3) процент ответов относительно общего числа респондентов.

В многомерных откликах и дихотомиях каждый респондент может дать многомерный ответ; таким образом, проценты пункта (2) и пункта (3), в общем, различны.

Подробные двувходовые таблицы

Опция строит таблицу результатов с двувходовыми таблицами. Если определено более одной таблицы, то открывается диалоговое окно, в котором можно выбрать таблицы для просмотра. Для таблиц с более чем двумя переменными (факторами) строится каскад таблиц результатов для последних двух факторов внутри уровней других факторов.


Итоговая табл. 2x2 (Общие равны числу респондент.)

Итоговая табл. 2x2 (Общие равны числу респондент.)

ОСНОВНЫЕ СТАТИСТ. Подмабл. внедри: GENDER:FEMALE

FOOD_1	CAR_1 DOM_SPRT	CAR_1 DOM_SEDN	CAR_1 FOR_SPRT	CAR_1 FOR_SEDN	Строк. Всего
HAMBURGR	12	5	19	4	20
SANDWICH	6	2	8	3	9
CHICKEN	3	1	8	2	8
PIZZA	16	8	19	3	23
MEXICAN	5	5	5	3	7
CHINESE	4	5	8	1	9
SEAFOOD	6	4	8	3	11
OTHER	3	3	3	1	5
Всего	23	13	30	8	36

Детали результатов определяются опциями *Таблицы сопряженности* (см ниже).

Отображать длинные метки значений

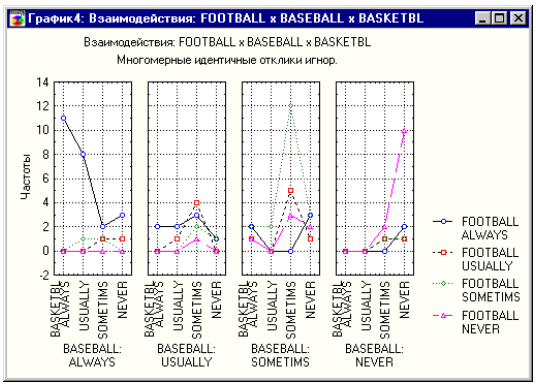
Если опция установлена, длинные метки значений будут отображены в первом столбце двувходовой таблицы результатов.

Отображать выбранные % в отдельной таблице

Опция доступна, если выбрана одна из опций *Проценты* содержащихся в поле *Таблицы сопряженности* (см. ниже). По умолчанию, если проценты запрошены, то они показываются в той же таблице результатов, что и частоты. При этом каждая ячейка таблицы результатов содержит несколько значений: частоты и проценты. Если выбрана данная опция, то проценты будут показаны в отдельной таблице результатов (что удобно, например, если нужно построить графики процентов непосредственно из таблиц результатов).

Графики взаимодействий частот

Один из способов представления частот на пересечении трех факторов заключается в построении линейного графика (*графика взаимодействий*).



На графике первый фактор представлен линиями различного вида и цвета; уровни следующего фактора обозначены “нижними-х” метками;

StatSoft®

Copyright © StatSoft, 1995

уровни последнего фактора – “верхними-x” метками. Если факторов более трех, то строится каскад графиков взаимодействия на каждой комбинации уровней оставшихся факторов.

Таблицы сопряженности

Выделить частоты > ... Если опция выбрана, то все частоты, превышающие введенное значение, будут выделены в таблице.

Проценты от общего числа. Опция вычисляет проценты в каждой ячейке в *Итоговых таблицах*, *Подробных двувходовых таблицах* и в *Таблице флагов и заголовков*. Проценты вычисляются относительно общего числа откликов или респондентов в текущей таблице.

Проценты по строке. Опция вычисляет проценты относительно суммы частот в соответствующей строке текущей таблицы в *Итоговых таблицах*, *Подробных двувходовых таблицах* и в *Таблице флагов и заголовков*.

Проценты по столбцу. Опция вычисляет проценты - относительно общего (суммарного) количества частот в соответствующем столбце текущей таблицы в *Итоговых таблицах*, *Подробных двувходовых таблицах* и в *Таблице флагов и заголовков*.

Частоты/проценты по ... Для переменных с многомерными откликами и многомерных дихотомий респондент может дать несколько ответов, поэтому несколько раз может учитываться при обработке анкет. Опция определяет, на чем будут основываться общие суммы и проценты в таблице результатов: на общем числе откликов или на общем числе респондентов.

Число откликов. Если опция выбрана, то маргинальные частоты вычисляются на основе откликов. Таким образом, они будут равны сумме (по строкам или по столбцам) соответствующих частот, содержащихся в ячейках. Так же вычисляются и проценты.

Число респондентов. Если опция выбрана, то маргинальные частоты основываются на числе респондентов.

Для многомерных дихотомий и многомерных откликов эти частоты обычно не совпадают с суммой частот в ячейках. Запрашиваемые проценты вычисляются в соответствии с этими маргинальными частотами.

Опция недоступна, если выбрана *Парная кросстабуляция* для многомерных откликов (см. *Задание многомерных откликов/дихотомий*, стр. 1563; см. также раздел *Парная кросстабуляция многомерных откликов*, стр. 1476).

Итоги задания таблицы

Процедура строит таблицу результатов со спецификациями факторов в текущей таблице. Для каждого фактора показывается:

(1) Имя и число уровней;

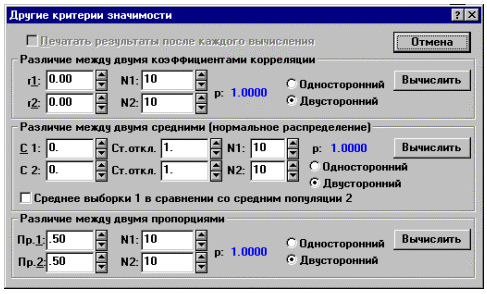
(2) Тип (обычная группирующая переменная, многомерный отклик, многомерная дихотомия);

(3) Имена всех исходных переменных, составляющих многомерные отклики и дихотомии.

Другие критерии значимости

При выборе процедуры *Другие критерии значимости* из стартовой панели, появляется диалоговое окно, в котором можно вычислить значимость различия между двумя коэффициентами корреляции, а также средними и пропорциями. Конкретные значения и размеры выборок задаются в полях ввода.

Copyright © StatSoft, 1995



Печатать результаты после каждого вычисления

Опция позволяет автоматически печатать результаты.

Различие между двумя коэффициентами корреляции

Процедура позволяет определить, имеется ли значимое различие между двумя коэффициентами корреляции Пирсона. Доступны *Односторонний* и *Двусторонний* критерии.

Различие между двумя средними (нормальное распределение)

Опции позволяют вычислить:

(1) Уровень значимости различия между двумя выборочными средними;

(2) Уровень значимости различия между средним выборки, и средним популяции, задаваемым в поле ввода (если выбрана опция *Среднее выборки 1 в сравнении со средним популяции 2*).

p-уровень рассчитывается с помощью *t*-критерия. Доступны *односторонний* и *двусторонний* критерии.

Различие между двумя пропорциями

Опция позволяет вычислить значимость разности двух пропорций, задаваемыми в полях ввода. Доступны *Односторонний* и *Двусторонний* критерии.

p-уровень основан на *t*-значении для соответствующего сравнения:

$$|t|=\sqrt{[(N_1*N_2)/(N_1+N_2)]}*|p_1-p_2|/\sqrt{(p*q)}$$

где

p равен $(p_1*N_1+p_2*N_2)/(N_1+N_2)$

q равен $(1-p)$.

Выделение значимых результатов

Статистически значимые результаты (*p<альфа*) выделяются в текущей таблице результатов. На панели инструментов появится кнопка *Параметры*, с помощью которой можно задать нужный уровень значимости.

После того, как введен новый уровень, значимые результаты будут выделены. Общая логика проверки гипотез описана в главе *Элементарные понятия статистики* (глава 8).

StatSoft®
Copyright © StatSoft, 1995

ПРЕДМЕТНЫЙ УКАЗАТЕЛЬ

А

Апостериорные критерии
- критерий Тьюки, 1552
- критерий Дункана, 1552
- критерий наименьшей значимой разности (НЗР), 1551
- критерий Ньюмена-Кеулса, 1552
- критерий Тьюки для выборок неравного размера, 1552
- критерий Шеффе, 1552

Апостериорные сравнения средних, 1463, 1544, 1551
Асимметрия, 1533

В

Вероятностные графики
- категоризованные нормальные вероятностные графики, 1530, 1539, 1546
- категоризованные нормальные вероятностные графики без тренда, 1539
- нормальные вероятностные графики, 1526, 1556
- нормальные вероятностные графики без тренда, 1527, 1556
- полунормальные вероятностные графики, 1527, 1556

Внутригрупповая вариация, 1460
Внутригрупповые корреляции, 1497, 1545
Выбор файла данных, 1484
Выбросы
- обзор, 1451
- количественный подход к выбросам, 1452

Вычисления с повышенной точностью, 1523, 1534

Г

Гамма статистика, 1473
Гистограммы
- 2М гистограмма, 1553
- 3М гистограмма, 1531, 1536, 1556, 1563
- категоризованные гистограммы, 1530, 1538, 1546, 1562

График средних и стандартных отклонений, 1548
Графики
- для таблиц результатов, 1487
- настройка, 1488

Графики взаимодействий
- график взаимодействий средних, 1547
- график взаимодействий частот, 1563, 1566
- категоризованные, 1529
- сопоставление факторов осям на графике, 1529

Графическое представление группировки, 1465
Группировка
- апостериорные сравнения средних, 1463, 1544
- внутригрупповые корреляции, 1463, 1497, 1541, 1542, 1545, 1548
- внутригрупповые описательные статистики, 1522, 1542, 1548
- дисперсионный анализ, 1544, 1550
- дисперсия, 1543, 1550
- итоговая таблица средних, 1495
- маргинальные средние, 1544
- моделирование структурными уравнениями (SEPATH), 1465
- описательные статистики для групп, 1462
- подробный анализ выбранных таблиц, 1542, 1549
- пример, 1494
- стандартные отклонения, 1543, 1550
- статистические тесты в группах, 1462
- суммы, 1543, 1550
- число наблюдений, 1543, 1550

Группировка в сравнении с анализом дискриминантных функций, 1464
Группировка и таблицы частот, 1464
Группирующая переменная, 1509
Группирующие переменные, 1489

Д

Диаграммы размаха
- для всех переменных, 1525, 1556
- категоризованная диаграмма размаха, 1529, 1546

Диаграммы рассеяния
- 2М диаграмма рассеяния (с именами), 1527, 1535
- 3М диаграмма рассеяния (с именами), 1528, 1536
- 3М диаграмма рассеяния с подогнанной поверхностью, 1528, 1536
- категоризованные диаграммы рассеяния, 1531, 1535, 1539, 1548
- матричная диаграмма рассеяния, 1528, 1535

Дисперсионный анализ, однофакторный, 1462, 1494, 1496, 1544, 1550
Доверительные границы для среднего, 1523
Другие критерии значимости, 1477, 1482, 1522, 1567

З

Зависимость
- прямая регрессии, 1451

Зависимые переменные, 1489
Заголовки
- интерпретация таблиц заголовков, 1469

И

Истинное среднее и доверительный интервал, 1447
Итеративная пропорциональная подгонка, 1560

К

Как использовать систему *STATISTICA*, 1483
Категоризованные графики
- гистограммы, 1530, 1538, 1546, 1562
- графики взаимодействий, 1529
- графики средних, 1529
- диаграмма размаха, 1529, 1546

Copyright © StatSoft, 1995

StatSoft®

диаграммы рассеяния, 1531, 1535, 1539, 1548
нормальные вероятностные графики, 1530, 1539, 1546
нормальные вероятностные графики без тренда, 1539
Квартили, 1523
Квартильный размах, 1533
Колмогорова-Смирнова критерий, 1524
Корреляции
обзор, 1449
R Спирмена, 1472
внутригрупповые корреляции, 1497, 1545, 1550
выделение значимых корреляций, 1486, 1534
гамма статистика, 1473
значимое различие между двумя корреляциями, 1457
значимость корреляции, 1451
измерение нелинейных зависимостей, 1454
как интерпретировать значение корреляции, 1451
корреляции в неоднородных группах, 1452
корреляция Пирсона, 1450, 1485
коэффициент d Соммера, 1473
линейная корреляция, 1450
ложные корреляции, 1457
матрицы, 1521, 1533
моделирование структурными уравнениями (SEPATH), 1457
назначение, 1449
нелинейные зависимости между переменными, 1453
отображать p и n, 1534
подробная таблица результатов, 1534
разведочный анализ корреляционных матриц, 1454
различие между двумя коэффициентами корреляции, 1568
смещение, вызванное попарным удалением пропущенных данных, 1456
тау Кендалла, 1473
упорядочивание данных, 1450
Корреляция r Пирсона, 1480
Коэффициент d Соммера, 1473
Коэффициент R-квадрат, 1534
Коэффициент сопряженности C, 1472
Коэффициенты корреляции, 1534
Коэффициенты неопределенности S(X,Y), S(X|Y), S(Y|X), 1473
Коэффициенты уравнения регрессии, 1534
Критерий Тьюки, 1552
Критерий Дункана, 1552
Критерий значимости, 1477
Критерий Колмогорова-Смирнова, 1555
Критерий Левена, 1538, 1540, 1544, 1550
Критерий Лиллиефорса, 1556
Критерий наименьшей значимой разности (НЗР), 1551
Критерий Ньюмена-Кеулса, 1552
Критерий разности дисперсий, 1491
Критерий хи-квадрат
Макнемара хи-квадрат, 1472
Пирсона хи-квадрат, 1471
поправка Йетса, 1471
Критерий Шеффе, 1552
Кросстабуляция
графическое представление, 1468
два на два таблицы, 1466
цель, 1466
Кросстабуляция данных, 1466, 1470, 1558, 1559
Кросстабуляция многомерных откликов и дихотомий, 1517
Кумулятивные проценты, 1555

Л

Логит преобразование частот, 1555
Ложные корреляции, 1457

М

Макнемара хи-квадрат, 1472
Маргинальные средние, 1544
Маргинальные частоты, 1467
Матричные графики
диаграммы рассеяния, 1528
Медиана, 1523
Меры сопряженности, 1472
Метод категоризации
все различные значения, 1554
заданные группирующие коды, 1554
начать с минимального значения, 1554
определенные пользователем категории, 1554
приближенные интервалы, 1554
размер шага, 1554
с текстовыми значениями, 1554
целые категории, 1554
число равных интервалов, 1554
Многовходовые таблицы
графическое представление, 1470
таблицы сопряженности, 1559
Многовходовые таблицы с контрольными переменными, 1469
Многомерные дихотомии, 1474
включить пропущенные данные как дополнительную категорию для каждого фактора, 1565
задание многомерной дихотомии, 1515, 1563
кросстабуляция многомерных дихотомий, 1476
определение простых группирующих переменых и переменных с многомерными откликами, 1511
определение факторов, 1513
парная кросстабуляция, 1476
переменные, 1510
пример, 1509
пропущенные данные, 1516
просмотр таблиц, 1565
соглашения об обозначениях, 1512
считать только уникальные отклики (игнорировать повторяющиеся отклики), 1565
трехвходовая таблица, 1519
Многомерные отклики
включить пропущенные данные как дополнительную категорию для каждого фактора, 1565
задание многомерных откликов, 1563
интерпретация таблиц частот с многомерными откликами, 1475
кодирование многомерных переменных, 1474
кросстабуляции многомерных откликов, 1476
многомерные идентичные отклики, 1513

StatSoft®

Copyright © StatSoft, 1995

определение простых группирующих переменных и переменных с многомерными откликами, 1511
определение факторов, 1513
парная кросстабуляция, 1476, 1513
переменные, 1474
повторные идентичные отклики, 1475
пример, 1509
пропущенные данные, 1513
просмотр таблиц, 1565
соглашения об обозначениях, 1512
считать только уникальные отклики (игнорировать повторяющиеся отклики), 1565
таблицы многомерных откликов, 1559
таблицы частот, 1514
трехвходовая таблица, 1519
Моделирование структурными уравнениями (SEPATH), 1457, 1460, 1465

Н

Набор многомерных откликов, 1564
Независимые переменные, 1489
Нормальная аппроксимация частот, 1524, 1555
Нормальность, 1448
Колмогорова-Смирнова критерий, 1524
Шапиро-Уилка W критерий, 1525

О

Объем выборки n, 1534
Однородности дисперсий, 1538, 1540, 1544, 1550
Однофакторный дисперсионный анализ
апостериорные сравнения средних, 1496
внутригрупповые корреляции, 1497
пример, 1494
Описательные статистики
обзор, 1447
асимметрия, 1533
верхняя квартиль, 1532
дисперсия, 1532
доверительные границы для среднего, 1523, 1532
квартили, 1523, 1532
квартильный размах, 1533
медиана, 1523, 1532
минимальные-максимальные значения, 1532
нижняя квартиль, 1532
размах, 1533
среднее, 1532
стандартная ошибка, 1532
стандартная ошибка асимметрии, 1533
стандартная ошибка эксцесса, 1533
стандартное отклонение, 1532
сумма, 1532
упорядочивание данных, 1447
цель, 1447
число наблюдений n, 1532
эксцесс, 1533
Определенные пользователем категории, 1557
Остаточные частоты, 1560

П

Пакетная обработка/печать, 1500, 1549
Пакетная обработка/печать таблиц группировки, 1542, 1549
Пакетная печать/построение графиков, 1501
Парная кросстабуляция, 1564
Переменная с многомерными откликами, 1509, 1510
Пирсона r, 1450, 1521
Подстановка среднего значения в пропущенные данные, 1456
Попарное удаление пропущенных данных, 1455
Попарные средние, 1551
Попарные стандартные отклонения, 1551
Поправка Йетса, 1471
Построение графиков для таблиц результатов, 1487
Построчное удаление пропущенных данных, 1455, 1523, 1534, 1538, 1541, 1542, 1549, 1555, 1565
Пробит преобразование частот, 1555
Проверка гипотез, 1567
различие между двумя коэффициентами корреляции, 1568
различие между двумя пропорциями, 1568
различие между двумя средними, 1568
Пропущенные данные
замена средним, 1456
попарное удаление, 1455
построчное удаление, 1455, 1523, 1534, 1538, 1541, 1542, 1549, 1555, 1565
Проценты по столбцу, 1562, 1567
Проценты по строке, 1562, 1567

Р

Разведочный анализ данных (РАД), 1453
Размах, 1533
Разность между средними (t-критерий), 1489

С

Свободный член, 1534
Смещение, вызванное попарным удалением пропущенных данных, 1456
Спирмена R, 1472, 1534
Стохастическая зависимость, 1473
Стьюдента t значение, 1534
Стьюдента t-критерий
t-критерий для зависимых выборок, 1460, 1481, 1521, 1541
t-критерий для независимых выборок, 1458, 1480, 1521, 1537, 1539, 1541
t-критерий с разделенными оценками дисперсий, 1538, 1540
графики, 1459, 1491
критерий Т-квадрат Хотеллинга, 1538
критерий Левена, 1538, 1540, 1544, 1550
матрицы t-критериев, 1461, 1541
моделирование структурными уравнениями (SEPATH), 1460
разности между средними, 1489
результаты, 1490
сравнение переменных, 1539

Copyright © StatSoft, 1995

Т

Таблицы 2 на 2, 1466
Таблицы сопряженности, 1506, 1563
Таблицы сопряженности для
    многомерных откликов и
    дихотомий, 1515
Таблицы флагов и заголовков, 1469,
    1502, 1522, 1558, 1561
Таблицы частот, 1465, 1498, 1522,
    1524, 1553, 1566
Тау Кендалла, 1473
Тетрахорическая корреляция, 1472
Точный критерий Фишера, 1471

У

Условия выбора наблюдений, 1557

Ф

Фи-квадрат, 1472

Х

Хи-квадрат Пирсона, 1471
Хотеллинга Т-квадрат критерий,
    1538

Ч

Частоты
    графики взаимодействий, 1563,
        1566
    кумулятивные, 1555
    маргинальные, 1467
    ожидаемые, 1562
    остатки, 1562
    относительные, 1555

Ш

Шапиро-Уилка W критерий, 1525,
    1556

Э

Эксцесс, 1533

StatSoft®

Copyright © StatSoft, 1995

Глава 10:

Быстрые основные статистики

Оглавление

Вводный обзор 1576
Выбор переменных для вычисления Быстрых основных статистик 1576
Описательные статистики, вычисленные с помощью опции Быстрые основные статистики 1577
Обзор 1577
Описательные статистики (из выпадающего меню Анализ) 1578
Описательные для… (из контекстного меню) 1579
Описательные по… (из контекстного меню) 1579
Корреляционные матрицы, вычисленные с помощью опции Быстрые основные статистики 1580
Обзор 1580
Корреляционные матрицы (из выпадающего меню Анализ) 1580
Корреляции для… (из контекстного меню) 1581
Корреляции по… (из контекстного меню) 1581
Таблицы частот, вычисленные с помощью опции Быстрые основные статистики 1582
Обзор 1582
Таблицы частот (из выпадающего меню Анализ) 1583
Частоты для… (из контекстного меню) 1584
Частоты по… (из контекстного меню) 1584
Быстрые основные статистики – расширенные возможности 1585
Переменные 1585
Коды 1586
Глобальные опции 1586
Описательные статистики 1586
Корреляции 1586
Распределение, нормальность 1587
Группировка 1588
Графические опции 1589
Вероятностный калькулятор 1594
Предметный указатель 1599

Подробное оглавление см. на следующей странице.

Copyright © StatSoft, 1995

Подробное оглавление

ВВОДНЫЙ ОБЗОР 1576
Выбор переменных для вычисления *Быстрых основных статистик* 1576

ОПИСАТЕЛЬНЫЕ СТАТИСТИКИ, ВЫЧИСЛЕННЫЕ С ПОМОЩЬЮ ОПЦИИ *БЫСТРЫЕ ОСНОВНЫЕ СТАТИСТИКИ* 1577
Обзор 1577
Описательные статистики (из выпадающего меню *Анализ*) 1578
Описательные для… (из контекстного меню) 1579
Описательные по… (из контекстного меню) 1579

КОРРЕЛЯЦИОННЫЕ МАТРИЦЫ, ВЫЧИСЛЕННЫЕ С ПОМОЩЬЮ ОПЦИИ *БЫСТРЫЕ ОСНОВНЫЕ СТАТИСТИКИ* 1580
Обзор 1580
Корреляционные матрицы (из выпадающего меню *Анализ*) 1580
Корреляции для… (из контекстного меню) 1581
Корреляции по… (из контекстного меню) 1581

ТАБЛИЦЫ ЧАСТОТ, ВЫЧИСЛЕННЫЕ С ПОМОЩЬЮ ОПЦИИ *БЫСТРЫЕ ОСНОВНЫЕ СТАТИСТИКИ* 1582
Обзор 1582
Таблицы частот (из выпадающего меню *Анализ*) 1583
Частоты для… (из контекстного меню) 1584
Частоты по… (из контекстного меню) 1584

***БЫСТРЫЕ ОСНОВНЫЕ СТАТИСТИКИ* – РАСШИРЕННЫЕ ВОЗМОЖНОСТИ 1585**
Переменные 1585
Коды 1586
Глобальные опции 1586
Построчное удаление пропущенных данных 1586
Отображать длинные имена переменных 1586
Описательные статистики 1586
Вычислить медианы и квартили 1586
Корреляции 1586
Сохранить 1587
p-уровни 1587
Построчное удаление ПД 1587
Распределение, нормальность 1587
Таблицы частот 1587
Гистограммы 1587
Ожидаемые нормальные частоты 1588
Критерии К-С и Лиллиефорса 1588
Группировка 1588

StatSoft®

Copyright © StatSoft, 1995

Число интервалов ... 1588
Целые интервалы ... 1589
Графические опции ... 1589
Диаграмма размаха для всех переменных ... 1589
Диаграммы размаха по группам ... 1590
Категоризованные гистограммы ... 1590
2М диаграмма рассеяния (с именами) ... 1591
Матричный ... 1591
3М диаграмма рассеяния (с именами) ... 1592
Поверхность ... 1592
Нормальные вероятностные графики ... 1592
Полунормальные вероятностные графики ... 1593
Нормальные вероятностные графики без тренда ... 1593
Категоризованные нормальные вероятностные графики ... 1593
Категоризованная диаграмма рассеяния ... 1593
3М гистограмма ... 1594

ВЕРОЯТНОСТНЫЙ КАЛЬКУЛЯТОР ... 1594
Калькулятор вероятностных распределений ... 1594
Вводный обзор ... 1594
Распределения ... 1595
Фиксированный масштаб ... 1595
Обратная функция распределения ... 1596
Двусторонняя ... 1596
1 – функция распределения (1 - ф.р.) ... 1596
Значение *x* ... 1596
p-значение ... 1596
Параметры распределения ... 1596
Печать ... 1596
График ... 1597
Вычислить ... 1597
Распределение коэффициента корреляции Пирсона ... 1597
Вводный обзор ... 1597
Параметры ... 1598
Двусторонний ... 1598
Вычислить *p* по *r* ... 1598
Вычислить *r* по *p* ... 1598
Вычислить *r* по z ... 1598
Печать ... 1598
Вычислить ... 1598

ПРЕДМЕТНЫЙ УКАЗАТЕЛЬ ... 1599

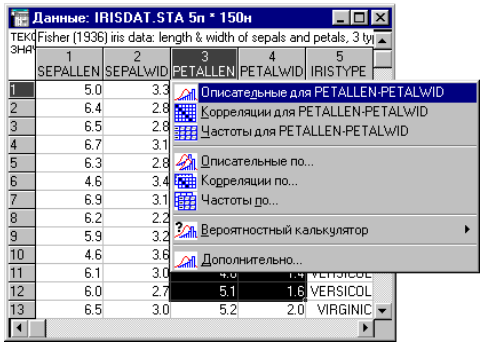
Copyright © StatSoft, 1995

Глава 10:

БЫСТРЫЕ ОСНОВНЫЕ СТАТИСТИКИ

ВВОДНЫЙ ОБЗОР

Быстрые основные статистики доступны из таблицы исходных данных или из таблицы результатов с помощью контекстного меню, вызываемого правой кнопкой мыши, а также из меню *Анализ* любого модуля системы *STATISTICA*. Их можно вызвать с помощью кнопки на панели инструментов. *Быстрые основные статистики* могут быть вызваны в любом моменте анализа (например, если нужно получить дополнительную информацию при просмотре результатов).



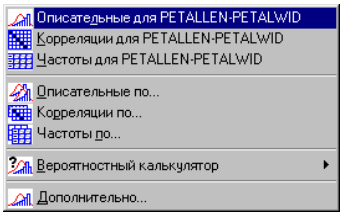
Эти процедуры включают выбор основных статистик, которые можно вычислить для больших списков переменных (например, вычислить корреляции для всех переменных в файле данных), а также для группированных переменных (для каждого значения (кода) выбранной группирующей переменной).

Меню *Быстрые основные статистики* также содержит быстрый доступ к *Вероятностному калькулятору* (детали см. на стр. 1594).

Выбор переменных для вычисления *Быстрых основных статистик*

Одно из основных преимуществ опций *Быстрые основные статистики* состоит в том, что для них не требуется выбор переменных для анализа с помощью стандартного диалогового окна. Вместо этого переменные автоматически выбираются из высвеченного блока таблицы. Система *STATISTICA* просит выбрать переменные, только если блок не содержит достаточного числа переменных, или если требуется группирующая переменная (Заметим, если *Быстрые основные статистики* вызваны из выпадающего меню *Анализ*, то переменные для анализа выбираются с помощью последовательности окон выбора переменных).

Во всех таблицах исходных данных (и тех таблицах результатов, которые включают только один список переменных или одну группирующую переменную и список переменных) контекстное меню имеет следующий вид:

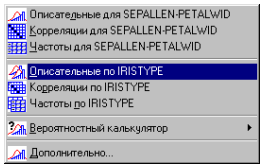


Если вы анализируете данные с группирующими переменными (*по...*), система *STATISTICA* откроет стандартное диалоговое окно *Выбора переменных,* в котором можно выбрать нужную группирующую переменную. Если таблица результатов имеет матричный формат, то предопределенный анализ данных, разбитых на группы, доступен из меню *Быстрые основные*

StatSoft

Copyright © StatSoft, 1995

статистики (при этом система *STATISTICA* не попросит выбрать группирующую переменную).



Если переменные не отмечены текущим положением курсора, программа просит выбрать переменные из списка.

При вычислении *Быстрых основных статистик* система *STATISTICA* рассматривает выбранные наблюдения анализируемых переменных (условия выбора наблюдений и веса можно изменить в диалоговом окне *Быстрые основные статистики - Расширенные возможности*, см. стр. 1585).

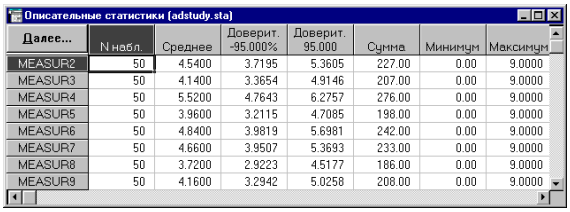
Смотрите также *Статистики блока* (глава 3), которые вычисляют описательные статистики и строят статистические графики для выделенных блоков данных.

ОПИСАТЕЛЬНЫЕ СТАТИСТИКИ, ВЫЧИСЛЕННЫЕ С ПОМОЩЬЮ ОПЦИИ *БЫСТРЫЕ ОСНОВНЫЕ СТАТИСТИКИ*

Обзор

STATISTICA вычисляет широкий набор описательных статистик для выбранных переменных. Для этого используйте *Описательные статистики* из меню *Анализ - Быстрые основные статистики*, *Описательные для…* или *Описательные по…* из контекстного меню вызываемого правой кнопкой мыши. Используйте также диалоговое окно *Быстрые основные статистики - Расширенные возможности*, чтобы настроить вычисления, например, чтобы выбрать способ обработки пропущенных значений, показать длинные имена переменных, вычислить медианы и квартили и т.д. (сделанный выбор сохранится при следующем обращении к процедурам *Быстрые основные статистики*).

На каждом уровне группирующей переменной (если она выбрана) будут построены таблицы с описательными статистиками выбранных данных.


Описательные статистики (adstudy.sta)

Далее...	N набл.	Среднее	Доверит. -95.000%	Доверит. 95.000	Сумма	Минимум	Максимум
MEASUR2	50	4.5400	3.7195	5.3605	227.00	0.00	9.0000
MEASUR3	50	4.1400	3.3654	4.9146	207.00	0.00	9.0000
MEASUR4	50	5.5200	4.7643	6.2757	276.00	0.00	9.0000
MEASUR5	50	3.9600	3.2115	4.7085	198.00	0.00	9.0000
MEASUR6	50	4.8400	3.9819	5.6981	242.00	0.00	9.0000
MEASUR7	50	4.6600	3.9507	5.3693	233.00	0.00	9.0000
MEASUR8	50	3.7200	2.9223	4.5177	186.00	0.00	9.0000
MEASUR9	50	4.1600	3.2942	5.0258	208.00	0.00	9.0000

Доступны следующие описательные статистики:

- число наблюдений N;
- среднее;
- доверительные границы для среднего;
- сумма;
- минимум;
- максимум;
- размах;
- дисперсия;
- стандартное отклонение;
- стандартная ошибка среднего;
- асимметрия;
- эксцесс;
- стандартные ошибки для асимметрии и эксцесса;
- медианы и квартили, выбранные ранее в окне *Расширенные возможности.*

Более подробная информация о вычислении этих статистик дана в *Приложении II*. Определенный метод вычисления медиан и квартилей можно выбрать в диалоговом окне системы *STATISTICA Параметры по умолчанию: общие настройки*, доступном из выпадающего меню *Сервис*, см. также *Окно таблицы исходных данных*, глава 3).

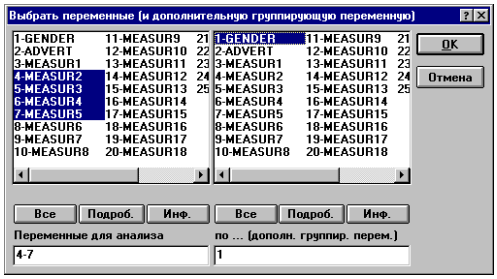
Copyright © StatSoft, 1995
StatSoft®

Настройка вычислений с помощью диалогового окна *Быстрые основные статистики – Расширенные возможности.* Диалоговое окно *Расширенные возможности* (доступное с помощью опции *Быстрые основные статистики - Дополнительно...*, см. стр. 1585) предоставляет дополнительные статистические процедуры (выбор, сделанный в этом окне, сохранится и при следующем обращении к процедурам *Быстрые основные статистики*).

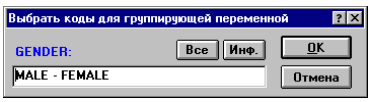
Например, с помощью опций данного окна можно вычислить медианы и квартили. Дополнительно, можно построчно удалить пропущенные данные, а также с помощью глобальной опции окна *Быстрые основные статистики - Расширенные возможности* добавить длинные имена переменных в таблицу результатов.

Описательные статистики (из выпадающего меню *Анализ*)

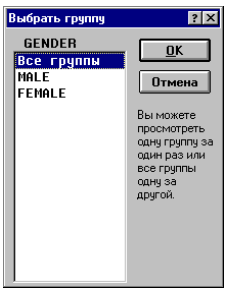
Выбор этой процедуры в меню *Анализ*, открывает диалоговое окно *Выбрать переменные (и дополнительную группирующую переменную)*, в котором можно выбрать переменные (до 4,096) для анализа и (дополнительно) одну группирующую переменную.



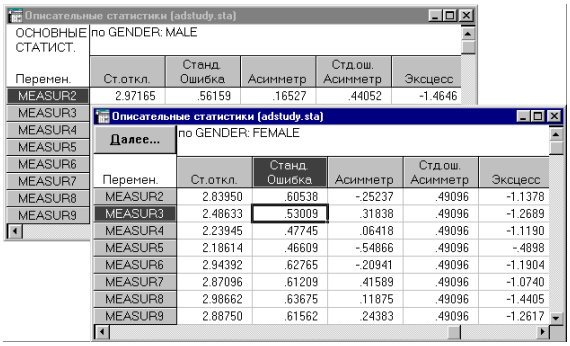
Если вы не выбираете группирующую переменную, нажмите *OK*, чтобы сразу вычислить таблицу с описательными статистиками. Если группирующая переменная выбрана, нажмите *OK* и далее в окне *Выбрать коды для группирующей переменной* выберите коды группирующей переменной.



После выбора кодов откроется окно *Выбрать группу*.



В этом окне можно выбрать группу, чтобы показать на экране каскад таблиц результатов с описательными статистиками для всех групп (нажмите *OK* для выбора опции по умолчанию) или только для определенных групп (по одной таблице результатов на группу). Нажмите *OK* для завершения анализа.



Настройка вычислений. Для настройки вычислений используйте диалоговое окно *Быстрые основные статистики - Расширенные возможности* (доступное с помощью опции *Быстрые основные статистики -*

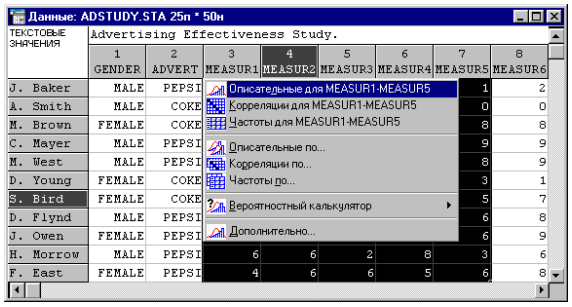
StatSoft®
Copyright © StatSoft, 1995

Дополнительно...). Оно позволяет, например, выбрать вычисление медиан и квартилей, построчное удаление пропущенных данных, а также включить отображение длинных имен переменных в таблице результатов (выбранные настройки будут использоваться при следующем использовании опции *Быстрые основные статистики - Расширенные возможности* без использования окна *Расширенные возможности*).

Описательные для… (из контекстного меню)

Если нужно вычислить некоторые описательные статистики выбранных переменных без промежуточных окон (например, для выбранных переменных или кодов), используйте опцию – *Описательные для…* из контекстного меню *Быстрые основные статистики*, доступного с помощью правой кнопки мыши.

STATISTICA сразу вычислит описательные статистики для выделенных значений в таблице исходных данных без запроса дополнительной информации.

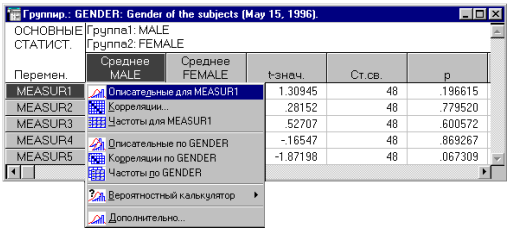


Для того чтобы вычислить описательные статистики при использовании группирующей переменной, используйте опцию *Описательные по…* или опцию *Дополнительно…* из контекстного меню, вызываемого правой кнопкой мыши или опцию *Быстрые основные статистики - Описательные статистики* из меню *Анализ*.

Описательные по… (из контекстного меню)

Опция позволяет вычислить описательные статистики для выделенных значений на разных уровнях группирующей переменной. После выбора опции, откроется диалоговое окно *Выбрать коды для группирующей переменной*, в котором выбираются коды (группы) (дополнительно, можно выбрать или изменить выбор анализируемых переменных, выбрав до 4,096 переменных). Нажмите *OK* и откройте окно *Выбрать группу*. В этом диалоговом окне можно выбрать для просмотра последовательности таблиц результатов с описательными статистиками для всех групп (нажмите *OK,* чтобы выбрать опцию по умолчанию) или только для определенных групп (по одной таблице результатов на каждую группу).

Если вы щелкнете правой кнопкой мыши в таблице результатов, а группирующая или вторая переменная используется в анализе, группирующая переменная будет автоматически выбрана и показана в меню.



Опция откроет окно *Выбрать группу*, в окне можно выбрать для просмотра таблицы результатов с описательными статистиками для всех групп или только для определенных групп (по одной таблице результатов на каждую группу). Нажмите *OK* для завершения анализа.

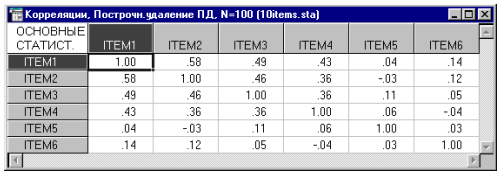
Copyright © StatSoft, 1995

StatSoft®

КОРРЕЛЯЦИОННЫЕ МАТРИЦЫ, ВЫЧИСЛЕННЫЕ С ПОМОЩЬЮ ОПЦИИ *БЫСТРЫЕ ОСНОВНЫЕ СТАТИСТИКИ*

Обзор

STATISTICA вычисляет корреляции между выбранными переменными на каждом уровне группирующей переменной (если такая выбрана). Для вычисления корреляций выберите либо опцию *Корреляционные матрицы* из меню *Анализ - Быстрые основные статистики*, либо *Корреляции для…*, либо *Корреляции по…* из меню, вызываемого правой кнопкой мыши.



ОСНОВНЫЕ СТАТИСТ.	ITEM1	ITEM2	ITEM3	ITEM4	ITEM5	ITEM6
ITEM1	1.00	.58	.49	.43	.04	.14
ITEM2	.58	1.00	.46	.36	-.03	.12
ITEM3	.49	.46	1.00	.36	.11	.05
ITEM4	.43	.36	.36	1.00	.06	-.04
ITEM5	.04	-.03	.11	.06	1.00	.03
ITEM6	.14	.12	.05	-.04	.03	1.00

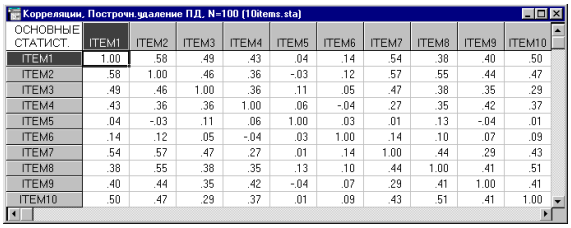
Размер корреляционной матрицы ограничен только объемом памяти компьютера.

Настройка вычислений с помощью диалогового окна *Быстрые основные статистики – Расширенные возможности*. Диалоговое окно *Расширенные возможности* (доступное с помощью опции *Быстрые основные статистики - Дополнительно...*, см. стр. 1585) обеспечивает вас дополнительными возможностями (сделанный выбор сохранится и при следующем обращении к процедурам *Быстрые основные статистики*). Можно, например, отобразить *p-уровни* для корреляций, выбрав опцию *p-уровни*. Можно выбрать построчный или парный способ удаления пропущенных данных (по умолчанию принято построчное удаление наблюдений с пропусками), а также включить длинные имена переменных в таблицу результатов с помощью глобальной опции в диалоговом окне *Быстрые основные статистики - Расширенные возможности*.

Корреляционные матрицы (из выпадающего меню *Анализ*)

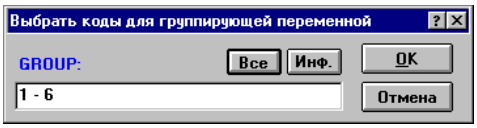
Когда выбрана эта опция из меню *Анализ*, будет открыто диалоговое окно *Выбрать переменные (и дополнительную группирующую переменную)*, в котором можно выбрать переменные для анализа и (дополнительно) группирующую переменную.

Если группирующая переменная не выбрана, нажмите *OK*, чтобы увидеть корреляционную матрицу.



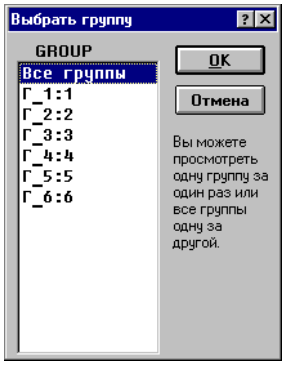
ОСНОВНЫЕ СТАТИСТ.	ITEM1	ITEM2	ITEM3	ITEM4	ITEM5	ITEM6	ITEM7	ITEM8	ITEM9	ITEM10
ITEM1	1.00	.58	.49	.43	.04	.14	.54	.38	.40	.50
ITEM2	.58	1.00	.46	.36	-.03	.12	.57	.55	.44	.47
ITEM3	.49	.46	1.00	.36	.11	.05	.47	.38	.35	.29
ITEM4	.43	.36	.36	1.00	.06	-.04	.27	.35	.42	.37
ITEM5	.04	-.03	.11	.06	1.00	.03	.01	.13	-.04	.01
ITEM6	.14	.12	.05	-.04	.03	1.00	.14	.10	.07	.09
ITEM7	.54	.57	.47	.27	.01	.14	1.00	.44	.29	.43
ITEM8	.38	.55	.38	.35	.13	.10	.44	1.00	.41	.51
ITEM9	.40	.44	.35	.42	-.04	.07	.29	.41	1.00	.41
ITEM10	.50	.47	.29	.37	.01	.09	.43	.51	.41	1.00

Если группирующая переменная выбрана, нажмите *OK* и далее выберите коды группирующей переменной в окне *Выбрать коды для группирующей переменной*.



После выбора кодов откроется окно *Выбрать группу*.

StatSoft®
Copyright © StatSoft, 1995

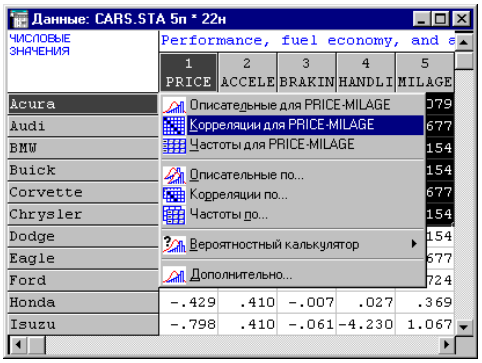


В этом диалоговом окне можно выбрать группы, чтобы отобразить на экране корреляционные матрицы для всех групп (нажмите *OK* для выбора опции по умолчанию) или только для определенных групп (одна таблица результатов для каждой группы). Нажмите *OK,* чтобы завершить анализ.

Настройка вычислений. Чтобы настроить вычисления, используйте окно *Быстрые основные статистики - Расширенные возможности* (доступное с помощью опции *Быстрые основные статистики - Дополнительно...* см. стр. 1585). Например, можно выбрать построчный или попарный способ удаления пропущенных данных (по умолчанию принято построчное удаление наблюдений с пропусками), а также включить длинные имена переменных в таблицу результатов с помощью глобальной опции в диалоговом окне *Быстрые основные статистики - Расширенные возможности*.

Корреляции для… (из контекстного меню)

Если нужно вычислить квадратную корреляционную матрицу высвеченных значений без промежуточных диалоговых окон (например, для выбранных переменных или кодов), используйте опцию *Быстрые основные статистики* из контекстного меню, вызываемого правой кнопкой мыши.

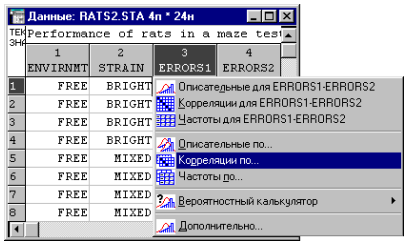


Чтобы вычислить корреляции для сгруппированных данных, используйте опцию *Корреляции для...* или опцию *Дополнительно...* из контекстного меню, вызываемого правой кнопкой мыши. Можно также воспользоваться опцией *Быстрые основные статистики - Корреляционные матрицы* из выпадающего меню *Анализ*. Чтобы вычислить прямоугольную корреляционную матрицу, используйте опцию *Корреляционные матрицы*.

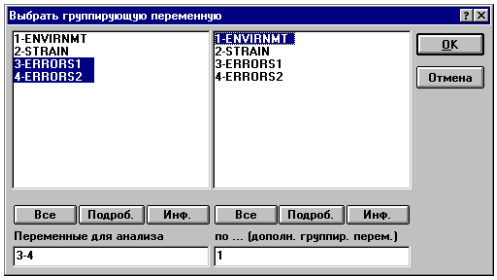
Корреляции по… (из контекстного меню)

Опция *Быстрые основные статистики* (в меню, вызываемом правой кнопкой мыши) вычисляет квадратную корреляционную матрицу выделенных переменных, разбитых на группы. Если вы выберете эту опцию из контекстного меню, вызываемого правой кнопкой мыши

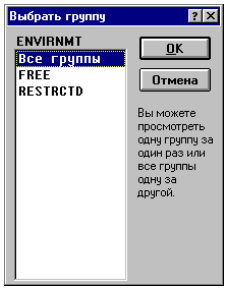
Copyright © StatSoft, 1995
StatSoft®



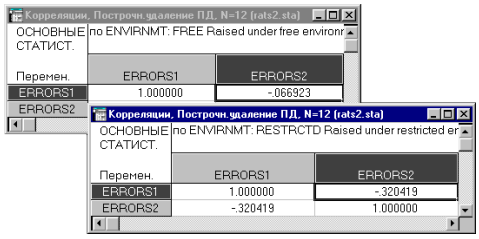
откроется окно *Выбрать группирующую переменную,* в котором можно выбрать единственную группирующую переменную (дополнительно, можно произвести или отменить выбор переменных для анализа).



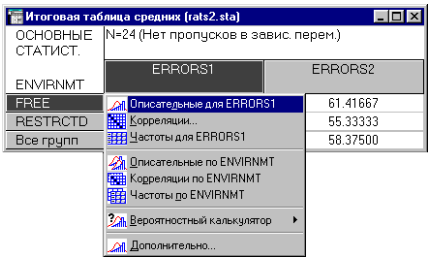
Нажмите *OK* и откройте окно *Выбрать группу*.



Нажав *OK,* завершите анализ.



Если щелкнуть по таблице результатов правой кнопкой мыши, группирующая переменная (если она присутствует в анализе), будет автоматически показана в меню *Быстрые основные статистики*.

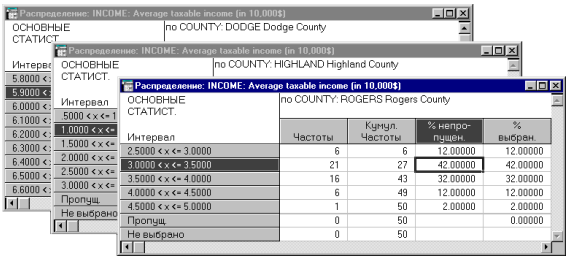


В этом случае будет открыто описанное выше окно *Выбрать группу*.

ТАБЛИЦЫ ЧАСТОТ, ВЫЧИСЛЕННЫЕ С ПОМОЩЬЮ ОПЦИИ *БЫСТРЫЕ ОСНОВНЫЕ СТАТИСТИКИ*

Обзор

Если вы выберете опцию *Таблицы частот* в меню *Анализ - Быстрые основные статистики* или опцию *Частоты для…*, или *Частоты по …* с помощью правой кнопки мыши, *STATISTICA* вычислит каскад таблиц частот для выбранных переменных, одну таблицу результатов для одной переменной, на каждом уровне группирующей переменной (если она выбрана).



В итоговые статистики для выделенных значений входят (для каждой группы):

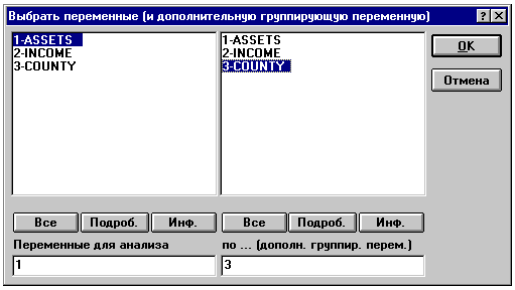
StatSoft®
Copyright © StatSoft, 1995

- частоты;
- кумулятивные частоты;
- процент значений без пропусков;
- процент выбранных значений;
- кумулятивный процент значений без пропусков;
- 100 % - процент значений без пропусков;
- кумулятивный процент выбранных значений;
- 100 % - процент выбранных значений.

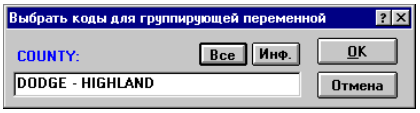
Настройка вычислений с помощью диалогового окна *Быстрые основные статистики – Расширенные возможности.* Диалоговое окно *Расширенные возможности* (доступное с помощью опции *Быстрые основные статистики - Дополнительно...*, см. стр. 1585) дает дополнительные возможности (сделанный выбор сохранится и при следующем обращении к процедурам *Быстрые основные статистики*). Можно включить длинные имена переменных в таблицу результатов с помощью глобальной опции в диалоговом окне *Быстрые основные статистики - Расширенные возможности* (выбор, сделанный в этом диалоговом окне останется в силе, когда вы снова будете использовать *Быстрые основные статистики - Таблицы частот*).

Таблицы частот (из выпадающего меню *Анализ*)

Если выбрана эта опция из меню *Анализ*, открывается окно *Выбрать переменные (и дополнительную группирующую переменную)*, в котором можно выбрать анализируемые переменные (до 4096, см. ранее) и (дополнительно) группирующую переменную.



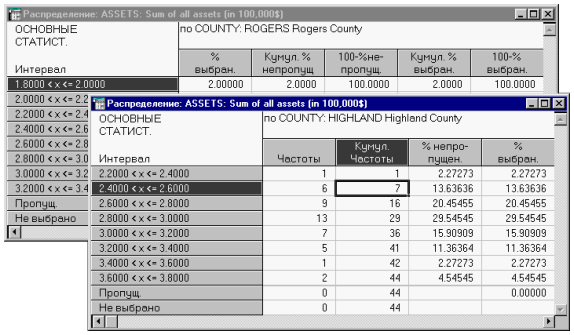
Если группирующая переменная не выбрана, нажмите *OK*, чтобы вычислить таблицу с описательными статистиками. Если группирующая переменная выбрана, нажмите *OK* и далее выберите коды группирующей переменной в окне *Выбрать коды для группирующей переменной*.



После выбора кодов откроется окно *Выбрать группу*.

В этом диалоговом окне можно выбрать вывод на экран таблиц результатов с частотами для всех групп (нажмите *OK* для выбора этой опции, установленной по умолчанию) или только для определенных групп (одна таблица результатов на одну группу). Нажмите *OK,* чтобы завершить анализ.

Copyright © StatSoft, 1995
StatSoft®

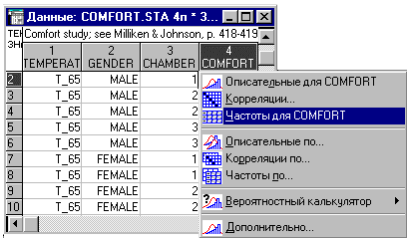


Настройка вычислений. Чтобы настроить вычисления, используйте диалоговое окно *Быстрые основные статистики - Расширенные возможности* (доступное с помощью опции *Быстрые основные статистики - Дополнительно...*, см. стр. 1585), чтобы настроить вычисления. Например, чтобы вычислить медианы и квартили, построчно удалить пропущенные данные, а также включить длинные имена переменных в таблицу результатов с помощью соответствующей опции в диалоговом окне *Быстрые основные статистики - Расширенные возможности* (выбор, сделанный в этом диалоговом окне сохранится, когда вы снова будете использовать *Быстрые основные статистики - Таблицы частот*).

Частоты для… (из контекстного меню)

Если хотите вычислить частоты для выбранных переменных без промежуточных диалоговых окон (например, для выбранных переменных или кодов), используйте опцию *Быстрые основные статистики* из контекстного меню, вызываемого правой кнопкой мыши.

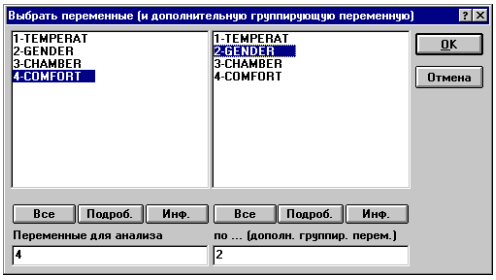
Система *STATISTICA* сразу вычислит частоты для выделенных в таблице значений.



Для того чтобы вычислить частоты для сгруппированных переменных, используйте опцию *Частоты по…* или опцию *Дополнительно...* из контекстного меню, вызываемого правой кнопкой мыши, или опцию *Быстрые основные статистики – Таблицы частот* из выпадающего меню *Анализ*.

Частоты по… (из контекстного меню)

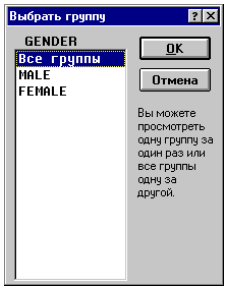
Эта опция позволяет вычислить частоты выделенных значений на разных уровнях группирующей переменной.



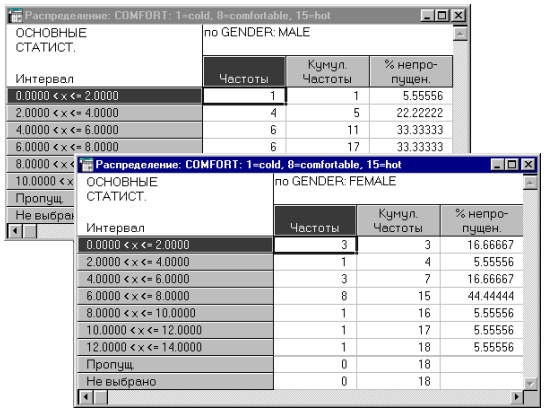
После выбора опции, откроется диалоговое окно *Выбрать коды для группирующей переменной*, в котором можно выбрать коды (группы) (дополнительно, можно выбрать или изменить выбор анализируемых переменных, выбрав до 4,096 переменных). Нажмите *OK* в диалоговом окне, откроется окно *Выбрать группу*. В этом диалоговом окне можно выбрать для дальнейшего просмотра последовательности таблиц результатов с частотами для всех групп (нажмите *OK*, чтобы выбрать опцию по

StatSoft®
Copyright © StatSoft, 1995

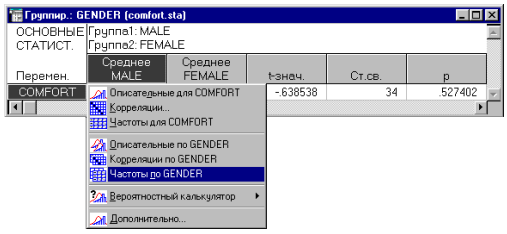
умолчанию) или только для определенных групп (по одной таблице результатов для каждой группы).



Нажмите *OK* для завершения анализа.



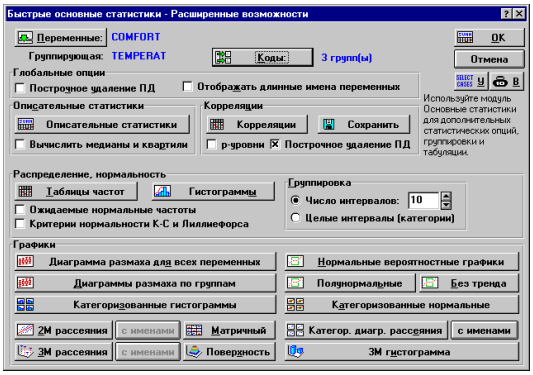
Если вы щелкнете по таблице результатов правой кнопкой мыши, а группирующая переменная уже использовалась в анализе, то она будет представлена в меню *Быстрые основные статистики*.



В этом случае после выбора этой опции откроется диалоговое окно *Выбрать группу,* в котором можно выбрать построение таблиц частот для всех групп или для определенных групп (отдельная таблица результатов для каждой группы). Нажмите *OK*, чтобы завершить анализ.

БЫСТРЫЕ ОСНОВНЫЕ СТАТИСТИКИ – РАСШИРЕННЫЕ ВОЗМОЖНОСТИ

Опция *Быстрые основные статистики* (из выпадающего меню *Анализ* или контекстного меню правой кнопки мыши) предлагает расширенные возможности для описательных статистик, корреляций, таблиц частот, а также дополнительные графические опции.



Параметры, установленные в этом окне, будут применены и при следующем использовании опции *Быстрые основные статистики* без повторного использования окна *Расширенные возможности* (например, из выпадающего меню *Анализ*). Доступные опции описаны ниже.

Переменные

Нажмите эту кнопку, чтобы открыть диалоговое окно *Выбрать переменные (и дополнительную группирующую переменную)*. Здесь можно выбрать переменные для анализа и (дополнительно) группирующую переменную.

Copyright © StatSoft, 1995

StatSoft®

(Заметьте, что имя группирующей переменной появится ниже этой кнопки.)

Коды

Нажмите кнопку, чтобы открыть диалоговое окно *Выбрать коды для группирующей переменной,* в котором можно выбрать коды.

Глобальные опции

За исключением построчного удаления пропущенных данных при вычислении корреляций (которое задается независимо от этих опций, см. *Корреляции*, ниже), следующие две опции применяются глобально ко всем другим опциям этого диалогового окна.

Построчное удаление пропущенных данных

Если выбрано построчное удаление пропущенных данных, *STATISTICA* игнорирует все наблюдения, имеющие пропуски в какой-либо переменной из выбранного списка. Если опция не выбрана, то производится попарное удаление пропущенных значений.

Отображать длинные имена переменных

Если выбрана эта опция, в таблице результатов показываются длинные имена переменных (если имеются) вместе с короткими именами.

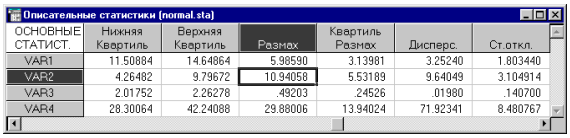
Описательные статистики

Эта опция позволяет вычислить описательные статистики для выбранных переменных (можно выбрать до 4096 переменных).

Ниже описана опция *Вычислить медианы и квартили*.

Вычислить медианы и квартили

Опция вычисляет, дополнительно к списку описательных статистик, приведенному на стр. 1577, медианы (50% процентиль), квартили (25% и 75% процентили) и *квартильный размах* для всех переменных.


Описательные статистики (normal.sta)

ОСНОВНЫЕ СТАТИСТ.	Нижняя Квартиль	Верхняя Квартиль	Размах	Квартиль Размах	Дисперс.	Ст.откл.
VAR1	11.50884	14.64864	5.98590	3.13981	3.25240	1.803440
VAR2	4.26482	9.79672	10.94058	5.53189	9.64049	3.104914
VAR3	2.01752	2.26278	.49203	.24526	.01980	.140700
VAR4	28.30064	42.24088	29.88006	13.94024	71.92341	8.480767

Медиана делит значения переменной на две равные части: 50% всех значений соответствующей переменной превосходят медиану, 50% - меньше медианы. Соответственно определяются квартили - 25% и 75% процентили (вместе с медианой они делят значения переменной на 4 части) В общем случае, процентиль - это значение, слева от которого находится заданная доля наблюдений. Например, 90% процентиль - это значение, слева от которого лежит 90% всех наблюдаемых значений переменной.

Процентили. Заметим, что процентили можно также вычислить в модуле *Основные статистики и таблицы.* Заданные пользователем процентили можно вычислить в модуле *Непараметрическая статистика* или используя язык *STATISTICA BASIC*.

Корреляции

Опции позволяют вычислить корреляции между выбранными переменными на каждом уровне группирующей переменной (если она имеется). Размер корреляционной матрицы ограничен только объемом памяти компьютера.

StatSoft®

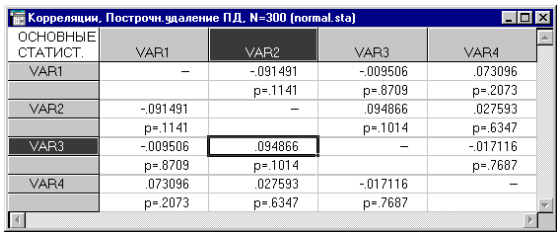
Copyright © StatSoft, 1995

Сохранить

Используйте эту опцию, чтобы вначале вычислить корреляционную матрицу (не показывая на экране), а затем сохранить ее. После нажатия кнопки будет открыто диалоговое окно *Сохранение файла*, в котором можно задать имя файла для сохранения матрицы. Если имеется группирующая переменная, то в одном файле будет вычислено и сохранено несколько корреляционных матриц (с корреляциями по группам) с возможностью дальнейшего использования в таких процедурах, как, например, *Моделирование структурными уравнениями*, см. том III.

p-уровни

Отметьте этот флажок, чтобы в таблице результатов вместе с коэффициентами корреляции отображались также их *p*-уровни.


Корреляции, Построчн.удаление ПД, N=300 (normal.sta)

ОСНОВНЫЕ СТАТИСТ.	VAR1	VAR2	VAR3	VAR4
VAR1	–	-.091491	-.009506	.073096
		p=.1141	p=.8709	p=.2073
VAR2	-.091491	–	.094866	.027593
	p=.1141		p=.1014	p=.6347
VAR3	-.009506	.094866	–	-.017116
	p=.8709	p=.1014		p=.7687
VAR4	.073096	.027593	-.017116	–
	p=.2073	p=.6347	p=.7687	

Заметим, что уровни значимости, вычисленные для отдельных коэффициентов, нужно интерпретировать с известной осторожностью (например, при *альфа=.05* можно ожидать, что чисто случайным образом корреляции между 5% пар переменных окажутся значимыми).

Построчное удаление ПД

В отличие от глобальной опции *Построчное удаление ПД*, эта опция применяется только при вычислении *корреляций*. Если она выбрана, то корреляции вычисляются только для наблюдений, не имеющих пропусков для всех выбранных переменных.

Если флажок не установлен (т.е. выбрано *попарное удаление пропущенных данных*), то попарные корреляции вычисляется для всех наблюдений, не имеющих пропусков для двух текущих переменных.

Распределение, нормальность

Эти опции позволяют вычислить последовательность таблиц частот для выбранных переменных, - одна таблица результатов для каждой переменной на каждом уровне группирующей переменной (если она выбрана). Можно выбрать метод группировки в разделе *Группировка* в диалоговом окне (см. ниже). Более детальная информация дана в *Таблицы частот - Обзор* (стр. 1582).

Таблицы частот

Нажав эту кнопку, вы откроете окно *Выбрать группу*. В этом окне можно выбрать для вычисления таблиц частот для всех групп или для определенной группы (одна таблица результатов для одной группы).

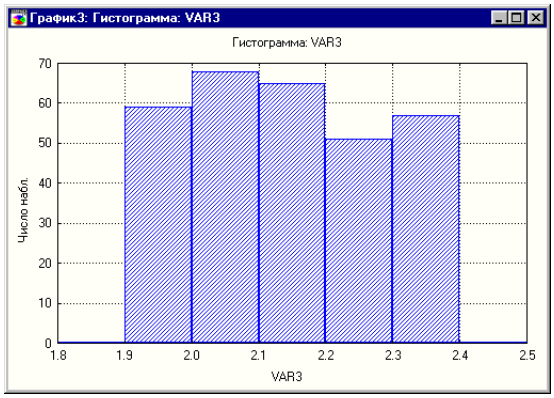
Распределение: VAR3 (normal.sta)

ОСНОВНЫЕ СТАТИСТ.	Частоты	Кумул. Частоты	% непро-пущен.	% выбран.
1.9000 < x <= 2.0000	59	59	19.66667	19.66667
2.0000 < x <= 2.1000	68	127	22.66667	22.66667
2.1000 < x <= 2.2000	65	192	21.66667	21.66667
2.2000 < x <= 2.3000	51	243	17.00000	17.00000
2.3000 < x <= 2.4000	57	300	19.00000	19.00000
Пропущ.	0	300		0.00000
Не выбрано	0	300		

Гистограммы

Нажав эту кнопку, вы откроете окно *Выбрать группу,* в котором можно выбрать построение гистограмм для всех групп или только для выбранных (по одной таблице результатов для каждой группы).

Copyright © StatSoft, 1995



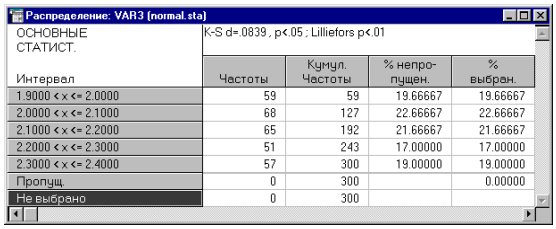
Другие типы гистограмм доступны с помощью опций *Статистические графики* [доступной из выпадающего меню *Графика* и из опции *Галерея графиков* (кнопка)].

Ожидаемые нормальные частоты

Если этот флажок установлен, для каждой группы в таблице частот ожидаемые частоты будут рассчитаны на основе нормального распределения.

Критерии К-С и Лиллиефорса

Опция вычисляет одновыборочную *D*-статистику Колмогорова-Смирнова и вероятности Лиллиефорса (см. Lilliefors, 1976).


Распределение: VAR3 (normal.sta)

ОСНОВНЫЕ СТАТИСТ.

K-S d=.0839, p<.05; Lilliefors p<.01

Интервал	Частоты	Кумул. Частоты	% непро-пущен.	% выбран.
1.9000 < x <= 2.0000	59	59	19.66667	19.66667
2.0000 < x <= 2.1000	68	127	22.66667	22.66667
2.1000 < x <= 2.2000	65	192	21.66667	21.66667
2.2000 < x <= 2.3000	51	243	17.00000	17.00000
2.3000 < x <= 2.4000	57	300	19.00000	19.00000
Пропущ.	0	300		0.00000
Не выбрано	0	300		

Если статистика Колмогорова-Смирнова значима, то гипотеза о нормальности распределения переменной отвергается. Вычисляются два значения вероятности для каждого критерия. Первое табулировано Massey (1951) и относится к случаю, когда параметры распределения – среднее и стандартное отклонение - известны априори и не оцениваются. Однако обычно эти параметры оцениваются по наблюдениям. В этом случае проверяется сложная условная гипотеза (с неизвестными заранее параметрами распределения) и вычисляются вероятности *Лиллиефорса* (Lilliefors, 1967, см. также программу *Lillief.stb* на языке *STATISTICA BASIC*, которая выполняет серию экспериментов Монте-Карло и табулирует критические значения статистики Лиллиефорса).

Группировка

Эти опции применяются, если таблицы частот и гистограммы построены с помощью опций *Распределение, нормальность*. Опция имеет только два режима группировки (разбиения на группы) значений переменных для таблиц частот. Дополнительные опции, графики и статистики доступны с помощью опции *Таблицы частот* (см. стартовую панель модуля *Основные статистики и таблицы*, глава 9). Заметим, что можно также использовать опцию *Перекодировать* (из меню *Правка - Переменные*), которая позволяет сгруппировать данные (т.е. задать новые значения группирующей переменной) на основании определяемого пользователем критерия (практически неограниченной сложности).

Число интервалов

Если выбрана эта опция, то диапазон значений переменных при построении таблиц частот или гистограмм будет разбит на примерно равное заданному число интервалов. Опция полезна при построении таблиц для непрерывных переменных. Критерии нормальности будут доступны, только если эта опция выбрана. Заметим, что число групп иногда может

StatSoft®

Copyright © StatSoft, 1995

отличаться от заданного, т.к. *STATISTICA* строит “приближенные” интервалы - границы интервала и ширина округляются, имеют последнюю десятичную цифру 1, 2 или 5 (например, 10.5, 11.0, 11.5 и т.д.). Такие интервалы легче интерпретировать, чем интервалы, содержащие много десятичных знаков (например, 10.12423, 10.13533 и т.д.). Полный контроль над методом группировки переменных доступен с помощью процедуры *Таблицы частот* (см. главу 9).

Целые интервалы

Выберите этот метод группировки, если значения табулируемой переменной можно рассматривать как целые номера категорий, или если она принимает только целые значения. Если выбран этот метод, все нецелые значения будут проигнорированы, если *Таблицы частот* или *Гистограммы* строятся с помощью опции *Распределение, нормальность*. Выбор в разделе *Группировка* не влияет на ход других вычислений (например, на *Описательные статистики*).

Заметим, что можно также использовать опцию *Перекодировать* (из меню *Правка - Переменные*), которая позволяет сгруппировать данные (т.е. задать новые значения группирующей переменной) на основании определяемого пользователем критерия (практически неограниченной сложности).

Графические опции

Диаграмма размаха для всех переменных

Опция строит диаграмму размаха для выбранных переменных, разбитых на группы (если выбрана группирующая переменная, по одной на каждую группу).

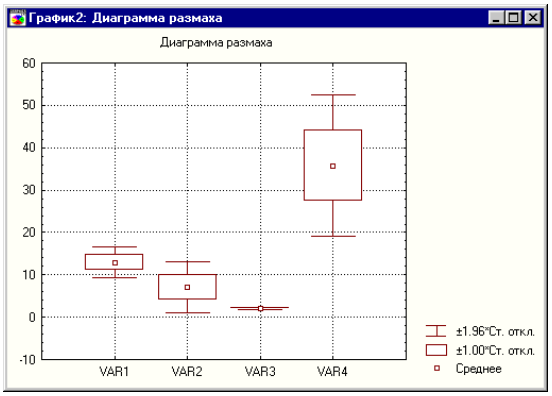


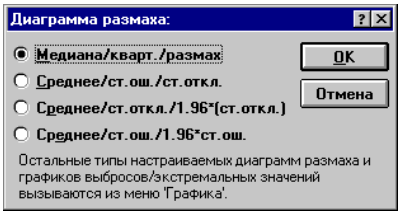
График имеет три компоненты:

(1) центральную точку, показывающую главную тенденцию или положение (медиана на графике вверху);

(2) прямоугольник, показывающий разброс значений относительно главной тенденции (25% -75% процентили);

(3) отрезки вокруг прямоугольника, показывающие диапазон значений переменной (Мин-макс).

После нажатия кнопки откроется диалоговое окно, где можно выбрать график для каждой переменной:



Здесь можно выбрать:

(1) медианы (центральная точка), квартили (прямоугольник), диапазоны (отрезки);

(2) средние, стандартные ошибки средних, стандартные отклонения;

(3) средние, стандартные отклонения, стандартные отклонения, умноженные на

Copyright © StatSoft, 1995
StatSoft

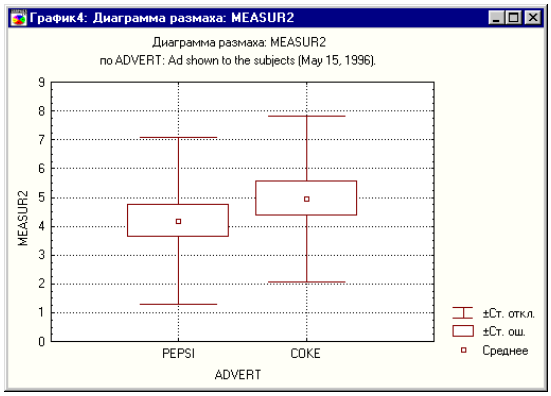
1.96 (95% доверительный интервал для среднего нормального распределения);

(4) средние, стандартные ошибки среднего, стандартные ошибки среднего, умноженные на 1.96.

Многие другие типы этих графиков доступны с помощью опции *Статистические графики* [из выпадающего меню *Графика* и из *Галереи графиков* (кнопка)].

Диаграммы размаха по группам

Опция строит диаграммы размаха для каждой выбранной переменной, разбитой на группы значениями другой, предварительно выбранной, (группирующей) переменной. (Если группирующая переменная не выбрана ранее, система предложит сделать выбор, после нажатия этой кнопки.)



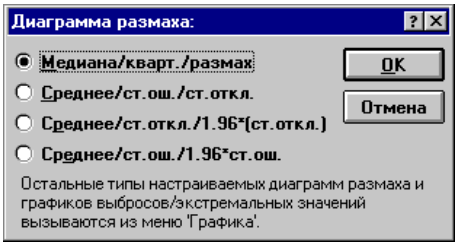
Этот график включает три компоненты (для каждой группы):

(1) точку, показывающую центр (медиана на графике вверху);

(2) прямоугольник, показывающий разброс значений относительно главной тенденции (25% -75% процентили);

(3) отрезки вокруг прямоугольника, показывающие диапазон значений переменной (Мин-макс).

После нажатия кнопки откроется диалоговое окно, в котором можно выбрать параметры графика для каждой переменной:



Здесь можно выбрать:

(1) медианы (центральная точка), квартили (прямоугольник), диапазоны (отрезки);

(2) средние, стандартные ошибки средних, стандартные отклонения;

(3) средние, стандартные отклонения, стандартные отклонения, умноженные на 1.96 (95% доверительный интервал для среднего нормального распределения);

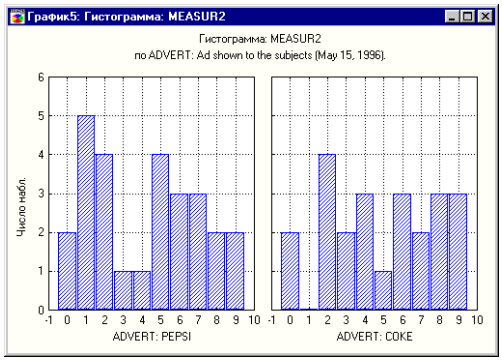
(4) средние, стандартные ошибки среднего, стандартные ошибки среднего, умноженные на 1.96.

Категоризованные гистограммы

Опция строит последовательность категоризованных гистограмм для выбранных переменных, один график для одной переменной. Для каждой переменной система *STATISTICA* построит гистограммы на уровнях группирующей переменной (например, отдельно для мужчин и женщин).

StatSoft

Copyright © StatSoft, 1995



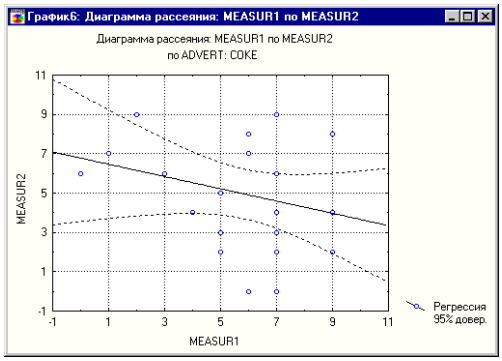
Если группирующая переменная не выбрана, то система предложит сделать выбор, когда нажмете эту кнопку.

Другие типы этих графиков доступны с помощью опции *Статистические графики* [из выпадающего меню *Графика* и из *Галереи графиков* (кнопка)].

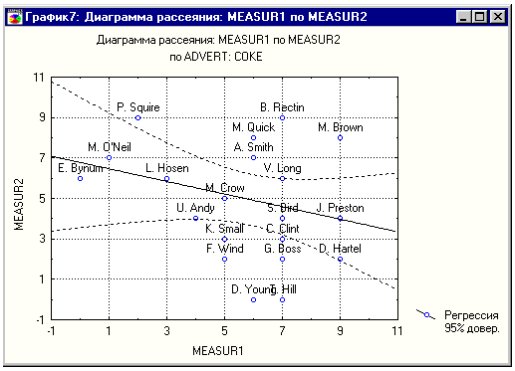
2M диаграмма рассеяния (с именами)

Опция строит последовательность диаграмм рассеяния для выбранных пар переменных на каждом уровне группирующей переменной. (Если группирующая переменная не выбрана, система предложит сделать выбор, когда нажмете эту кнопку.)

Если выбрана опция *2M рассеяния*, то на графике будет построена регрессионная прямая и 95% доверительные границы для нее.



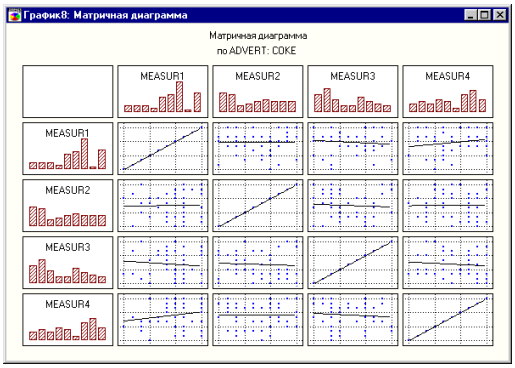
Если использована кнопка *с именами*, на диаграмме рассеяния точки будут снабжены именами или номерами наблюдений (если имена не определены в файле данных).



Другие типы графиков доступны с помощью опции *Статистические графики* [из выпадающего меню *Графика* и из *Галереи графиков* (кнопка)].

Матричный

Опция строит матрицу диаграмм рассеяния для выбранных переменных, одна матрица диаграмм рассеяния для каждого уровня группирующей переменной.



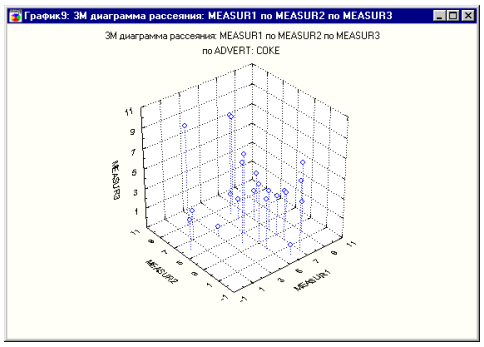
Многие другие типы этих графиков доступны с помощью опции *Статистические графики* [из выпадающего меню *Графика* и из *Галереи графиков* (кнопка)].

Copyright © StatSoft, 1995

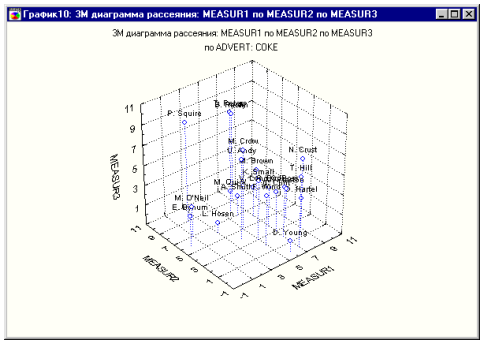
StatSoft®

3М диаграмма рассеяния (с именами)

Опция строит 3М диаграмму рассеяния для выбранной тройки переменных на каждом уровне группирующей переменной.



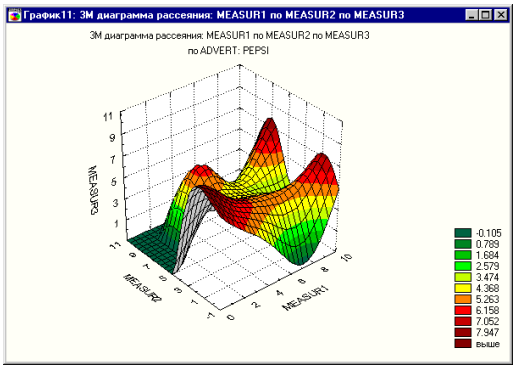
(Если группирующая переменная не выбрана, то система предложит ее выбрать.) Если использована кнопка *с именами*, точки диаграммы рассеяния будут отмечены именами или номерами наблюдений (если имена наблюдений определены в файле данных).



Другие типы этих графиков доступны с помощью опции *Статистические графики* [из выпадающего меню *Графика* и из *Галереи графиков* (кнопка)].

Поверхность

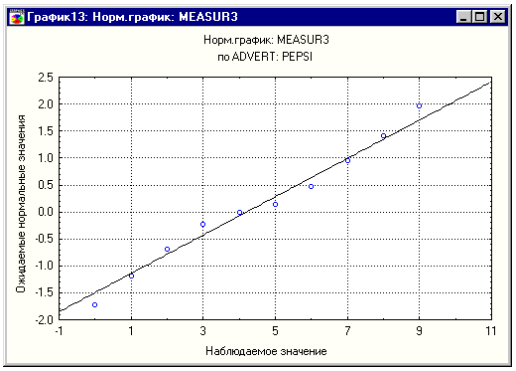
Опция строит: 3М диаграмму рассеяния для выбранной тройки переменных на каждом уровне группирующей переменной, а также поверхность по методу наименьших квадратов.



Другие типы этих графиков доступны с помощью опции *Статистические графики* [из выпадающего меню *Графика* и из *Галереи графиков* (кнопка)].

Нормальные вероятностные графики

Опция строит последовательность нормальных вероятностных графиков для выбранных сгруппированных переменных (если выбрана группирующая переменная).

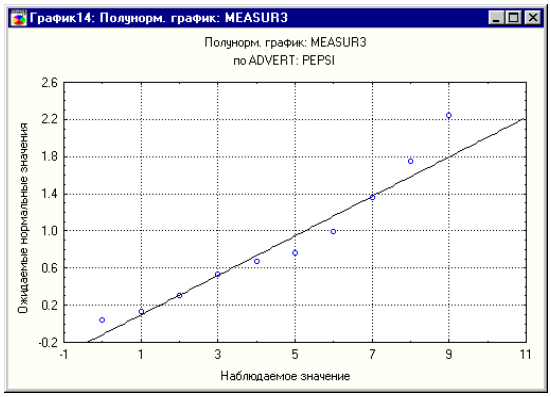


См. также опцию *Статистические графики* [из выпадающего меню *Графика* и из *Галереи графиков* (кнопка)].

StatSoft
Copyright © StatSoft, 1995

Полунормальные вероятностные графики

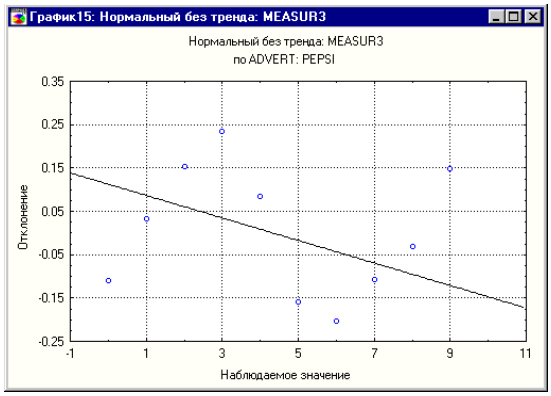
Опция строит последовательность полунормальных вероятностных графиков для выбранных сгруппированных переменных (если выбрана группирующая переменная).



См. также опцию *Статистические графики* [из выпадающего меню *Графика* и из *Галереи графиков* (кнопка)].

Нормальные вероятностные графики без тренда

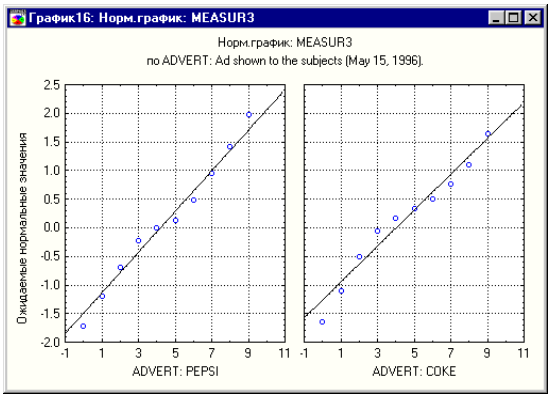
Опция строит последовательность нормальных вероятностных графиков без тренда для выбранных сгруппированных переменных.



См. также опцию *Статистические графики* [из выпадающего меню *Графика* и из *Галереи графиков* (кнопка)].

Категоризованные нормальные вероятностные графики

Опция строит последовательность нормальных вероятностных графиков для выбранных переменных на каждом уровне группирующей переменной.



Если группирующая переменная не выбрана, система предложит сделать выбор. См. также опцию *Статистические графики* [из выпадающего меню *Графика* и из *Галереи графиков* (кнопка)].

Категоризованная диаграмма рассеяния

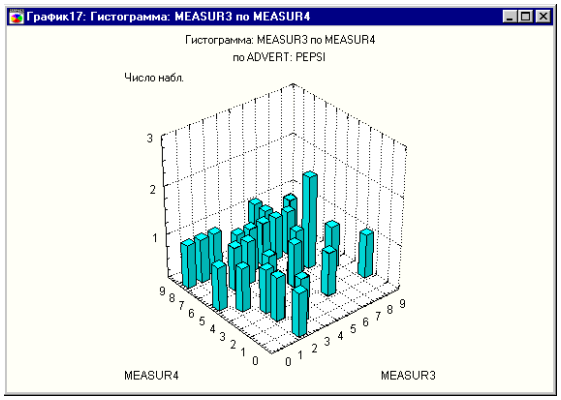
Опция строит последовательность категоризованных диаграмм рассеяния для выбранных пар переменных. После нажатия кнопки пользователю предлагается выбрать два списка переменных (из списка зависимых переменных). Диаграммы рассеяния будут построены для каждой переменной в первом списке и каждой переменной во втором списке, категоризованные уровнями группирующей переменной.

Copyright © StatSoft, 1995

Если используется кнопка *с именами*, точки на диаграмме рассеяния будут отмечены соответствующими номерами или именами наблюдений (если имена имеются).

3М гистограмма

Опция строит последовательность 3М гистограмм для выбранных пар переменных в каждой группе.

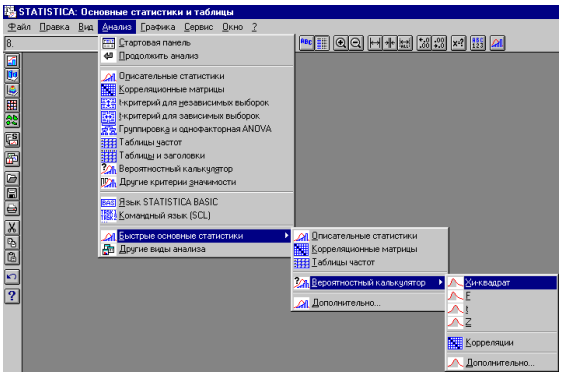


После нажатия кнопки, прежде всего, нужно выбрать группы, для которых будут строиться гистограммы, а затем два списка переменных. 3М гистограмма строится для каждой переменной из первого списка и для каждой переменной из второго списка на выбранных уровнях группирующей переменной.

Чтобы построить гистограммы, нормальные вероятностные графики или диаграммы размаха с использованием наблюдений (а не переменных), используйте опцию *Блочные статистические графики,* доступную из контекстного меню, выпадающего меню *Графика* или из *Галереи графиков* (вместо того, чтобы транспонировать данные; см. глава 3, *Окно таблицы исходных данных*).

ВЕРОЯТНОСТНЫЙ КАЛЬКУЛЯТОР

Если выбрана опция *Быстрые основные статистики - Вероятностный калькулятор* (из меню *Анализ*, из контекстного меню правой кнопки мыши или из меню панели инструментов), можно открыть *Калькулятор вероятностных распределений* (выбрать распределения *Chi*^{2}, *F*, *Z*, или *Дополнительно...* из расширенного меню) или окно *Распределение коэффициента корреляции Пирсона* (выбрать *Корреляции* из расширенного меню).



Калькулятор вероятностных распределений

Вводный обзор

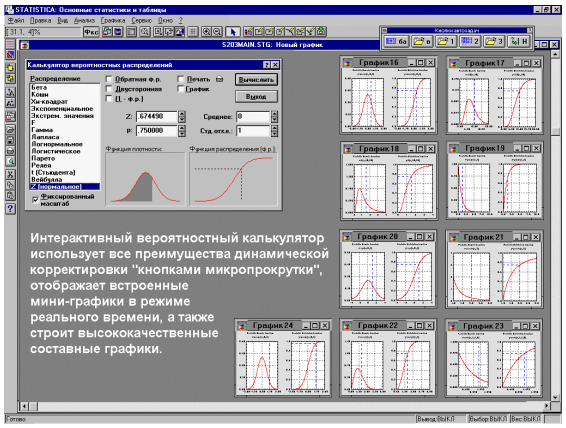
Калькулятор вероятностных распределений или *Вероятностный калькулятор* позволяет вычислить для различных распределений (например, для распределения Фишера *F* с параметрами *df 1*, *df 2*):

(1) верхние или нижние квантили для заданного значения (для симметричных распределений, например, для нормального, также можно

StatSoft
Copyright © StatSoft, 1995

вычислить двусторонние значения вероятности),

(2) критические значения для заданного пользователем значения вероятности.



Одной из полезных функций *Калькулятора вероятностных распределений* является вычисление критических значений для критериев значимости, а также вычисление *p*-уровней для различных статистик, имеющих определенное распределение (обычно это распределение Фишера *F*, распределение Стьюдента *t* или нормальное распределение *z*). В главе *Элементарные понятия статистики* (глава 8) описана логика статистических критериев и понятия уровень значимости.

Более формально действия *Вероятностного калькулятора* можно описать следующим образом. С каждым распределением случайной величины со значениями на действительной прямой связана единственная функция распределения. Это возрастающая функция, которая точке *x* действительной прямой ставит в соответствие точку *p* отрезка [0, 1] - вероятность того, что случайная величина примет значение меньше *x*. *Вероятностный калькулятор* позволяет: по любому *x* вычислить *p*, по любому *p* из отрезка [0,1] вычислить *x*. Можно провести вычисления для обратной функции распределения, для двух хвостов (в случае симметричных распределений), а также использовать *1 - p* вместо *p* (т.е. рассматривать левый хвост функции распределения). Доступные распределения описаны в следующем разделе.

Распределения

В этом диалоговом окне доступны следующие распределения. Относительно распределений более подробно см. *Приложение I*.

- *Бета*
- Коши
- *Хи-квадрат*
- Экспоненциальное
- Экстремального значения
- *F*
- *Гамма*
- Лапласа
- Логистическое
- Логнормальное
- Парето
- Релея
- *t* (Стьюдента)
- Вейбулла
- *Z* (Нормальное)

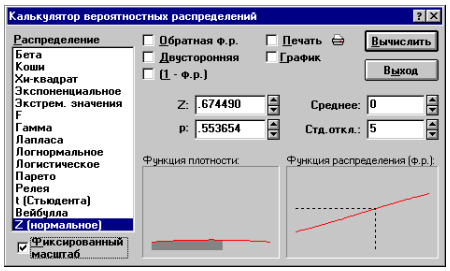
Фиксированный масштаб

Опция позволяет изменить масштаб графиков (показываемых в окне и выводимых с помощью опции *График*). Если опция выбрана, программа использует фиксированный масштаб для графиков и при изменении параметров распределения, можно увидеть изменения на графике в том же самом масштабе. Заметим, что при некоторых значениях параметров результаты могут выйти за пределы графика. Чтобы избежать этого, отмените опцию

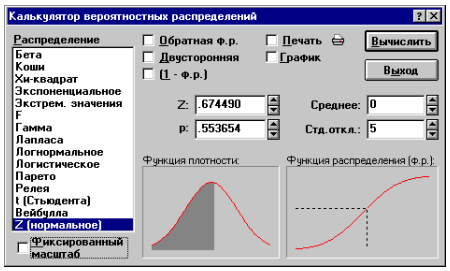
Copyright © StatSoft, 1995
StatSoft

Фиксированный масштаб. После этого программа выберет оптимальный масштаб.

Фиксированный масштаб: ВКЛЮЧИТЬ



Фиксированный масштаб: ВЫКЛЮЧИТЬ



Иногда для некоторых распределений форма не меняется при изменении параметров. Это происходит, например, если изменить *Стд.откл.* для нормального (*z*) распределения, при выключенной опции *Фиксированный масштаб*.

Обратная функция распределения

Эта опция позволяет провести вычисления для обратной функции распределения (ф.р.). Таким образом, если ввести уровень значимости, то можно вычислить критическое значение соответствующей статистики. Опция будет выбрана автоматически, если изменить значение *p*.

Двусторонняя

Опция выбирает двусторонний критерий (т.е. рассматриваются оба хвоста функции распределения критериальной статистики). Доступна только для симметричных распределений.

1 – функция распределения (1 - ф.р.)

Опция вычисляет критическое значение выбранного распределения по величине (1-ф.р.) или, наоборот, вычисляет (1-ф.р.) по критическому значению (т.е. рассматривается левый хвост функции распределения критериальной статистики).

Значение *x*

В этом поле отображается значение аргумента *x* выбранной функции распределения. После ввода значения *x* (вручную или с использованием нажатием *стрелок*), *STATISTICA* вычислит соответствующее значение вероятности *p*. Дополнительная информация о распределениях и их свойствах дана в *Приложении I*.

p-значение

Для выбранного распределения вычисляется *p*-значение или, наоборот, вводится *p*-значение и вычисляется критическое значение *x*.

Параметры распределения

В этом поле ввода программа отображает параметры текущего распределения. Если параметры (число степеней свободы, среднее, стандартное отклонение и т.п.) изменились, программа пересчитывает *p*-значение. Дополнительная информация о распределениях дана в *Приложении I*.

Печать

Используйте эту опцию, чтобы распечатать (направив на принтер, в файл и/или в *Окно вывода*) *p*-значение и значение случайной величины. Эти значения будут автоматически печататься, если нажать кнопку *Вычислить*. Можно редактировать параметры, используя стрелки и просматривать результаты в

StatSoft
Copyright © StatSoft, 1995

диалоговом окне. После выбора параметров, нажмите кнопку *Вычислить* для печати результатов. Чтобы печатать графики, созданные с помощью опции *График*, можно использовать опцию *Печать графика* или кнопку на панели инструментов. Если ранее выбрана опция *Автоматически печатать все графики*, то все создаваемые графики будут автоматически печататься.

График

Опция позволяет создать составной график плотности и функции распределения (см. том II о составных графиках).

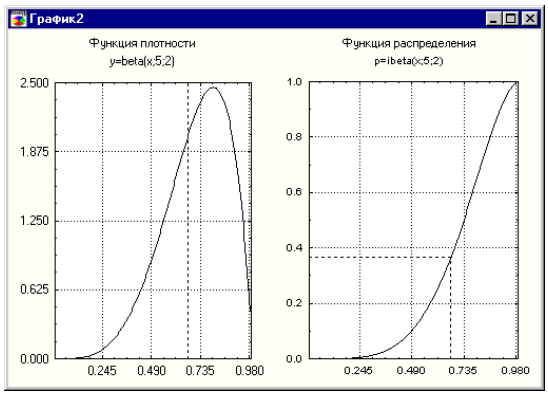


График создается, после нажатия кнопки *Вычислить*. Поэтому, перед тем как создать график, можно изменять параметры распределения и вычисляемые характеристики, используя *стрелки,* и просматривать результаты в диалоговом окне. Чтобы построить график, нажмите кнопку *Вычислить*. Для того чтобы распечатать графики, когда выбрана опция *График*, можно использовать либо опцию *Печать графика*, либо кнопку на панели инструментов. Если ранее выбрана опция *Автоматически печатать все графики*, то создаваемые графики будут печататься автоматически.

Вычислить

Нажатие этой кнопки вводит текущее редактируемое значение (если не используются *стрелки*), после чего программа вычисляет *p*-значение или значение переменной, имеющей заданное распределение. После этого будут выполнены опции:

- *Печать* значения переменной и соответствующего *p*-значения на принтере, в файл и/или в *Окно вывода* (в соответствии с установками в окне *Параметры страницы/вывода*);
- *Построить* составной график с функцией плотности и функцией распределения.

Распределение коэффициента корреляции Пирсона

Вводный обзор

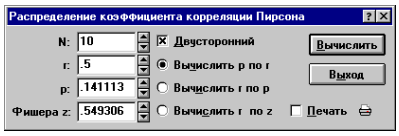
Наиболее широко используется коэффициент корреляции – *Пирсона r*, называемый также линейной корреляцией, т.к. он оценивает силу линейной зависимости между переменными. Это основной тип корреляции в *Быстрых основных статистиках* (в каждом модуле) и в модуле *Основные статистики и таблицы* (см. глава 9, где дан обзор различных корреляций). Другими словами, можно сказать, что коэффициент корреляции определяет степень “пропорциональности” изменения двух величин. Значение коэффициента корреляции не зависит от масштаба, в котором выполнены измерения. Например, корреляция между ростом и весом будет одной и той же, независимо от единиц измерения: *дюймы и фунты* или *сантиметры и килограммы*. *Пропорциональность* означает *линейную зависимость*. Корреляция высокая, если зависимость можно аппроксимировать прямой линией. Эта прямая называется *регрессионной прямой* или *прямой наименьших*

Copyright © StatSoft, 1995

квадратов, потому что она построена так, что сумма *квадратов* расстояний от нее до наблюдаемых точек минимальна. Корреляция Пирсона вычисляется по формуле:

$$r_{12} = \Sigma(Y_{i1}-\text{Y-bar}_1)*(Y_{i2}-\text{Y-bar}_2)/ [\Sigma(Y_{i1}-\text{Y-bar}_1)^2*\Sigma(Y_{i2}-\text{Y-bar}_2)^2]^{1/2}$$

Диалоговое окно *Распределение коэффициента корреляции Пирсона* позволяет вычислить *p*-значение для данного коэффициента корреляции. Можно сделать, наоборот: ввести *p* и вычислить коэффициент корреляции (для выборки определенного объема N). Критерий значимости основан на предположении, что остатки зависимой переменной *y* имеют нормальное распределение с постоянной дисперсией.



Опции этого диалогового окна обсуждаются ниже.

Параметры

Имеются следующие параметры:

- *N* (объем выборки),
- *r* (коэффициент корреляции),
- *p* (уровень значимости),
- *Z* Фишера (критическое значение распределения).

Эти параметры можно либо задать, либо вычислить (см. ниже).

Двусторонний

Опция позволяет выбрать двусторонний критерий (по умолчанию используется односторонний).

Вычислить *p* по *r*

Опция вычисляет *p*-уровень (уровень значимости) и *Z*- значение Фишера по значению *r* и объему выборки *N*.

Вычислить *r* по *p*

Опция вычисляет *r* и *Z* значение по *p*-уровню и объему выборки *N*.

Вычислить *r* по z

Опция вычисляет *r* и *p*-уровень по *Z*-значению и объему выборки *N*.

Печать

Используйте опцию, чтобы отпечатать (направить на принтер, в файл и/или в *Окно вывода*) *p*-значение и критическое значение. Это будет сделано, когда вы нажмете кнопку *Вычислить*. Можно редактировать параметры, используя инструмент стрелки и просматривать результаты в диалоговом окне. Когда будете готовы, нажмите кнопку *Вычислить*, чтобы печатать результаты.

Вычислить

Нажмите кнопку, чтобы выполнить вычисления и *Печатать* результаты (если выбрана опция *Печать*, см. выше).

StatSoft®

Copyright © StatSoft, 1995

ПРЕДМЕТНЫЙ УКАЗАТЕЛЬ

Б

Бета распределение, 1595
Быстрые основные статистики, 1576
  3М гистограмма, 1594
  вероятностные графики, 1592
  выбор переменных, 1576
  гистограмма, 1587, 1590, 1594
  глобальные опции, 1586
  график поверхности, 1592
  диаграмма размаха, 1589
  диаграммы рассеяния, 1591, 1593
  длинные имена переменных, 1586
  квартили, 1586
  корреляции, 1586
  корреляции по..., 1581
  корреляционные матрицы, 1580
  К-С критерий, 1588
  Лиллиефорса критерий, 1588
  медианы, 1586
  ожидаемые нормальные частоты, 1588
  описательные по..., 1579
  описательные статистики, 1577, 1586
  опции группировки, 1588
  попарное удаление пропущенных данных, 1586, 1587
  построчное удаление пропущенных данных, 1586, 1587
  распределение, нормальность, 1587
  расширенные возможности, 1585
  таблицы частот, 1582, 1587
  целые интервалы, 1589
  частоты по…, 1584
  число интервалов, 1588
Быстрые основные статистики - расширенные возможности, 1585

В

Вейбулла распределение, 1595

Г

Гамма распределение, 1595
Группировка
  целые интервалы, 1589
  число интервалов, 1588

К

Калькулятор вероятностных распределений
  р-значение, 1596
  график, 1597
  двусторонний тест, 1596
  значение *x*, 1596
  критическое значение, 1596
  обзор, 1594
  обратное распределение, 1596
  параметры, 1596
  печать, 1596
  распределения, 1595
  фиксированный масштаб, 1595
Корреляции
  Пирсона, 1597
Корреляционные матрицы, 1580, 1581, 1586
Коши распределение, 1595
К-С критерий, 1588

Л

Лапласа распределение, 1595
Лиллиефорса критерий, 1588
Логистическое распределение, 1595
Логнормальное распределение, 1595

Н

Нормальное распределение, 1595

О

Описательные по..., 1579
Описательные статистики, 1577, 1586

П

Парето распределение, 1595
Попарное удаление пропущенных данных, 1587
Построчное удаление пропущенных данных, 1586, 1587

Р

Распределение коэффициента корреляции Пирсона
  обзор, 1597
Распределение экстремального значения, 1595
Распределения
  бета, 1595
  Вейбулла, 1595
  гамма, 1595
  Коши, 1595
  Лапласа, 1595
  логистическое, 1595
  логнормальное, 1595
  нормальное, 1595
  Парето, 1595
  Релея, 1595
  Стьюдента t, 1595
  Фишера F, 1595
  хи-квадрат, 1595
  экспоненциальное, 1595
  экстремального значения, 1595
Релея распределение, 1595

С

Стьюдента t распределение, 1595

Т

Таблицы частот, 1582
Таблицы частот по..., 1584

Ф

Фишера F распределение, 1595

Х

Хи-квадрат распределение, 1595

Э

Экспоненциальное распределение, 1595

Copyright © StatSoft, 1995

StatSoft®

StatSoft®

Copyright © StatSoft, 1995

Глава 11:

НЕПАРАМЕТРИЧЕСКАЯ СТАТИСТИКА И ПОДГОНКА РАСПРЕДЕЛЕНИЯ

Оглавление

Вводный обзор ... 1605
Обзор программы ... 1610
Диалоговые окна, опции, статистики ... 1613
  Стартовые панели ... 1613
  Таблицы частот 2 x 2 ... 1623
  Наблюдаемые частоты в сравнении с ожидаемыми ... 1624
  Коэффициенты корреляции ... 1624
  Сравнение двух групп ... 1626
  Краскела-Уоллиса ANOVA и медианный тест ... 1627
  Сравнение двух переменных ... 1628
  Ранговый дисперсионный анализ Фридмана и конкордация Кендалла ... 1629
  Кохрена *Q* критерий ... 1630
  Описательные статистики ... 1630
  Подгонка распределения ... 1631
Примеры ... 1634
Технические замечания ... 1650
Предметный указатель ... 1651

Подробное оглавление

ВВОДНЫЙ ОБЗОР ... 1605
  Основная цель ... 1605
  Краткий обзор непараметрических процедур ... 1606
  Какой метод использовать ... 1607
  Подгонка распределения ... 1608
    Подгонка распределения с помощью *2M графиков* и модуля *Анализ процессов* ... 1609
  Структура главы ... 1609
ОБЗОР ПРОГРАММЫ ... 1610
  Непараметрическая статистика ... 1610
  Подгонка распределений ... 1610
  Альтернативные процедуры ... 1611

Copyright © StatSoft, 1995

StatSoft®

ДИАЛОГОВЫЕ ОКНА, ОПЦИИ, СТАТИСТИКИ 1613
Стартовые панели 1613
Стартовая панель - Непараметрические статистики 1613
Таблицы частот 2 на 2: статистики *Хи/V/Фи-квадрат*, Макнемара, точный критерий Фишера 1613
Наблюдаемые частоты в сравнении с ожидаемыми 1613
Корреляции (Спирмена R, *тау* Кендалла, *Гамма*) 1613
Критерий серий Вальда-Вольфовица 1614
U критерий Манна-Уитни 1614
Двухвыборочный критерий Колмогорова-Смирнова 1615
ANOVA Краскела-Уоллиса и медианный тест 1615
Критерий знаков 1616
Критерий Вилкоксона 1616
ANOVA Фридмана и коэффициент конкордации Кендалла 1617
Q критерий Кохрена 1617
Стартовая панель – Описательные статистики 1618
Медиана 1618
Мода 1618
Геометрическое среднее 1618
Гармоническое среднее 1619
Дисперсия и стандартное отклонение 1619
Среднее отклонение 1619
Размах 1619
Квартильный размах 1619
Асимметрия 1619
Эксцесс 1620
Стартовая панель – Подгонка распределения 1620
Общее 1620
Нормальное распределение 1620
Равномерное распределение 1621
Экспоненциальное распределение 1621
Гамма распределение 1621
Логнормальное распределение 1621
Хи-квадрат распределение 1622
Биномиальное распределение 1622
Распределение Пуассона 1622
Геометрическое распределение 1622
Распределение Бернулли 1622
Таблицы частот 2 x 2 1623
Наблюдаемые частоты в сравнении с ожидаемыми 1624
Коэффициенты корреляции 1624
Переменные 1624
Корреляция 1624
Вычислить 1625
Матричная диаграмма 1625
Сравнение двух групп 1626

StatSoft®

Copyright © StatSoft, 1995

Переменные ... 1627
Диаграмма размаха ... 1627
Категоризованная гистограмма ... 1627
ANOVA Краскела-Уоллиса и медианный тест ... 1627
Переменные ... 1628
Коды ... 1628
Пакетная обработка/ печать ... 1628
Диаграмма размаха ... 1628
Категоризованная гистограмма ... 1628
Сравнение двух переменных ... 1628
Переменные ... 1629
Диаграмма размаха ... 1629
Ранговый дисперсионный анализ Фридмана и конкордация Кендалла ... 1629
Переменные ... 1629
Диаграмма размаха ... 1629
Кохрена *Q* критерий ... 1630
Переменные ... 1630
Коды ... 1630
Описательные статистики ... 1630
Переменные ... 1630
Вычислить процентили ... 1630
Вычисление описательных статистик ... 1631
Подгонка распределения ... 1631
Переменная ... 1632
Распределение ... 1632
Число групп ... 1632
Критерий согласия *хи-квадрат* ... 1632
Одновыборочный критерий согласия Колмогорова-Смирнова ... 1632
График ... 1633

ПРИМЕРЫ ... 1634
Пример 1: Таблица 2x2 ... 1634
Пример 2: Критерий знаков ... 1635
Пример 3: Наблюдаемые значения в сравнении с ожидаемыми: *хи-квадрат* ... 1637
Пример 4: Корреляции ... 1637
Пример 5: Критерий серий Вальда-Вольфовица, Манна-Уитни *U* критерий, двухвыборочный критерий Колмогорова-Смирнова ... 1639
Пример 6: Ранговый дисперсионный анализ Краскела-Уоллиса и медианный тест ... 1640
Пример 7: Дисперсионный анализ Фридмана и конкордация Кендалла ... 1643
Пример 8: Критерий Вилкоксона парных сравнений ... 1644
Пример 9: Q критерий Кохрена ... 1645
Пример 10: Описательные статистики ... 1646
Пример 11: Подгонка распределения ... 1647

ТЕХНИЧЕСКИЕ ЗАМЕЧАНИЯ ... 1650
Совпадающие ранги ... 1650

Copyright © StatSoft, 1995

Критерии значимости для выборок малого объема..... 1650
Подгонка распределения Вейбулла и Гомпертца для цензурированных данных 1650
Подгонка распределения с помощью *2M графиков* и в модуле *Анализ процессов* 1650

ПРЕДМЕТНЫЙ УКАЗАТЕЛЬ.....1651

StatSoft®

Copyright © StatSoft, 1995

Глава 11:

НЕПАРАМЕТРИЧЕСКАЯ СТАТИСТИКА И ПОДГОНКА РАСПРЕДЕЛЕНИЯ

ВВОДНЫЙ ОБЗОР

Основная цель

Краткий обзор понятия "критерий значимости". Для того чтобы понять идею *непараметрической* статистики, следует понять идею *параметрической* статистики. Глава *Элементарные понятия статистики* (глава 8) знакомит с понятием статистической значимости критерия, основанного на выборочном распределении определенной статистики (просмотрите эту главу, прежде чем продолжить чтение). Говоря кратко, если вы знаете исходное распределение наблюдаемой переменной, то можете предсказать, как в повторных выборках равного объема будет "вести себя" используемая статистика - т.е. каким образом она будет распределена.

Пусть, например, имеется 100 случайных выборок, из одной популяции по 100 взрослых человек в каждой. Вычислим средний рост субъектов в каждой выборке, т.е. построим выборочное среднее. Тогда распределение выборочных средних можно хорошо аппроксимировать нормальным распределением (более точно, *t* распределением Стьюдента с 99 степенями свободы, см. ниже). Теперь представьте, что случайным образом извлечена еще одна выборка из числа жителей некоего города ("Вышгород"), где, по вашим представлениям, проживают люди с ростом выше среднего. Если средний рост людей в этой выборке попадает в верхнюю 95% критическую область *t* распределения, то можно сделать обоснованный вывод, что жители Вышгорода действительно, в среднем более высокие (чем в целом в популяции).

Действительно ли большинство переменных имеют нормальное распределение? В рассмотренном примере использовался тот факт, что в повторных выборках равного объемы средние значения (роста людей) будут иметь *t* распределение (с определенным средним и дисперсией). Однако это верно лишь, когда рассматриваемая переменная (рост) имеет нормальное распределение. Для многих изучаемых переменных невозможно сказать с уверенностью, что это действительно так. Например, является ли доход нормально распределенной величиной? - скорее всего. Случаи редких болезней не являются нормально распределенными в популяции, число автомобильных аварий также не является нормально распределенным, как и многие переменные, интересующие исследователя.

Объем выборки. Другим фактором, часто ограничивающим применимость критерия, основанного на предположении нормальности, является объем или размер выборки, доступной для анализа (*размер выборки*; *n*). До тех пор пока выборка достаточно большая (например, 100 или больше наблюдений), можно считать, что выборочное распределение нормально, даже если вы не уверены, что распределение переменной в популяции является нормальным. Тем не менее, если выборка очень мала, то эти критерии (основанные на предположении нормальности) следует использовать только при наличии уверенности, что переменная действительно имеет нормальное распределение. Однако нет способа проверить это предположение на малой выборке.

Проблемы измерения. Использование критериев, основанных на предположении

Copyright © StatSoft, 1995

StatSoft®

нормальности, кроме того, ограничено точностью измерений. Например, рассмотрим исследование, в котором средний балл успеваемости (*GPA*) является основной переменной. Можно ли сказать, что средняя успеваемость студента A в два раза выше, чем успеваемость студента C? Является ли различие между средним баллом студентов B и A сравнимым с различием между студентами D и C? GPA является грубой мерой, позволяющей только ранжировать студентов в порядке "хороший" - "плохой". Эта общая задача измерений обычно обсуждается в учебниках по статистике в терминах *типов измерений* или *шкалы измерения* (см. также *Элементарные понятия статистики*, глава 8). Не вдаваясь в детали, отметим, что наиболее общие статистические методы, такие как дисперсионный анализ (*t*-критерий), регрессия и т.д. предполагают, что исходные измерения выполнены, по крайней мере, в *интервальной* шкале, в которой интервалы можно разумным образом сравнивать между собой (например, *B* минус *A* равняется *D* минус *C*). Тем не менее, как в данном примере, такие предположения часто неестественны, и данные скорее просто *ранжированы* (измерены в *порядковой* шкале) или упорядочены, чем измерены точно.

Параметрические и непараметрические методы. Надеемся, что после этого введения становится ясной необходимость статистических процедур, позволяющих обрабатывать данные "низкого качества" из выборок малого объема с переменными, про распределение которых ничего не известно. Непараметрические методы как раз и разработаны для того, чтобы использовать их в ситуациях, когда исследователь ничего не знает о параметрах исследуемой популяции (отсюда и название *непараметрические*).

Говоря более специальным языком, непараметрические методы не основываются на оценке параметров (таких как среднее или стандартное отклонение) при описании выборочного распределения интересующей величины. Поэтому эти методы иногда также называют (более подходящим термином) *свободными от параметров* или *свободно распределенными* методами.

Краткий обзор непараметрических процедур

По существу, для каждого параметрического критерия имеется, по крайней мере, один непараметрический эквивалент. Эти тесты попадают в одну из следующих категорий:

- критерии различия между группами (независимые выборки);
- критерии различия между переменными (зависимые выборки);
- критерии зависимости между переменными.

Различия между независимыми группами. Обычно, когда имеются две выборки (например, мужчины и женщины), которые вы хотите сравнить относительно их среднего значения для некоторой изучаемой переменной, вы можете использовать *t*-тест для независимых выборок (*Основные статистики и таблицы*, глава 9); непараметрическими альтернативами этому тесту являются критерий *серий Вальда-Вольфовица*, *Манна-Уитни U* тест, *двухвыборочный критерий Колмогорова-Смирнова*. Если вы имеете несколько групп, то можете использовать дисперсионный анализ (см. *ANOVA/MANOVA*, глава 13). Непараметрическими эквивалентами этого метода являются ранговый критерий *Краскела-Уоллиса* и *медианный* тест.

Различия между зависимыми группами. Если вы хотите сравнить две переменные, относящиеся к одной и той же выборке (например, математические успехи студентов в начале и в конце семестра), то обычно

StatSoft®

Copyright © StatSoft, 1995

используется *t-критерий* для зависимых выборок (*Основные статистики и таблицы*, глава 9). Альтернативным непараметрическим тестом является *критерий знаков* и *критерий Вилкоксона*. Если рассматриваемые переменные по природе своей категориальны, то подходящим является *хи-квадрат Макнемара.* Если рассматривается более двух переменных, относящихся к одной и той же выборке, то обычно используется дисперсионный анализ (ANOVA) с повторными измерениями. Альтернативным непараметрическим методом является *Ранговый дисперсионный анализ Фридмана* и *Q критерий Кохрена* (последний, если переменная измерена в категориальных терминах). *Q* критерий Кохрена используется, например, для оценки изменений частот (долей).

Зависимости между переменными. Для того, чтобы оценить зависимость между двумя переменными, обычно вычисляют коэффициент корреляции. Непараметрическими эквивалентами стандартного коэффициента корреляции Пирсона являются статистики:*Спирмена R*, *Кендалла tau* и коэффициент *Гамма* (см. *Непараметрические корреляции*, стр. 1624). Если две рассматриваемые переменные по природе своей категориальны или являются категоризованными, то подходящими непараметрическими критериями для проверки наличия связей (зависимостей) между ними будут: *хи-квадрат, фи* коэффициент, *точный критерий Фишера* (см. *Стартовая панель - Непараметрическая статистика и распределения*, стр. 1613). Дополнительно в *STATISTICA* доступен критерий зависимости между несколькими переменными: *коэффициент конкордации Кендалла*. Этот тест часто используется для оценки согласованности мнений независимых экспертов (судей), в частности, баллов, выставленных одному и тому же субъекту.

Описательные статистики. Если данные не являются нормально распределенными, а измерения, в лучшем случае, содержат ранжированную информацию, то вычисление обычных описательных статистик (например, среднего, стандартного отклонение) иногда не слишком информативно. Например, в психометрии хорошо известно, что воспринимаемая интенсивность стимулов (например, воспринимаемая яркость света) представляет собой логарифмическую функцию реальной интенсивности (яркости, измеренной в объективных единицах - *люксах*).

В данном примере, обычная оценка среднего (сумма значений, деленная на число стимулов) не дает верного представления о среднем значении действительной интенсивности стимула. (В обсуждаемом примере скорее следует вычислить геометрическое среднее.) Модуль *Непараметрическая статистика* вычислит разнообразный набор мер положения (среднее, медиана, мода и т.д.) и рассеяния (дисперсия, гармоническое среднее, квартильный размах, т.д.) позволяющих представить более “полную картину” данных (см. *Описательные статистики*, стр. 1618).

Какой метод использовать

Нелегко дать простой совет, касающийся использования непараметрических процедур. Каждая непараметрическая процедура имеет свои достоинства и свои недостатки. Например, двухвыборочный критерий Колмогорова-Смирнова чувствителен не только к различию в положении двух распределений, например, к различиям средних, но также чувствителен и к форме распределения. Тест Вилкоксона парных сравнений предполагает, что можно ранжировать различия между сравниваемыми наблюдениями. Если это не так, лучше использовать критерий знаков. В общем, если результат исследования является важным (например, оказывает ли людям помощь определенная очень дорогая и болезненная

Copyright © StatSoft, 1995

StatSoft®

лекарственная терапия?), то всегда целесообразно применить различные непараметрические тесты. Возможно, результаты тестирования (разными тестами) будут различны. В таком случае следует попытаться понять, почему разные тесты дали разные результаты. С другой стороны, непараметрические тесты имеют меньшую статистическую мощность (менее чувствительны), чем их параметрические конкуренты, и если важно обнаружить даже слабые эффекты (например, является ли данная пищевая добавка опасной для людей?) следует особенно внимательно выбирать статистику критерия.

Большие массивы данных и непараметрические методы. Непараметрические методы наиболее приемлемы, когда объем выборок мал. Если данных много (например, $n > 100$) часто имеет мало смысла использовать непараметрическая статистика. Глава *Элементарные понятия статистики* (глава 8) содержит краткое обсуждение *центральной предельной теоремы*. Главное здесь состоит в том, что когда выборки становятся очень большими, то выборочные средние подчиняются нормальному закону, даже если исходная переменная не является нормальной или измерена с погрешностью. Таким образом, параметрические методы, являющиеся более чувствительными (имеющие большую статистическую мощность), всегда подходят для больших выборок. Большинство критериев значимости многих непараметрических статистик, описанных далее, основываются на асимптотической теории (большой выборки) поэтому соответствующие тесты часто не выполняются, если размер выборки становится слишком малым. Обратитесь к описаниям определенных критериев, чтобы узнать больше об их мощности и эффективности.

Подгонка распределения

В некоторых исследовательских проектах можно сформулировать гипотезы относительно распределения рассматриваемой переменной. Например, переменные, значения которых определяются бесконечным числом независимых факторов, распределены по нормальному закону: можно предположить, что рост индивидуума является результатом воздействия многих независимых факторов, таких как различные генетические предрасположенности, болезни, перенесенные в раннем возрасте и т.д. Как следствие, рост имеет тенденцию к нормальному распределению в популяции США. С другой стороны, если наблюдаемые значения переменной являются результатом очень редких событий, то переменная будет иметь распределение *Пуассона* (которое иногда называется распределением редких событий). Например, несчастные случаи на производстве можно рассматривать как результат пересечения ряда неудачных событий (на житейском языке стечением маловероятных обстоятельств), поэтому их частота приближенно описывается распределением Пуассона. Эти и другие полезные распределения подробно описываются в этой главе.

Модуль *Непараметрическая статистика* позволяет строить критерии согласия данных с соответствующим гипотетическим распределением. Вы можете также ввести ряд произвольных ожидаемых значений (т.е. “построить” гипотетическое распределение) и проверить согласие с наблюдаемым множеством значений (*наблюдаемые в сравнении с ожидаемыми* частотами).

Гипотеза нормальности. Другим обычным приложением процедуры подгонки распределения является проверка гипотезы нормальности до того, как использовать какой-либо параметрический тест (см. выше).

StatSoft®

Copyright © StatSoft, 1995

Подгонка распределения с помощью *2М графиков* и модуля *Анализ процессов*

Начиная с 5-й версии, *STATISTICA* включает несколько альтернативных процедур для подгонки непрерывных распределений к данным. Можно выбрать (категоризованные или не категоризованные) графики *Квантиль-квантиль*, *Вероятность-вероятность*, или *2М гистограммы* (с кривой, показывающей ожидаемые частоты) из меню *Графика* и затем выбрать тип распределения: *Нормальное*, *Бета*, *Экстремального значения*, *Гамма*, *Геометрическое*, *Лапласа*, *Логистическое*, *Логнормальное*, *Пуассона*, *Релея*, и *Вейбулла* распределения. Программа вычислит оценку максимального правдоподобия для параметров соответствующего распределения (см. том II)

Модуль *STATISTICA Анализ процессов* (см. том IV) также содержит несколько опций для подгонки этих распределений и вычисления таблиц с ожидаемыми и наблюдаемыми частотами. Кроме того, этот модуль также дает возможность подогнать одновременно все эти непрерывные распределения к данным. Таблицы результатов будут содержать соответствующие значения статистики Колмогорова-Смирнова *d-max*, расположенные в возрастающем порядке (иными словами, от наилучшей подгонки к наихудшей).

Структура главы

Эта глава руководства структурно отличается от глав, в которых описаны другие статистические процедуры. По существу, модуль *Непараметрическая статистика* сам является полным статистическим пакетом, содержащим много различных процедур, предназначенных для разного типа данных и решения различных исследовательских задач. В связи с этим, справочный раздел (начинающийся на стр. 1613), в котором каждый метод обсуждается подробно в сравнении с альтернативными методами, помещается первым. Далее даются примеры использования каждой статистической процедуры (начиная со стр. 1634).

Copyright © StatSoft, 1995

StatSoft®

ОБЗОР ПРОГРАММЫ

Непараметрическая статистика

Модуль *Непараметрическая статистика* содержит всесторонний набор описательных статистик для анализа порядковых (ординальных) и номинальных данных. Дополнительно более специализированные методы доступны в других модулях *STATISTICA*; обратитесь к разделу *Альтернативные процедуры* ниже.

Альтернативы t-критерию для независимых выборок. Доступные альтернативы *t-критерию* для независимых выборок - критерий серий Вальда-Вольфовица, Манна-Уитни *U* тест, двухвыборочный критерий Колмогорова-Смирнова.

Альтернативы t- критерию для зависимых выборок. Доступные альтернативы t-критерию для зависимых выборок - критерий знаков и критерий Вилкоксона парных сравнений.

Альтернативы для межгрупповой ANOVA. Доступные альтернативы межгрупповой ANOVA - ранговый дисперсионный анализ Краскела-Уоллиса и медианный тест.

Альтернативы для ANOVA с повторно измеренными факторами. Доступные альтернативы - ранговая корреляция Спирмена *R*, *Кендалла тау*, *Гамма*, коэффициент конкордации Кендалла.

Процедуры для таблиц частот. Модуль *Непараметрическая статистика и распределения* включает различные статистики для анализа. Доступны статистики таблиц частот 2 x 2: *хи-квадрат*, *V-квадрат*, точный критерий Фишера, статистика *фи-квадрат* (альтернатива корреляции), критерий *хи-квадрат* Макнемара (используется, например, для проверки изменений частот); Кохрена *Q* тест - альтернатива ANOVA с повторными измерениями, если рассматриваемая переменная категориальная. Отметим, что другие модули *STATISTICA* также содержат много дополнительных процедур для анализа данных, представленных в виде частот (см. раздел *Альтернативные процедуры*, ниже).

Описательные статистики. Модуль *Непараметрическая статистика и распределения* вычисляет полный набор описательных статистик. Это - среднее, медиана, мода, процентили (определенные пользователем), геометрическое среднее, гармоническое среднее, стандартное отклонение, дисперсия, гармоническое среднее, размах, квартильный размах, асимметрия, эксцесс и т.д.

Подгонка распределений

Модуль *Непараметрическая статистика и распределения* позволяет табулировать наблюдаемые данные и вычислять соответствующие ожидаемые значения для различных непрерывных и дискретных распределений. Графики подгонки распределения доступны из таблиц результатов. Чтобы оценить согласие наблюдаемых данных с гипотетическим распределением, программа вычисляет критерий *хи-квадрат* и одновыборочный критерий Колмогорова-Смирнова. Непрерывные распределения, доступные в этом модуле, включают: нормальное распределение, равномерное распределение, экспоненциальное распределение, *Гамма* распределение, логнормальное распределение, *хи-квадрат* распределение; дискретные распределения включают: биномиальное распределение, распределение Пуассона, геометрическое распределение и распределение Бернулли. Пользователь также может оценить согласие

StatSoft®

Copyright © StatSoft, 1995

наблюдаемых данных с произвольным (заданным) набором ожидаемых частот. Дополнительно специальные распределения (например, Вейбулла, Гомпертца) доступны в модуле *Анализ выживаемости* (см. том III) и в модуле *Анализ процессов* (см. том IV), а также с помощью *Статистических 2М графиков*: *Квантиль-квантиль, Вероятность-вероятность, Гистограммы* (см. том II; см. также раздел *Альтернативные процедуры*, ниже).

Альтернативные процедуры

Корреляции. Вы можете вычислить стандартные корреляции Пирсона в каждом модуле *STATISTICA* с помощью опций *Быстрые основные статистики* (доступны из панели инструментов и выпадающего меню *Анализ*; см. главу 10). Также их можно вычислить в других модулях (например, в модуле *Основные статистики и таблицы*, глава 9; *Множественная регрессия*, глава 12; *Факторный анализ*, том III; *Менеджер мегафайлов*, том III). Тетрахорические корреляции (для категориальных переменных) доступны в модуле *Надежность/позиционный анализ* (см. том III). Различные меры расстояний доступны в модуле *Кластерный анализ*. Разнообразные меры зависимости между категориальными (номинальными) переменными доступны в разделе кросстабуляции модуля *Основные статистики и таблицы* (глава 9).

Модуль *Нелинейное оценивание* (том III) содержит различные нелинейные модели для анализа зависимости категориальных переменных (пробит и логит анализ); пользователь может также задать свой тип уравнения регрессии и функции потерь. Модуль *Анализ выживаемости* (том III) содержит специализированные регрессионные процедуры для обработки данных типа времени жизни (см. опцию *Регрессионные модели* в стартовой панели этого модуля); некоторые из этих процедур являются также непараметрическими (свободными от параметров) (например, регрессионная модель Кокса).

Таблицы частот (сопряженности). Дополнительные непараметрические меры связи между двумя категориальными (номинальными) переменными можно вычислить в модуле *Основные статистики и таблицы* (*хи-квадрат* тесты, статистики Кендалла *may b*, *may c*, энтропийные меры, коэффициенты зависимости; см. главу 9). Модуль *Логлинейного анализа* (том III) представляет собой продвинутую программу для анализа многоуровневых, многовходовых таблиц сопряженности (предельно большой размерности). *STATISTICA* также содержит программы *STATISTICA BASIC* (файл *Manthaen.stb* в поддиректории *STBASIC*) вычисления критерия Ментела-Хенцела для сравнения двух групп (см. Mantel and Haenszel, 1959; см. также Lee, 1992). Обратитесь к комментариям, приведенным в файле *Manthaen.stb* для дополнительных сведений.

Подгонка распределения. Модуль *Анализ выживаемости* (том III) подгоняет распределения Вейбулла и Гомпертца к наблюдаемым временам выживания и отказов. Можно обрабатывать также неполные (цензурированные) данные.

Гистограммные опции *Статистических 2М графиков* (см. меню *Графика*) предлагают разнообразные возможности для построения простых, кумулятивных и категоризованных графиков частот распределений. Доступны распределения: *Нормальное*, *Бета*, *Экстремального значения*, *Гамма*, *Геометрическое*, *Лапласа*, *Логистическое*, *Логнормальное*, *Пуассона*, *Релея* и *Вейбулла* (см. приложение I для описания этих распределений). Вы можете также построить графики *Квантиль-квантиль* и *Вероятность-вероятность* для этих распределений и оценить качество подгонки распределений к наблюдаемым данным. Эти опции подробно описаны в томе II руководства.

Copyright © StatSoft, 1995

Модуль *Анализ процессов* (см. том IV) содержит несколько опций для подгонки этих распределений и вычисления таблиц с ожидаемыми и наблюдаемыми частотами. Модуль предлагает также опцию подгонки одновременно всех выбранных непрерывных распределений к данным; при этом таблица результатов содержит соответствующие значения статистики Колмогорова-Смирнова *d-max*, расположенные в возрастающем порядке (от “лучшего” к “худшему”; т.е. от распределения, имеющее лучшее согласие к распределению, имеющему худшее согласие с данными).

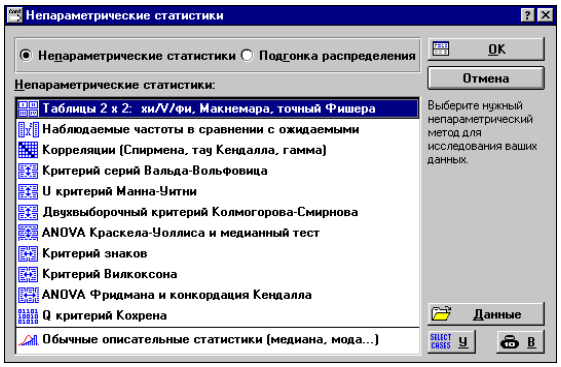
StatSoft®

Copyright © StatSoft, 1995

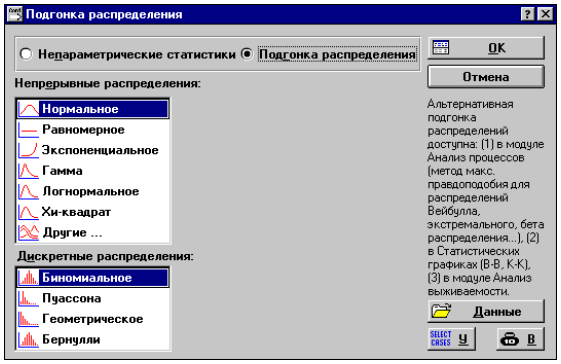
ДИАЛОГОВЫЕ ОКНА, ОПЦИИ, СТАТИСТИКИ

Стартовые панели

Стартовая панель этого модуля имеет два уровня: непараметрические статистики и подгонка распределения.



Вы можете переключаться между уровнями, выбирая соответствующую опцию в верхней части окна.



Каждая из этих опций описана в следующих разделах.

Стартовая панель - Непараметрические статистики

Таблицы частот 2 на 2: статистики *Хи/V/Фи-квадрат*, Макнемара, точный критерий Фишера

Опция открывает диалоговое окно, где можно ввести частоты в таблицу 2 x 2 (состоящую из двух строк и двух столбцов) и вычислить различные статистики, позволяющие оценить зависимость между переменными - *дихотомиями*, принимающими два значения. Более подробная информация об этой опции дана на стр. 1623.

Опция 2 x 2 может быть использована как альтернатива корреляциям, если обе рассматриваемые переменные являются дихотомиями. *STATISTICA* также вычислит критерий значимости *хи-квадрат* Макнемара, позволяющий проверить изменение частот. В этом случае можно рассматривать опцию 2 x 2 как альтернативу *t*-критерию для зависимых выборок, когда рассматриваемые переменные - дихотомии.

Наблюдаемые частоты в сравнении с ожидаемыми

Опция позволяет оценить согласие наблюдаемых частот с произвольным набором ожидаемых частот (см. также стр. 1624). Процедура сделает подсказку пользователю ввести 2 переменные: одна содержит ожидаемые, другая - наблюдаемые частоты. Для проверки согласия наблюдаемых и ожидаемых частот вычисляется стандартный критерий *хи-квадрат*.

Корреляции (Спирмена R, *тау* Кендалла, *Гамма*)

Опция позволяет вычислить три различных альтернативы коэффициенту корреляции Пирсона: корреляцию Спирмена *R*, статистику *тау Кендалла*, статистику *Гамма* (см. стр. 1624

Copyright © StatSoft, 1995

StatSoft

для описания каждой). После выбора опции на экране появится диалоговое окно, в котором можно выбрать переменные и определенный тип корреляции для вычисления. Можно вычислить одну непараметрическую корреляцию или матрицу непараметрических корреляций. Матрица диаграмм рассеяния для всех переменных будет построена, если нажать кнопку *Матричная диаграмма*. Таким образом пользователь может отобразить зависимости на графике. Альтернативные непараметрические меры связи между двумя переменными доступны также в модуле *Основные статистики и таблицы*, глава 9.

Критерий серий Вальда-Вольфовица

Критерий серий Вальда-Вольфовица представляет собой непараметрическую альтернативу *t*-критерию для независимых выборок (см. *Основные статистики и таблицы*, глава 9). Данные имеют тот же вид, что и для *t*-критерия для независимых выборок (см. *Основные статистики и таблицы*, глава 9). Файл должен содержать группирующую (*независимую*) переменную, имеющую, по крайней мере, два различных значения (кода), чтобы однозначно определить, к какой группе относится каждое наблюдение в файле данных.

Программа открывает диалоговое окно выбора группирующей (кодовой) переменной и списка зависимых переменных (переменных, по которым две группы сравниваются между собой), а также кодов для группирующей переменной (опция *Коды*).

Предположения и интерпретация. Критерий серий Вальда-Вольфовица устроен следующим образом: Представьте, вы хотите сравнить мужчин и женщин, описываемых некоторой переменной. Вы можете отсортировать данные с помощью этой переменной (например, расположить по возрастанию) и найти те случаи, когда субъекты одного и того же пола примыкают друг к другу в отсортированных данных (иными словами, образуют серию). Если нет различия между мужчинами и женщинами, то число и длина “серий”, относящиеся к одному и тому же полу, будут более или менее случайными. В противном случае, две группы (мужчины и женщины) отличаются друг от друга.

Критерий предполагает, что рассматриваемые переменные являются непрерывными и измерены, по крайней мере, в порядковой шкале (ранжированы). Критерий серий Вальда-Вольфовица проверяет гипотезу о том, что две независимые выборки извлечены из двух популяций, которые в чем-то существенно различны между собой, иными словами, различаются не только средними, но также формой распределения. Нулевая гипотеза состоит в том, что обе выборки извлечены из одной и той же популяции. В этом смысле критерий отличается от параметрического *t*-критерия, который предназначен для проверки отличия в положении (среднем) двух выборок.

U критерий Манна-Уитни

Критерий Манна-Уитни представляет непараметрическую альтернативу *t*-критерию для независимых выборок (см. *Основные статистики и таблицы*, глава 9). Опция предполагает, что данные расположены тем же образом, что в и *t*-критерии для независимых выборок. В частности, файл должен содержать кодовую (*независимую*) переменную, имеющую, по крайней мере, два разных кода для однозначной идентификации принадлежности каждого наблюдения к определенной группе.

Выбор опции открывает диалоговое окно выбора кодовой (группирующей) переменной и списка зависимых переменных (переменных, по которым две группы сравниваются), а также кодов, используемых в кодовой переменной для идентификации двух групп (опция *Коды*).

StatSoft®

Copyright © StatSoft, 1995

Предположения и интерпретация.
Критерий *U* Манна-Уитни предполагает, что рассматриваемые переменные измерены, по крайней мере, в порядковой шкале (ранжированы). Интерпретация теста по существу похожа на интерпретацию результатов *t*-критерия для независимых выборок, за исключением того, что *U* критерий вычисляется, как сумма индикаторов попарного сравнения элементов первой выборки с элементами второй выборки. *U* критерий - наиболее мощная (чувствительная) непараметрическая альтернатива *t*-критерию для независимых выборок; фактически, в некоторых случаях он имеет даже *большую* мощность, чем *t-критерий*.

Двухвыборочный критерий Колмогорова-Смирнова

Критерий Колмогорова-Смирнова – другая непараметрическая альтернатива *t-критерию* для независимых выборок (см. *Основные статистики и таблицы*, глава 9). Данные расположены таким же образом, как и для *t-критерия* для независимых выборок. Именно, файл данных должен содержать кодовую (*независимую*) переменную, имеющую, по крайней мере, два различных кода для однозначного определения к какой группе принадлежит каждое наблюдение.

Опция открывает диалоговое окно выбора кодовой переменной и списка зависимых переменных (переменных, по которым две группы сравниваются между собой), а также кодов, используемых в кодовой переменной для идентификации двух групп (опция *Коды*).

Предположения и интерпретация.
Критерий Колмогорова-Смирнова проверяет гипотезу о том, что выборки извлечены из одной и той же популяции, против альтернативной гипотезы: выборки извлечены из разных популяций. В отличие от параметрического *t*-критерия для независимых выборок и от *U* критерия Манна-Уитни (см. выше), который проверяет различие в положении двух выборок (различие в средних и различие в среднем рангов, соответственно), критерий Колмогорова-Смирнова также чувствителен к различию *общих форм* распределений двух выборок (в частности, различия в рассеянии, асимметрии и т.д.). Таким образом, интерпретация этого критерия похожа на интерпретацию критерия серий Вальда-Вольфовица (см. выше).

ANOVA Краскела-Уоллиса и медианный тест

Эти два теста являются непараметрическими альтернативами однофакторного дисперсионного анализа. *STATISTICA* предполагает, что файл данных имеет тот же вид, что и файл в модуле *ANOVA/MANOVA*. Именно, файл данных должен содержать кодовую (независимую) переменную для однозначной идентификации групповой принадлежности каждого наблюдения в файле. Процедура позволяет сравнивать до 10 групп.

Выбор опции открывает диалоговое окно выбора кодовой переменной (*переменные*) и списка зависимых переменных (переменных, по которым две группы сравниваются), а также кодов, используемых в кодовой переменной для идентификации сравниваемых групп (опция *Коды*).

Предположения и интерпретация.
Критерий Краскела-Уоллиса основан на рангах, (а не на исходных наблюдениях) и предполагает, что рассматриваемая переменная непрерывна и измерена как минимум в порядковой шкале. Критерий проверяет гипотезу: сравниваемые выборки имеют одно и тоже распределение или распределения с одной и той же медианой. Таким образом, интерпретация критерия схожа с интерпретацией параметрической однофакторной ANOVA, за исключением того, что этот критерий основан на рангах, а не на средних значениях.

Copyright © StatSoft, 1995

StatSoft®

Медианный тест - это “грубая” версия критерия Краскела-Уоллиса. *STATISTICA* просто подсчитывает число наблюдений каждой выборки, которые попадают выше или ниже общей медианы выборок, и вычисляет затем значение *хи-квадрат* для таблицы сопряженности 2 x *k*. При нулевой гипотезе (все выборки извлечены из популяций с равными медианами) ожидается, что примерно 50% всех наблюдений в каждой выборке попадают выше (или ниже) общей медианы. Медианный тест особенно полезен, когда шкала содержит искусственные границы, и многие наблюдения попадают в ту или иную крайнюю точку (оказываются “вне шкалы”). В этом случае медианный тест фактически единственный метод сравнения выборок.

Критерий знаков

Это непараметрическая альтернатива *t*-критерию для зависимых выборок (см. *Основные статистики и таблицы*, глава 9). Критерий применяется в ситуациях, когда исследователь проводит два измерения (например, при разных условиях) одних и тех же субъектов и желает установить, дают измерения разный результат или нет (имеется эффект обработки или нет).

После выбора опции на экране появится диалоговое окно, в котором можно выбрать переменные из двух списков. Каждая переменная первого списка сравнивается с каждой переменной во втором списке.

Предположения и интерпретация.
Требуются очень слабые предположения для применения этого критерия (например, однозначная определенность медианы для разности значений). Не нужно никаких предположений о природе или форме распределения. Критерий основан на следующих простых соображениях: подсчитывает сколько раз определенное значение первой переменной (*A*) больше соответствующего значения переменной (*B*), иными словами, определяется количество положительных разностей между значениями переменной (*A*) и значениями переменной (*B*). Учитываются только знаки разностей (а не их значения).

При нулевой гипотезе (состоящей в том, что две переменные не отличаются друг от друга, отсутствует эффект обработки) высчитанное значение будет равно примерно половине от общего числа наблюдений (положительных разностей будет примерно столько же сколько отрицательных). Основываясь на биномиальном распределении можно вычислить *z* значение для числа наблюдений, в которых $A > B$, а также вычислить соответствующую вероятность для этого *z* значения. Для малых объемов выборки *n* (меньше 20) предпочтительнее использовать значения, табулированные Siegel and Castellan (1988), чтобы оценить статистическую значимость.

Критерий Вилкоксона

Критерий Вилкоксона парных сравнений является непараметрической альтернативой *t*-критерию для зависимых (коррелированных) выборок (см. *Основные статистики и таблицы*, глава 9). После выбора опции на экране появится диалоговое окно, в котором можно выбрать переменные из двух списков. Каждая переменная первого списка сравнивается с каждой переменной второго списка. Это то же самое расположение данных, что и в *t*-критерии (зависимые выборки) в модуле *Основные статистики и таблицы*.

Предположения и интерпретация.
Предполагается, что рассматриваемые переменная переменные измерены, по крайней мере, в порядковой шкале (ранжированы), и что можно также ранжировать *различия* между переменными (этот тип шкалы иногда называют *упорядоченной метрической шкалой,* см. Coombs, 1950). Интерпретация теста по существу идентична интерпретации результатов *t*-критерия для независимых выборок, за

StatSoft®

Copyright © StatSoft, 1995

исключением того, что *U* критерий вычисляется по ранговым суммам, а не по средним. *U* критерий – это наиболее мощная (чувствительная) непараметрическая альтернатива *t*-критерию для независимых выборок; фактически, он имеет в некоторых случаях даже большую мощность, чем *t-критерий*.

Требования для критерия более строгие, чем для знакового (см. выше). Однако если они удовлетворены, то он обеспечивает большую мощность, чем знаковый. Фактически, если предположения параметрического *t-критерия* для зависимых выборок (интервальная шкала) выполнены, то *U* критерий имеет почти такую же мощность, как и *t-критерий*.

ANOVA Фридмана и коэффициент конкордации Кендалла

Эти тесты в чем-то различны по своей природе, однако они требуют похожих исходных данных. Фридмана ANOVA - непараметрическая альтернатива однофакторному дисперсионному анализу с повторными измерениями (см. *ANOVA/MANOVA*, глава 13). Коэффициент конкордации Кендалла похож на *R* Спирмена (непараметрический коэффициент корреляции между двумя переменными, см. выше) за исключением того, что он выражает зависимость между *многомерными* переменными.

В процедуре ANOVA Фридмана предполагается, что данные имеют тот же вид, что и для внутригруппового (с повторными измерениями) дисперсионного анализа в модуле *ANOVA/MANOVA*. В особенности, значения для каждого уровня фактора с повторными измерениями должны содержаться в различных переменных. Выбор опции открывает диалоговое окно спецификации списка переменных.

Предположения и интерпретация: Фридмана ранговая ANOVA. Предполагается, что рассматриваемые переменные (уровни фактора) измерены, как минимум, в порядковой шкале (ранжированы). Проверяемая нулевая гипотеза состоит в том, что различные столбцы данных (в частности, переменные *STATISTICA*) содержат выборки, извлеченные из одной и той же популяции, или популяций с равными медианами. Таким образом, интерпретация результатов здесь та же, что для ANOVA с повторными измерениями.

Предположения и интерпретация: конкордация Кендалла. Коэффициент конкордации Кендалла выражает одновременную взаимосвязь между *k* наблюдениями (коррелированной выборки). Эта статистика часто используется, чтобы оценить степень согласия нескольких экспертов. По существу, коэффициент конкордации - это усреднение всех коэффициентов *R* Спирмена по наблюдениям; более точно:

```
усред. Спирмена R = (k*конкордация-1)/(k-1)
```

Таким образом, общие предположения для этого критерия те же, что для коэффициента ранговой корреляции Спирмена (см. выше).

Q критерий Кохрена

Q критерий Кохрена - это развитие критерия *хи-квадрата* Макнемара для изменений (см. опцию *Таблиц 2 x 2*, выше) в частотах или долях *k* (*k* больше двух) зависимых выборок. Более точно, тестируется, значимо или нет различаются между собой несколько сравниваемых частот или долей. После выбора опции *Q–критерий Кохрена* в стартовой панели, программа предложит определить список переменных и коды, идентифицирующие две категории или два уровня переменной - дихотомии. Критерий предполагает, что переменные закодированы как единицы и нули, и коды, определенные пользователем, соответственно преобразуются в эти значения (только для данного анализа, сам по себе файл не будет изменен).

Предположения и интерпретация. Кохрена *Q* критерий требует либо данных в номинальной

Copyright © StatSoft, 1995

StatSoft®

шкале, либо искусственно дихотомированных данных. Следующий типичный пример иллюстрирует применение *Q* критерия: пусть хотят сравнить различные позиции дихотомического опросника, имеющего ответы типа да - нет. Здесь каждая переменная в файле данных представляет одну позицию и содержит нули (нет) и единицы (да). Если *Q* критерий значим, то вы можете заключить, что позиции опросника имеют существенно разную степень трудности, т.к. корректно ответили на них только немногие респонденты.

Стартовая панель – Описательные статистики

Выбор этой опции позволяет вычислить разнообразные описательные статистики (медиана, процентили, квартили, размах, квартильный размах) и другие описательные статистики (среднее, гармоническое среднее, геометрическое среднее, стандартное отклонение, асимметрия, эксцесс, дисперсия, гармоническое среднее, сумма) для выбранных переменных. Пользователь может также задать определенные процентили, чтобы они были вычислены и отображены в таблице результатов; по умолчанию, программа вычисляет квартили (25%, 75% процентили).

Дополнительно стандартные описательные статистики (минимум, максимум, среднее, число наблюдений), а также описанные ниже статистики вычисляются для каждой переменной.

Медиана

Медиана разбивает выборку на две равные части. Пятьдесят процентов наблюдений лежит ниже медианы, пятьдесят процентов - выше медианы. Если значение медианы существенно отличается от среднего, то форма распределения скошена.

Вычисление медианы. Если число наблюдений четно или имеются совпадающие наблюдения, определение медианы следует уточнить. Метод вычисления медиан, квартилей и других процентилей можно выбрать в диалоговом окне *STATISTICA Параметры по умолчанию: общие настройки* (см. главу 3; см. также *Приложение II*), доступном из меню *Сервис*. Выбранный метод далее используется во всех соответствующих местах *STATISTICA* (например, когда строятся диаграммы размаха или медианы и квартили с помощью соответствующих функций в *STATISTICA BASIC*; см. том V с описанием этих функций).

Мода

Мода - это максимально часто встречающееся значение в выборке. Частота встречаемости также отображается; если имеется несколько значений с максимальной частотой, то соответствующий столбец имеет отметку *множественная* (имеется несколько мод, распределение мультимодально).

Геометрическое среднее

Геометрическое среднее – это произведение всех значений переменной, возведенное в степень *1/n* (единица, деленная на число наблюдений). Геометрическое среднее полезно, например, когда шкала измерений нелинейная. Так в психометрии хорошо известно, что воспринимаемая интенсивность стимулов (например, воспринимаемая яркость света) часто представляет собой логарифмическую функцию реальной интенсивности стимулов (яркости, измеренной в объективных единицах - *люксах*).

В данном примере обычная оценка среднего (сумма значений, деленная на число стимулов), очевидно, не дает верного представления о среднем значении действительной интенсивности стимула, и следует вычислить геометрическое среднее. Геометрическое

StatSoft®

Copyright © StatSoft, 1995

среднее возникает в экономике (например, при вычислении некоторых индексов).

STATISTICA вычисляет геометрическое среднее с помощью логарифмического преобразования:

log(геометрическое среднее) = {Σ[log(x_i)]}/n

где

x_i *i*'е значение

n число наблюдений

Если переменная содержит отрицательные значения или нуль (*0*), геометрическое среднее нельзя вычислить.

Гармоническое среднее

Гармоническое среднее иногда используют для усреднения частот. Гармоническое среднее вычисляется как:

ГС = n/Σ(1/x_i)

где

ГС гармоническое среднее

n число наблюдений

x_i значение наблюдения с номером *i*

Если переменная содержит нуль (*0*), гармоническое среднее нельзя вычислить.

Дисперсия и стандартное отклонение

Выборочная дисперсия и стандартное отклонение - наиболее часто используемые меры изменчивости или разброса значений переменной (см. *Основные статистики и таблицы*, глава 9). *STATISTICA* вычисляет дисперсию как сумму квадратов отклонений значений переменной от выборочного среднего, деленную на *n-1* (но не на *n*). Стандартное отклонение вычисляется как корень квадратный из оценки дисперсии.

Среднее отклонение

Среднее отклонение представляет собой другую меру изменчивости. Вычисляется как сумма абсолютных отклонений значений переменной от выборочного среднего, деленная на *n* (число наблюдений).

Размах

Размах переменной также является показателем изменчивости. Он вычисляется как наибольшее значение переменной минус наименьшее значение.

Квартильный размах

Квартильный размах равен: верхняя квартиль минус нижняя квартиль (75% процентиль минус 25% процентиль). Так как 75% процентиль (верхняя квартиль) - это значение, слева от которого находятся 75% наблюдений, а 25% процентиль (нижняя квартиль) - это значение, слева от которого находится 25% наблюдений, то квартильный размах – представляет собой интервал вокруг медианы, который содержит 50% наблюдений (значений переменной).

Асимметрия

Асимметрия - характеристика формы распределения, а именно степени скошенности. Распределение скошено влево, если значение асимметрии отрицательно. Распределение скошено вправо, если асимметрия положительна. Асимметрия стандартного нормального распределения равна *0*. Асимметрия связана с *третьим моментом* и определяется как:

Асимметрия = n*$М^3$/[(n-1)*(n-2)*σ^3]

где

$М^3$ равно: $\Sigma(x_i\text{-Среднее}_x)^3$

σ^3 стандартное отклонение, возведенное в третью степень

Copyright © StatSoft, 1995

StatSoft®

n число наблюдений

Эксцесс

Эксцесс - характеристика формы распределения, а именно мера остроты его пика (относительно нормального распределения, эксцесс которого равен *0*). Как правило, распределения с более острым пиком, чем у нормального, имеют положительный эксцесс; распределения, пик которых менее острый, чем пик нормального распределения, имеют отрицательный эксцесс. Эксцесс связан с *четвертым моментом* и определяется формулой:

$$\text{Эксцесс} = [n*(n+1)*M^4 - 3*M^2*M^2*(n-1)]/[(n-1)*(n-2)*(n-3)*\sigma^4]$$

где

M^j равно: $\Sigma(x_i\text{-Среднее}_x)^j$

σ^4 стандартное отклонение в четвертой степени

n число наблюдений

Стартовая панель – Подгонка распределения

Общее

После выбора любого из приведенных далее распределений, вы можете указать переменную, к значениям которой распределение будет подогнано. Независимо от выбранного распределения, вы можете переключаться между всеми другими распределениями одного и того же типа (*непрерывное* или *дискретное*) во всех последующих окнах.

После того как переменная задана, вычисляются параметры, обеспечивающие наилучшую подгонку распределения данного типа. Эти параметры отображаются, по умолчанию. Пользователь может также ввести свои значения параметров распределения.

Подгонка распределения с помощью опции Статистические 2М графики

Начиная с версии 5, *STATISTICA* имеет несколько альтернативных процедур подгонки непрерывных распределений к данным. В частности, можно выбрать (категоризованные или некатегоризованные) графики *квантиль-квантиль*, *вероятность-вероятность* графики, или *2М гистограммы* (с кривой, показывающей ожидаемые частоты) из меню *Графика,* и затем построить соответствующие графики, основываясь на распределениях *Бета*, *Экспоненциальное*, *Экстремального значения*, *Гамма*, *Лапласа*, *Логистическое*, *Логнормальное*, *Релея* или *Вейбулла*. Программа вычисляет оценки максимального правдоподобия параметров распределения (см. том II).

Нормальное распределение

Нормальное распределение (так называемая “колоколообразная кривая”, симметричная относительно среднего) - это теоретическая функция, особенно часто используемая в статистике (см. *Элементарные понятия статистики*, глава 8). Вообще, нормальное распределение дает хорошую модель, если:

(1) Имеется сильная тенденция данных принимать центральное значение;

(2) Положительные и отрицательные отклонения от этого центрального значения равновероятны;

(3) Частота отклонений быстро падает, когда отклонения становятся большими.

Механизм, лежащий в основе нормального распределения, можно представить следующим образом. Имеется бесконечное число независимых случайных событий, которые вносят вклад в значения наблюдаемой переменной. Например, имеется практически бесконечное число факторов, определяющих вес человека (тысячи генов, предрасположенность,

StatSoft®

Copyright © StatSoft, 1995

болезни и т.д.). Таким образом, можно ожидать нормальное распределение для веса в популяции всех людей. Плотность нормального распределения имеет вид:

$$f(x)=\{1/[(2*\pi)^{1/2}*\sigma]\} * e^{\{-1/2*[(x-\mu)/\sigma]^2\}}$$

где

μ среднее

σ стандартное отклонение

e константа Эйлера (2.71...)

π константа *Pi* (3.14...)

Равномерное распределение

Равномерное распределение полезно для описания переменных, в которых каждое значение равновероятно:

$$f(x) = a$$

где

a константа

Экспоненциальное распределение

Если *T* - время между наступлениями редких событий, происходящих в среднем с интенсивностью λ, то величина *T* имеет экспоненциальное распределение с параметром λ (*лямбда*). Экспоненциальное распределение часто используется для описания временных интервалов между последовательными случайными событиями, например, временных интервалов между автомобилями, пересекающими перекресток, длительностей жизни электронных приборов или приходов покупателей в кассу в бакалейном магазине. Плотность экспоненциального распределения имеет вид:

$$f(x) = \lambda*e^{-\lambda x}$$

где

λ (*лямбда*) параметр

e константа Эйлера (2.71...)

Гамма распределение

Плотность экспоненциального распределения имеет моду 0. Во многих примерах, *a priori* известно, что мода рассматриваемой случайной переменной не равна 0 (например, когда моделируется распределение длительности жизни электрических ламп или реальная (измеренная секундомером) длительность баскетбольного матча). В этих случаях *Гамма* распределение более подходит для описания распределения. Плотность *Гамма* распределения имеет вид:

$$f(x) = [1/(\Gamma(\alpha)*\beta^{\alpha})] * x^{\alpha-1} * e^{-x/\beta}$$

где

Γ (*гамма*) *Гамма* функция (аргумента *альфа*)

α (*альфа*) параметр формы

β (*Бета*) параметр масштаба

e константа Эйлера (2.71...)

Логнормальное распределение

Логнормальное распределение часто используется в моделировании таких переменных как персональные доходы, возраст новобрачных (точнее первобрачных – первый раз вступающих в брак) или допустимое отклонение от стандарта вредных веществ в продуктах питания. В общем, если величина *x* имеет нормальное распределение, то $y = e^x$ имеет логнормальное распределение. Плотность логнормального распределения имеет вид:

$$f(x) = \{1/[x*\sigma*(2*\pi)^{1/2}]\} * e^{\{-1/2*[(\ln(x)-\mu)/\sigma]^2\}}$$

Copyright © StatSoft, 1995

StatSoft®

$$e^{-\{1/[2*(\sigma)]\}*[\ln(x)-\mu]}$$

где

π константа *Pi* (3.14...)

σ стандартное отклонение

e константа Эйлера (2.71...)

μ среднее

Хи-квадрат распределение

Сумма квадратов *v* независимых случайных переменных, каждая из которых имеет стандартное нормальное распределение, имеет распределение *хи-квадрат* с *v* степенями свободы. Это распределение наиболее часто используется в моделировании случайных переменных в статистических приложениях. Плотность *хи-квадрат* распределения имеет вид:

$$f(x)=\{1/[2^{v/2}*\Gamma(v/2)]\} * x^{(v/2)-1} * e^{-x/2}$$

где

v степень свободы

e константа Эйлера (2.71...)

Γ(y) *Гамма* функция (аргумента *y*)

Биномиальное распределение

Биномиальное распределение полезно для описания распределения биномиальных событий, таких как число мужчин и женщин в случайно выбранных компаниях или число дефектов в выборке продукции из 20 единиц. Биномиальное распределение имеет вид:

$$f(x) = [n!/(x!*(n-x)!)]*p^x * q^{n-x}$$

где

p вероятность наступления события

q равно *1-p*

n число испытаний

Распределение Пуассона

Распределение Пуассона иногда называют распределением редких событий. Примеры переменных, распределенных по закону Пуассона: число несчастных случаев, число фатальных дефектов в производственном процессе. Распределение Пуассона определяется формулой:

$$f(x) = (\lambda^x * e^{-\lambda})/x!$$

где

λ (*лямбда*) среднее значение

e константа Эйлера (2.71...)

Геометрическое распределение

Если проводятся независимые испытания Бернулли (см. ниже) и подсчитывается количество испытаний до наступления “успеха”, то это число имеет геометрическое распределение. [Если вы бросаете монету, то число подбрасываний, которое вам нужно сделать до выпадения очередного герба, подчиняется геометрическому закону.]

Геометрическое распределение определяется формулой:

$$f(x) = p*(1-p)^{x-1}$$

где

p вероятность того, что определенное событие (например, успех) произойдет

Распределение Бернулли

Это распределение наилучшим образом описывает ситуации, где “испытание” имеет результат “успех”, либо “неуспех”, например,

StatSoft®

Copyright © StatSoft, 1995

при бросании монеты, или при моделировании удачной или неудачной хирургической операции.

Распределение Бернулли определяется формулой:

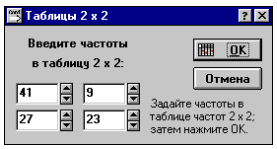
$$f(x) = p^{x}*(1-p)^{1-x}$$

где

p вероятность наступления определенного события (например, успеха)

Таблицы частот 2 x 2

Задайте частоты для *Таблицы 2 x 2* в этом диалоговом окне (см. также стр. 1613).



[Типичный пример таких таблиц: число мужчин и женщин, предпочитающих рекламу ПЕПСИ или КОКИ, или число заболевших и не заболевших людей, из числа сделавших и не сделавших прививки и т.д.]

Дополнительно к стандартному критерию *хи-квадрат* Пирсона и скорректированному *хи-квадрат* (*V-квадрат*, см. стр. 1634), можно вычислить следующие статистики (см. таблицу результатов ниже):

Таблица 2x2 (intrview.sta)

НЕПАРАМ. СТАТИСТ.	Столб. 1	Столб. 2	Сумма по строкам
Частоты, строка 1	41	9	50
Процент от общего	41.000%	9.000%	50.000%
Частоты, строка 2	27	23	50
Процент от общего	27.000%	23.000%	50.000%
Сумма по столбцу	68	32	100
Процент от общего	68.000%	32.000%	
Хи-квадрат (ст.св.=1)	9.01	p= .0027	
V-квадрат (ст.св.=1)	8.92	p= .0028	
Поправка Йетса	7.77	p= .0053	
Фи коэффициент	.09007		
Фишера р, односторонний		p= .0025	
двусторонний		p= .0049	
Макнемара Хи-квадрат (A/D)	4.52	p= .0336	
Хи-квадрат (B/C)	8.03	p= .0046	

Поправка Йетса. Свойства критерия *хи-квадрат* (точность аппроксимации распределения статистики распределением хи-квадрат) для таблиц 2 на 2 с малыми ожидаемыми частотами могут быть улучшены уменьшением абсолютного значения разностей между ожидаемыми и наблюдаемыми частотами на величину 0.5 перед возведением в квадрат (*Поправка Йетса*). Это так называемая поправка Йетса на непрерывность для таблиц частот 2на 2, которая обычно применяется, когда ячейки содержат только малые частоты, и некоторые ожидаемые частоты становятся меньше 10 (для дальнейшего обсуждение см. Conover, 1974; Everitt, 1977; Hays, 1988; Kendall and Stuart, 1979; Mantel, 1974).

***Фи* коэффициент.** Статистика *фи-квадрат* представляет собой меру связи (зависимости) между номинальными переменными (для которых не определено упорядочивание).

Точный критерий Фишера. Пусть даны маргинальные частоты в таблице 2 на 2, предположим, что оба фактора в таблице независимы. Зададимся вопросом: какова вероятность получить наблюдаемые частоты, исходя из данных маргинальных? Эта вероятность вычисляется *точно,* подсчетом всех возможных таблиц, которые можно построить, основываясь на данных маргинальных частотах. Таким образом, критерий Фишера вычисляет *точную* вероятность появления при нулевой гипотезе - когда факторы независимы - наблюдаемых частот в ячейке. Вычисляются односторонние и двухсторонние вероятности.

Критерий *хи-квадрат* Макнемара. Этот критерий применяется, когда частоты в таблице 2 на 2 получены по *зависимым* выборкам. Например, наблюдения до и после воздействия на одном и том же экспериментальном материале. Вы можете подсчитывать число студентов, имеющих минимальные успехи по математике в начале и конце семестра. Два значения *хи-квадрата* вычисляются: *A/D* и *B/C*.

Copyright © StatSoft, 1995

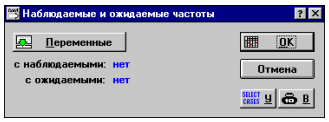
StatSoft

A/D хи-квадрат проверяет гипотезу о том, что частоты в ячейках *A* и *D* (верхняя левая, нижняя правая) одинаковы. *B/C хи-квадрат* проверяет гипотезу о равенстве частот в ячейках *B* и *C* (верхняя правая, нижняя левая).

Отметим, что *STATISTICA* также включает два модуля, предназначенные для анализа частот в многомерных таблицах практически неограниченной размерности и сложности (*Таблицы* в *Модуле Основные статистики и таблицы* и модуль *Логлинейный* анализ). Процедура *Таблицы* модуля *Основные статистики и таблицы* вычисляет всесторонний набор описательных статистик. *Логлинейный* анализ выполняет полный логлинейный анализ многовходовых таблиц. *STATISTICA* содержит пример *STATISTICA BASIC* программы (файл *Manthaen.stb* в каталоге *STBASIC*) для вычисления Критерий Ментела-Хенцела для сравнения двух групп (см. Mantel and Haenszel, 1959; см. также Lee, 1992). Обратитесь к комментариям в файле *Manthaen.stb* за дополнительной информацией.

Наблюдаемые частоты в сравнении с ожидаемыми

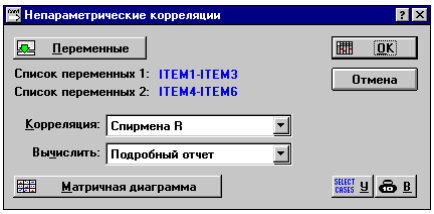
Задайте две переменные: одну, содержащую наблюдаемые частоты, вторую, содержащую ожидаемые частоты.



Пример непараметрического критерия см. на стр. 1637.

Коэффициенты корреляции

Если выбрана опция *Корреляции (Спирмена, may Кендалла, гамма*) из стартовой панели *Непараметрические статистики*, будет открыто диалоговое окно *Непараметрические корреляции*.



Переменные

Выберите как минимум две переменные, между которыми будут вычислены корреляции (если переменных в списках несколько, будут вычислены корреляции между каждой переменной в первом списке и каждой переменной во втором списке).

Корреляция

Можно выбрать следующие ранговые корреляции.

Статистика Спирмена *R*. Статистика *R* Спирмена предполагает, что рассматриваемые переменные измерены как минимум в порядковой шкале, иными словами индивидуальные наблюдения ранжированы.

Статистика Кендалла *may*. Статистика *may* Кендалла эквивалентна *R* Спирмена при выполнении некоторых основных предположений. Критерии, основанные на этих статистиках, также сравнимы по мощности. Однако обычно значения *R* Спирмена и *may* Кендалла различны, потому что они существенно отличаются как по своей внутренней логике, так и по способу вычисления. Siegel и Castellan (1988) выразили соотношение между ними следующим неравенством:

```
-1 ≤ 3*тау Кендалла - 2*R Спирмена ≤ 1
```

Более важно, что *may* Кендалла и R Спирмена по-разному интерпретируются. R Спирмена

StatSoft®

Copyright © StatSoft, 1995

можно мыслить как прямой аналог *r* Пирсона, вычисленный по рангам (а не по исходным наблюдениям). Тогда как *may* Кендалла представляет *вероятность,* точнее, разность между вероятностью того, что значения двух переменных располагаются в одном и том же порядке, минус вероятность того, что значения переменных располагаются в различном порядке (или вероятность того, что ранги двух переменных совпадают минус вероятность того, что они различны).

Гамма. *Гамма* статистика предпочтительнее статистики *R* Спирмена или *may* Кендалла в тех случаях, когда в данных имеется много совпадающих значений. С точки зрения основных предположений статистика *гамма* эквивалентна *R* Спирмена или *may* Кендалла. Ее интерпретация и вычисление более похожи на *may Кендалла,* чем на R Спирмена. *Гамма* представляет собой также *вероятность*; более точно, вероятность того, что ранговый порядок двух переменных совпадает минус вероятность того, что не совпадает, деленная на выражение: 1 минус вероятность совпадений. Таким образом, *Гамма* в основном эквивалентна *may* Кендалла, за исключением того, что совпадения рангов явно принимаются во внимание.

Вычислить

Выбрав опцию, можно вычислить как единственную непараметрическую корреляцию, так и матрицу корреляций. Таблица результатов, по умолчанию, изображает диаграмму рассеяния. Таким образом, вы можете легко посмотреть зависимости на графике.

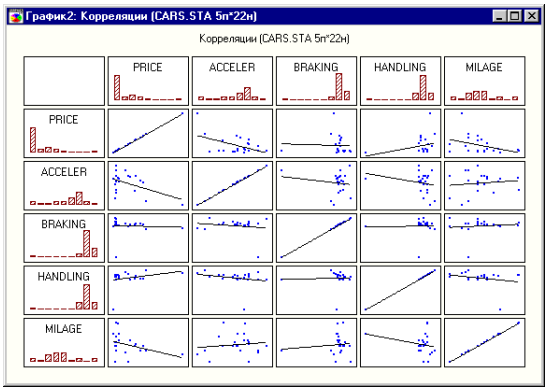
Подробный отчет. Опция вычисляет не только корреляции между выбранными переменными (см. выше), но также *число наблюдений n*, *t-значение* и *p-уровень* для каждой корреляции.

Матрицы двух списков. Опция вычисляет только корреляции между переменными, заданными в первом списке с переменными, заданными во втором списке.

Квадратная матрица. Опция вычисляет корреляции для одного списка переменных (квадратная матрица). Заметим, если выбраны два списка переменных, а затем выбрана эта опция, то списки будут “объединены” в один.

Матричная диаграмма

Нажмите кнопку, чтобы построить матричную диаграмму рассеяния для выбранных переменных.



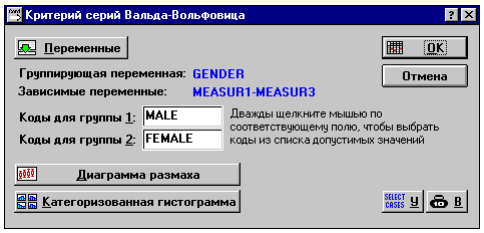
Этот график полезен тем, что позволяет быстро оценить и сравнить распределения выбранных переменных и форму зависимости между ними (например, коэффициент ранговой корреляции *R* Спирмена может измерять нелинейную зависимость между переменными, при условии, что она монотонна).

Copyright © StatSoft, 1995

StatSoft®

Сравнение двух групп

Следующие непараметрические критерии доступны для сравнения двух групп.



Критерий серий Вальда-Вольфовица. Критерий серий Вальда-Вольфовица устроен следующим образом: Представьте, вы хотите сравнить мужчин и женщин, описываемых некоторой переменной, например, измеряющей рост или кровяное давление. Вы можете отсортировать данные с помощью этой переменной, расположив их в порядке возрастания, иными словами, построить вариационный ряд, и найти те последовательности вариационного ряда, в которых субъекты одного и того же пола примыкают друг к другу (образуют серию). Если между мужчинами и женщинами нет различия, то число и длина таких “серий” (относящиеся к одному и тому же полу) будут более или менее случайно расположены в вариационном ряде. В противном случае, две группы, в данном примере мужчины и женщины, чем-то существенно отличаются друг от друга.

Критерий предполагает, что рассматриваемые переменные являются непрерывными и измерены, по крайней мере, в порядковой шкале (иными словами, их можно упорядочить). Критерий Вальда-Вольфовица проверяет гипотезу о том, что две независимые выборки извлечены из двух популяций, которые в чем-то существенно различны между собой, иными словами, различаются не только средним, но также формой распределения. Нулевая гипотеза состоит в том, что обе выборки извлечены из одной и той же популяции. В этом тест отличается от параметрического *t*-критерия, который предназначен для проверки различия в положении (в средних) двух выборок.

Манна-Уитни *U* критерий. Манна-Уитни *U* критерий предполагает, что рассматриваемые переменные измерены, по крайней мере, в порядковой (ординальной) шкале (ранжированы). Интерпретация теста по существу идентична интерпретации результатов *t*-критерия для независимых выборок, за исключением того, что *U* критерий вычисляется по рангам, а не по средним. Две группы объединяются, наблюдения упорядочиваются по возрастанию, и подсчитывается, сколько раз ранг одной группы предшествует рангу другой группы. *U* критерий – это наиболее мощная (чувствительная) непараметрическая альтернатива *t*-критерию для независимых выборок; фактически, в некоторых случаях он имеет даже *большую* мощность, чем *t-критерий*.

Двухвыборочный критерий Колмогорова-Смирнова. Критерий Колмогорова-Смирнова проверяет гипотезу о том, что выборки извлечены из одной и той же популяции, против альтернативы, что выборки извлечены из разных популяций. Он основан на максимуме абсолютного значения разности эмпирических функций первой и второй выборки. В отличие от параметрического *t*-критерия для независимых выборок и от критерия Манна-Уитни (см. выше), который проверяет различие в положении двух выборок (различие в средних и различие в средних рангах, соответственно), критерий Колмогорова-Смирнова также чувствителен к различию *общей формы* распределений двух выборок (в частности, различие в дисперсии, асимметрии и т.д.). Таким образом, его интерпретация похожа на интерпретацию критерия серий Вальда-Вольфовица (см. выше).

StatSoft®

Copyright © StatSoft, 1995

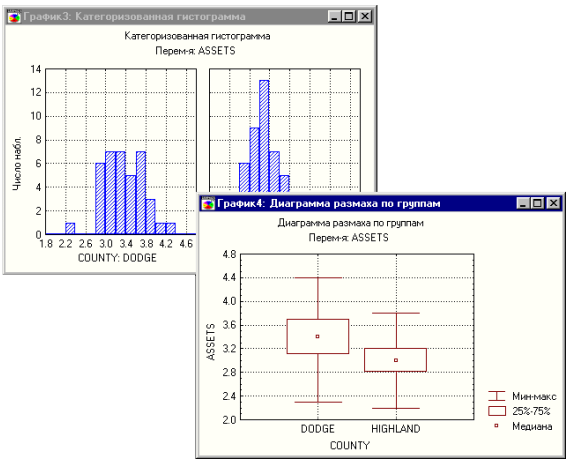
Переменные

Опция открывает обычное окно выбора двух списков переменных. Процедура ожидает, что данные расположены в том же порядке, что и для *t*-критерия для независимых выборок. Файл должен содержать кодовую переменную (*независимую* или *группирующую* переменную), имеющую как минимум два значения для однозначной идентификации групповой принадлежности каждого наблюдения. Вначале выберите в первом списке коды групп; во втором - переменные для сравнения.

Затем определите коды группирующей переменной, чтобы однозначно определить групповую принадлежность каждого наблюдения. Можно также дважды щелкнуть мышью на поле ввода, чтобы открыть окно *значений переменной* и выбрать в нем коды.

Диаграмма размаха

Диаграмма размаха – полезный визуальный метод сравнения распределения двух групп данных.

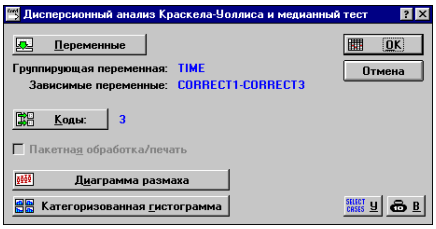


Категоризованная гистограмма

Опция строит категоризованные гистограммы распределения выбранных переменных, отдельно для каждой группы (как показано выше).

ANOVA Краскела-Уоллиса и медианный тест

Краскела-Уоллиса ANOVA, основанная на рангах, предполагает, что рассматриваемая переменная непрерывна и измерена как минимум в порядковой шкале. Критерий проверяет гипотезу о том, что сравниваемые выборки имеют одно и тоже распределение или распределения с одной и той же медианой. Таким образом, его интерпретация схожа с интерпретацией параметрической однофакторной ANOVA, за исключением того, что дисперсионный анализ Краскела-Уоллиса основан на рангах (наблюдений), а не на средних.



Медианный тест - это “грубая” версия Краскела-Уоллиса ANOVA в том смысле, что вычисления производятся в терминах таблицы сопряженности. *STATISTICA* просто подсчитывает число наблюдений каждой выборки, которые попадают выше или ниже общей медианы, затем вычисляет значение *хи-квадрат* для 2 x *k* таблицы сопряженности. При нулевой гипотезе (все выборки извлечены из популяций с равными медианами) ожидается, что примерно 50% всех наблюдений в каждой выборке попадают выше (или ниже) общей медианы. Медианный тест особенно полезен,

Copyright © StatSoft, 1995

StatSoft

когда шкала измерений содержит искусственно заданные границы, и многие наблюдения попадают в ту или иную крайнюю точку шкалы (т.е. лежат “вне шкалы”). В этом случае медианный критерий, фактически, единственный метод сравнения выборок.

Переменные

Опция открывает обычное окно выбора двух списков переменных. Предполагается, что данные расположены, как и для *t*-критерия для независимых выборок. Файл должен содержать кодовую переменную (*независимую* или *группирующую* переменную), имеющую как минимум два значения для однозначной идентификации групповой принадлежности каждого наблюдения. Таким образом, выберите в первом списке коды групп; во втором - переменные для сравнения.

Коды

Кнопка открывает диалоговое окно *выбора кодов*, в котором можно выбрать коды для однозначного определения группы, к которой принадлежит каждый наблюдение.

Пакетная обработка/ печать

Опция применяется, если ранее выбраны устройства вывода *Принтер, Файл вывода* и/или *Окно* в диалоговом окне *Параметры страницы/вывода*. Если выбрана опция *Пакетная обработка/печать,* то все переменные в списке зависимых переменных будут подвергнуты обработке, а результаты отпечатаны без дополнительного ввода исходных данных.

Диаграмма размаха

Диаграмма размаха – полезный визуальный метод сравнения распределения в двух группах (в опциях графика доступны медианы и квартили или средние и другие меры рассеяния).

Категоризованная гистограмма

Опция строит категоризованные гистограммы распределения выбранных переменных, отдельно для каждой группы (как показано выше).

Сравнение двух переменных

Следующие два непараметрических теста используются для сравнения двух переменных.

Критерий знаков. Требуется только одно предположение для применения этого теста - распределение рассматриваемой переменной должно быть непрерывным. Не нужно никаких дополнительных предположений о природе или форме распределения. Критерий устроен следующим образом: из значений первой переменной (*A*) вычитаются значения второй переменной (*B*) и подсчитывается количество положительных разностей.

При нулевой гипотезе (состоящей в том, что две переменные не отличаются друг от друга) высчитанное значение будет равно примерно половине от общего числа наблюдений. Основываясь на биномиальном распределении, можно вычислить *z* значение для наблюдаемого числа случаев, в которых $A > B$, также вычислить соответствующую вероятность хвоста и рассчитать уровень значимости. Для малых размеров выборки *n* (меньше 20) лучше использовать значения, табулированные Siegel and Castellan (1988), чтобы оценить статистическую значимость критерия знаков.

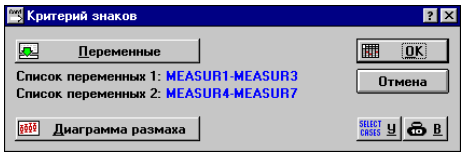
Критерий Вилкоксона парных сравнений. Предполагается, что рассматриваемые переменные измерены, как минимум, в порядковой шкале (ранжированы) и что можно

StatSoft®

Copyright © StatSoft, 1995

также ранжировать *различия* между переменными (этот тип шкалы иногда называют *упорядоченной метрической шкалой,* см. Coombs, 1950). Интерпретация теста, по существу, схожа с интерпретацией результатов *t*-критерия для независимых выборок, за исключением того, что *U* тест вычисляется по рангам (наблюдаемых величин), а не по средним. *U* тест – это наиболее мощная (чувствительная) непараметрическая альтернатива *t*-критерию для независимых выборок, фактически, в некоторых случаях имеющая даже большую мощность, чем *t-критерий*.

Требования к данным для этого теста более строгие, чем для критерия знаков (см. выше). Однако если они выполняются, он обеспечивает большую мощность, чем критерий знаков. Фактически, если предположения параметрического *t*-критерия для зависимых выборок (интервальная шкала) выполняются, то этот тест имеет почти такую же мощность, как и *t-критерий*.



Переменные

Опция открывает стандартное окно выбора двух списков переменных.

Диаграмма размаха

Диаграмма размаха – полезный визуальный метод сравнения распределения в двух группах. После нажатия кнопки, выберите две переменные для графика (в опциях графика доступны медианы и квартили или средние и другие меры рассеяния).

Ранговый дисперсионный анализ Фридмана и конкордация Кендалла

Фридмана ранговая ANOVA. Предполагается, что рассматриваемые переменные (уровни) измерены, как минимум, в порядковой шкале (ранжированы). Проверяемая нулевая гипотеза состоит в том, что разные столбцы данных (в частности, переменные *STATISTICA*) содержат выборки, извлеченные или из одной и той же популяции, или извлеченные из популяций, имеющих равные медианы. Таким образом, интерпретация результатов критерия та же, что для ANOVA с повторными измерениями.

Конкордация Кендалла. Коэффициент конкордации Кендалла выражает взаимосвязь между *k* наблюдениями (коррелированной выборки) одновременно. Часто эта статистика используется для того, чтобы оценить согласованность мнений экспертов. По существу, коэффициент конкордации - это усреднение всех коэффициентов *R* Спирмена; более точно:

```
усред. Спирмена R = (k*конкордация-1)/(k-1)
```

Таким образом, общие предположения для этого критерия те же, что для ранговой корреляции Спирмена.

Переменные

Опция открывает стандартное окно выбора одного списка переменных для анализа.

Диаграмма размаха

Диаграмма размаха - полезный график, позволяющий сравнить распределения двух или

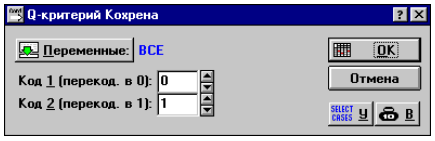
Copyright © StatSoft, 1995

StatSoft®

более переменных. После нажатия кнопки, выберите две переменные для этого графика (на графике доступны медианы и квартили или средние и другие меры рассеяния).

Кохрена Q критерий

Q критерий Кохрена - это развитие *хи-квадрата* Макнемара для тестирования изменений (см. опция *Таблиц 2x2* выше) в частотах или долях *k* (больше двух) зависимых выборок. Более точно, проверяется, значимо или нет различаются между собой несколько сравниваемых частот или долей. После выбора опции *Q*-критерий Кохрена в стартовой панели, программа предложит задать список переменных и коды, идентифицирующие две категории или два уровня дихотомийной меры. Критерий предполагает, что переменные, кодированы как единицы и нули, и коды, определенные пользователем, соответственно преобразуются (только для этого анализа, сам по себе файл не будет изменен). Типичный пример использования *Q* критерия, следующий: пусть хотят сравнить трудность различных позиций дихотомического опросника с ответами типа: правильно или неправильно (да-нет). Здесь каждая переменная в файле данных представляет одну позицию и содержит нули (нет) и единицы (да). Если *Q* критерий значим, то вы можете заключить, что позиции опросника имеют разную степень трудности, т.к. корректно ответили на них только несколько респондентов.



Переменные

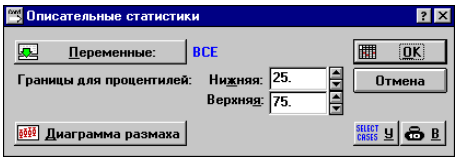
Опция открывает стандартное окно выбора одного списка переменных для анализа. Заметим, критерий требует, чтобы переменные содержали два целых значения или коды (см. ниже).

Коды

Q критерий Кохрена предполагает, что выбранные переменные имеют значения 0 или 1 (и только). Если используются другие коды (отличные от 0-1), задайте их в поле ввода; программа временно преобразует их в 0-1 (исходный файл не изменится).

Описательные статистики

Выбрав *Обычные описательные статистики* из стартовой панели *Непараметрические статистики,* вы откроете окно *Описательные статистики.*



Эта опция позволяет вычислить несколько описательных статистик (каждая из которых объясняется в разделе *Стартовая панель - Описательные статистики*, стр. 1618).

Переменные

Опция открывает стандартное окно выбора одного списка переменных для анализа.

Вычислить процентили

Задайте значения процентилей для вычисления и отображения в таблице результатов, по определению, процентиль - это значение, левее которого находится заданный процент наблюдений, рассматриваемой переменной; по умолчанию, программа вычисляет квартили, т.е. 25%, 75% процентили.

StatSoft®
Copyright © StatSoft, 1995

Вычисление медианы, квартилей, процентилей. Если число наблюдений четно, или имеются совпадающие наблюдения, в вычислениях процентилей появляется неточность. Метод вычисления может быть выбран в окне *STATISTICA Параметры по умолчанию: общие настройки STATISTICA* (см. главу 3; см. также *Приложение II* для технических деталей) доступном из меню *Сервис*. Однажды выбранный метод будет использоваться на всем протяжении работы со *STATISTICA* (например, когда строятся диаграммы размаха, или когда вычисляются медианы и квартили в *STATISTICA BASIC*; см. том V).

Вычисление описательных статистик

В дополнении к стандартным описательным статистикам *STATISTICA* вычислит несколько дополнительных описательных статистик, когда Вы нажмете *OK* в этом диалоговом окне.

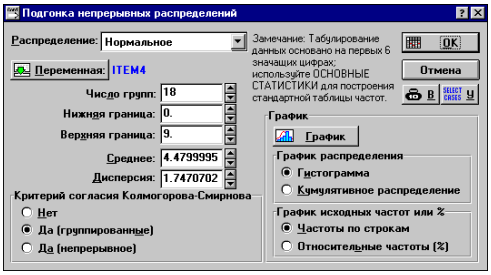
Описат. статистики [10items.sta]

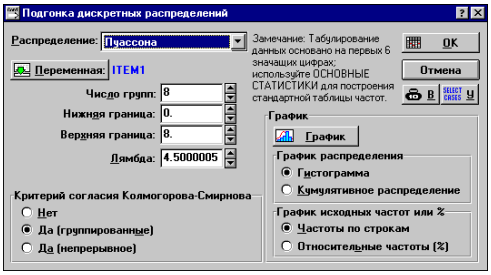
НЕПАРАМ. СТАТИСТ.	среднее	N	медиана	75.000ая процент.	геометр. среднее	гармон. среднее	ст. откл
ITEM1	4.500000	100	4.000000	6.000000	4.222662	3.863668	1.445998
ITEM2	4.740000	100	5.000000	6.000000	4.559502	4.361824	1.260271
ITEM3	4.700000	100	5.000000	6.000000	4.466244	4.161712	1.352140
ITEM4	4.480000	100	5.000000	5.000000	4.254885	3.975202	1.321768
ITEM5	4.590000	100	5.000000	6.000000	4.305502	3.916084	1.477747
ITEM6	4.550000	100	4.000000	6.000000	4.302523	4.048193	1.479660
ITEM7	4.650000	100	5.000000	6.000000	4.413615	4.091573	1.366075
ITEM8	4.780000	100	5.000000	6.000000	4.540550	4.231880	1.396822
ITEM9	4.670000	100	5.000000	6.000000	4.445617	4.207012	1.421729
ITEM10	4.450000	100	5.000000	5.000000	4.173693	3.793010	1.416889

Полный список этих статистик приведен в разделе *Стартовая панель - Описательные статистики*, стр. 1618.

Подгонка распределения

После щелчка мышью на нужное распределение в стартовой панели *Подгонка распределения* на экране появится диалоговое окно подгонки распределений (непрерывных или дискретных).





Доступные распределения обсуждались в разделе *Стартовая панель - Подгонка распределения*, стр. 1620.

Другие процедуры подгонки распределения. В *STATISTICA* имеется несколько других процедур для подгонки непрерывных распределений. Модуль *Анализ выживаемости* (том III) подгоняет распределения Вейбулла и Гомпертца к наблюдаемым временам выживаемости и отказов. Модуль позволяет также обрабатывать цензурированные наблюдения.

Гистограммы в *Статистических 2М графиках* (см. меню *Графики*) предлагают разнообразные возможности для построения простых, кумулятивных и категоризованных графиков частот распределений; доступные распределения включают: *Нормальное*, *Бета*, *Экстремального значения*, *Гамма*, *Геометрическое*, *Лапласа*, *Логистическое*, *Логнормальное*, *Пуассона*, *Релея* и *Вейбулла* (см. Приложение I для описания этих распределений). Вы можете также построить графики *Квантиль-квантиль* и *Вероятность-вероятность* этих распределений и оценить

Copyright © StatSoft, 1995

StatSoft®

качество подгонки к наблюдаемым данным. Эти опции детально описаны в томе II руководства.

Модуль *Анализ процессов* (см. том IV) содержит несколько опций для подгонки этих распределений и вычисления таблиц с ожидаемыми и наблюдаемыми частотами. Этот модуль также предлагает опцию, подогнать одновременно все эти непрерывные распределения к данным; таблица результатов содержит соответствующие значения статистики Колмогорова-Смирнова *d-max*, расположенные в порядке возрастания (от распределения с лучшим согласием к распределению с худшим согласием).

Переменная

Для подгонки распределения к значениям переменной вначале следует задать переменную. Нажмите для этого кнопку *Переменная*.

Распределение

Как только вы определили переменную (к значениям которой будет производиться подгонка), можно выбрать тип распределения: *непрерывное* или *дискретное* с помощью поля *Распределение*. Дискретное распределение приписывает ненулевую вероятность каждой точке, например, распределение Бернулли. Непрерывное распределение приписывает нулевую вероятность каждой точке, например нормальное распределение. Распределение задается с точностью до параметра. После выбора распределения вычисляются параметры, обеспечивающие наилучшую подгонку для соответствующего распределения. Эти параметры отображаются, по умолчанию, однако пользователь может ввести также собственные значения параметров распределения и проверить, насколько данные согласуются с этим распределением.

Число групп

Способом группировки можно управлять. Опция *Число групп* задает число групп в таблице частот или в гистограмме. Используйте поля *Нижняя граница* или *Верхняя граница,* чтобы задать нижнюю и верхнюю границу соответственно. Заметим, что там, где нужно, программа переопределит крайние интервалы в интервалы с бесконечной границей.

Критерий согласия *хи-квадрат*

По умолчанию, программа вычисляет критерий *хи-квадрат*, основанный на наблюдаемых и ожидаемых частотах. Группы, где ожидаемая частота меньше 5, объединяются в большие группы. Если критерий значим, гипотеза о соответствии наблюдаемых данных гипотетическому распределению отвергается.

Одновыборочный критерий согласия Колмогорова-Смирнова

Критерий можно вычислить либо по группированным значениям (используйте опцию *группированные*) либо из исходных данных (что более медленно; используйте опцию *непрерывное*).

Критерий Колмогорова-Смирнова для группированных данных. В этом случае программа вычисляет статистику *d-max* по табулированным данным. Если критерий Колмогорова-Смирнова значим, гипотеза, согласно которой наблюдения имеют определенное гипотетическое распределение, - отвергается (см. однако замечания о *вероятностях Лиллиефорса* ниже!).

Непрерывное распределение. В этом случае, программа отсортирует наблюдаемые данные и вычислит кумулятивную ожидаемую частоту в каждой точке (т.е. построит эмпирическую функцию распределения); статистика Колмогорова *d-max* равна

StatSoft®

Copyright © StatSoft, 1995

максимальной абсолютной разности между гипотетической функцией распределения и эмпирической функцией распределения.

Известные параметры, вероятности Лиллиефорса. Уровень значимости для статистики *d-max* Колмогорова-Смирнова основывается на значениях табулированных Massey (1951). Эти вероятности вычислены, когда параметры распределения известны заранее и не оцениваются. Однако обычно эти параметры неизвестны и оцениваются по наблюдениям. В таком случае проверяется сложная гипотеза (с неизвестными заранее параметрами распределения), например, гипотеза о том, что наблюдаемые величины имеют нормальное распределение и нужно использовать значения *Лиллиефорса* (Lilliefors, 1967).

Для других распределений следует с осторожностью подходить к стандартному уровню значимости, основанному на таблицах Massey (1951), потому что часто это влечет переоценку степени согласия распределения с данными (иными словами, стандартный *p* уровень может быть незначимым, когда, на самом деле, соответствующее распределение не обеспечивает хорошую подгонку, это связано с тем, что параметры распределения оцениваются по данным).

Если вам нужно оценить “поведение” статистики *d-max* для различных распределений с помощью методов Монте-Карло, воспользуйтесь программой на языке *STATISTICA BASIC Lillief.stb*, которая генерирует случайные последовательности Монте-Карло, и табулирует критические значения Лиллиефорса для нормального распределения. Программа может быть легко приспособлена и для других распределений (см. также том V руководства).

График

Если переменные еще не выбраны, нажмите кнопку *График* и выберите их. После этого снова нажмите кнопку, чтобы построить графики, задаваемые следующими опциями:

График распределения. Выберите опцию: график *Гистограмма* или *Кумулятивное распределение*.

График исходных частот или %. Если в опции выше выбрано *Гистограмма*, то можно построить или график *Исходных частот*, или график *Относительных частот (%)*.

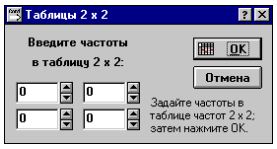
Copyright © StatSoft, 1995

StatSoft®

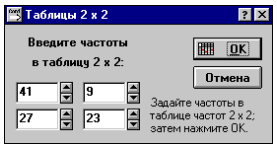
ПРИМЕРЫ

Пример 1: Таблица 2x2

После выбора опции *Таблицы 2x2: X^2/V^2/Фи^2, Макнемара, точный Фишера* из стартовой панели *Непараметрические статистики* появится диалоговое окно, в котором вводятся частоты в 4 клетки (ячейки) таблицы сопряженности 2 x 2.



Хи-квадрат, V-квадарат. Предположим, вы пытаетесь понять, имеет ли новая марка безалкогольного напитка успех. До того, как принять окончательное решение, вы проводите опрос, в котором просите респондентов: мужчин и женщин, выразить свое отношение к старому и новому напитку. Пусть из 50 мужчин 41 предпочитают новую марку; из 50 женщин ее предпочитают только 27. Введите эти данные:



Нажмите *OK,* чтобы получить таблицу с результатами.

Значение *хи-квадрат* для этих данных равно *9.01* (см. таблицу далее), что высоко значимо (см. *Элементарные понятия статистики*, глава 8, где дано объяснение статистической значимости). Таким образом, предпочтение мужчин значимо отличается от предпочтения женщин (очевидно, новая марка наиболее предпочтительна для мужчин). Статистика Пирсона хи-квадрат привлекательна тем, что ее распределение достаточно точно аппроксимируется распределением хи-квадрат. Однако в ряде случаев качество аппроксимации ухудшается, и используют скорректированные варианты хи-квадрат. Статистика *V-квадрат*– это статистика *хи-квадрат*, скорректированная на размер выборки (Кендалла and Stuart, 1979; см. также Rhoades and Overall, 1982). Если частоты в таблице относительно малы (например, в среднем меньше 10), следует больше полагаться на статистику *V-квадрат*, чем на статистику *хи-квадрат*.

Таблица 2x2 (10items.sta)

Далее...	Столб. 1	Столб. 2	Сумма по строкам
Частоты, строка 1	41	9	50
Процент от общего	41.000%	9.000%	50.000%
Частоты, строка 2	27	23	50
Процент от общего	27.000%	23.000%	50.000%
Сумма по столбцу	68	32	100
Процент от общего	68.000%	32.000%	
Хи-квадрат (ст.св.=1)	9.01	p= .0027	
V-квадрат (ст.св.=1)	8.92	p= .0028	
Поправка Йетса	7.77	p= .0053	
Фи коэффициент	.09007		
Фишера p, односторонний		p= .0025	
двусторонний		p= .0049	
Макнемара Хи-квадрат (A/D)	4.52	p= .0336	

Разумное объяснение статистики *V-квадрат* (и других поправок хи-квадрата для малых *n*) следующее: Представьте, вы наблюдаете значения 3 и 7 в одной строке таблицы 2 x 2. Тогда вероятность попадания наблюдения в первую ячейку может быть оценена как 30%, точно также вероятность попадания наблюдения во вторую ячейку как 70%. Однако эти вероятности могли быть оценены также как 34% и 66%, соответственно; наблюдаемые данные вовсе не противоречат этому. Другими словами, при малых *n* у вас имеется очевидная неуверенность в оценке вероятностей (и ожидаемых значений), - оценка неточна и *V-квадрат* делает поправку на это. Значение статистики *V-квадрат* всегда меньше *хи-квадрат*.

Фи-квадрат. Другой взгляд на эти данные состоит в том, что *Пол - Gender* зависим с *Предпочтением - Preference*. Эту корреляцию и принимает во внимание *фи-квадрат*.

StatSoft®

Copyright © StatSoft, 1995

Точный критерий Фишера. Пусть даны маргинальные частоты в таблице 2 на 2, (в данном примере это *50*, *50,* 68 и 32 – суммы частот в строках и столбцах). Предположим, что мужчины и женщины в популяции не различаются предпочтениями (факторы в таблице независимы). Зададимся вопросом: насколько вероятно получить частоты в ячейках большие или меньшие тех, что обнаружены в опросе? Для малых *n*, эта вероятность вычисляется *точно,* подсчетом всех возможных таблиц, которые можно построить по данным маргинальным частотам. Таким образом, критерий Фишера вычисляет *точную* вероятность появления при нулевой гипотезе (согласно которой переменные независимы) наблюдаемых частот в ячейке (или частот больших). Вычисляются односторонние и двухсторонние вероятности.

Макнемара *хи-квадрат*. Этот критерий применяется, когда частоты в таблице 2 на 2 получены из *зависимых* выборок. Например, значения переменной до и после эксперимента на одном и том же материале. Вы можете подсчитывать число студентов, имеющих минимальные успехи по математике в начале и конце семестра. Два значения *хи-квадрата* вычисляются: *A/D* и *B/C*. *A/D хи-квадрат* проверяет гипотезу о том, что частоты в ячейках *A* и *D* (верхняя левая, нижняя правая) одинаковы. *B/C хи-квадрат* проверяет гипотезу о равенстве частот в ячейках *B* и *C* (верхняя правая, нижняя левая).

Отметим, что *STATISTICA* имеет также два модуля, специально предназначенные для анализа частот и наблюдений в многомерных таблицах практически неограниченной размерности и сложности (модули *Основные статистики и таблицы* и *Логлинейный анализ*). Опция *Таблицы* модуля *Основные статистики и таблицы* вычисляет всесторонний набор описательных статистик. *Логлинейный анализ* выполняет полный логлинейный анализ многовходовых таблиц. *STATISTICA* содержит пример *STATISTICA BASIC* программы (файл *Manthaen.stb* в каталоге *STBASIC*) вычисления критерия Ментела-Хенцела для сравнения двух групп (см. Mantel and Haenszel, 1959; см. также Lee, 1992). Обратитесь к комментариям, включенным в файл *Manthaen.stb* за дополнительной информацией.

Пример 2: Критерий знаков

Критерий знаков – это непараметрическая альтернатива *t*-критерия для зависимых выборок (см. *Основные статистики и таблицы*, глава 9). Обсуждение логики критерия, предположений и сравнение с критерием Вилкоксона дано на стр. 1616.

Следующий пример основан на исследовании Додда (1979). При произнесении речи, большую роль играют визуальные сигналы; более точно, произносимые слова понимаются гораздо легче, когда видно лицо говорящего. В этом смысле все люди “читают по губам”, по крайней мере, за небольшим исключением. Додд пытался выяснить, осознают ли младенцы в возрасте 10 - 16 недель связь между произнесенными словами и движениями губ говорящего. Для этой цели Додд поместил младенцев в комнату, так что они могли наблюдать через окно за лицом человека, читающего речь. Эта речь передавалась в комнату или без задержки (*синхронно*) или с задержкой 400 миллисекунд (*асинхронно*). Зависимая переменная фиксировала время, которое ребенок наблюдал за лицом в окне (время внимания). Не формулировалось никаких определенных гипотез относительно условия, которое должно вызвать наибольшее внимание (асинхронная речь могла быть более интересной или внимание ребенка могло привлечь что-то другое, потому что, по самой идее опыта, лицо не выглядело источником речи).

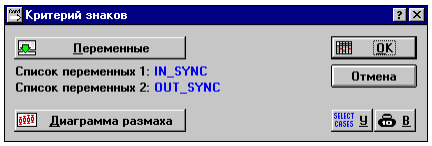
Copyright © StatSoft, 1995

Результаты исследования содержатся в файле *Synchron.sta* (две буквы, идентифицирующие ребенка, введены в имя переменной).

Данные: SYNCHR...

ЧИСЛОВЫЕ ЗНАЧЕНИЯ	Sensitivity of infants to vi	
	1 IN_SYNC	2 OUT_SYNC
DC	20.3	50.4
MK	17.0	87.0
VH	6.5	25.1
JM	25.0	28.5
SB	5.4	26.9
MM	29.2	36.6
RH	2.9	1.0
DJ	6.6	43.8
JD	15.8	44.2
ZC	8.3	10.4
CW	34.0	29.9
AF	8.0	27.7

Определение анализа. Критерий знаков - это непараметрическая альтернатива *t*-критерию для зависимых выборок. Так как один и тот же ребенок слушал речь в двух разных условиях (синхронно и асинхронно), то этот тест вполне здесь годится. Детали, касающиеся предположений и альтернативных процедур, описаны на стр.1616, 1616 и 1644. После запуска модуля *Непараметрическая статистика* откройте файл *Synchron.sta*. Далее выберите опцию *Критерий знаков* на стартовой панели. Можно выполнить тест для списка переменных (точнее, списка пар); однако, здесь имеются только 2 переменные. Поэтому нажмите кнопку *Переменные* и выберите переменные *In_sync* и *Out_sync*.



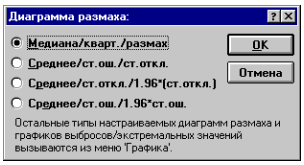
Нажмите *OK* для начала анализа. Ниже дана таблица с результатами.

Критерий знаков (synchron.sta)

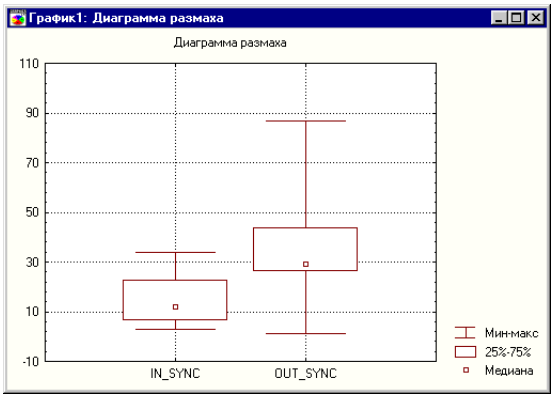
Далее...	Число несовп.	Процент v<V	Z	p-уров.
IN_SYNC & OUT_SYNC	12	83.33334	2.020726	.043308

p-уровень в таблице значим на уровне *.05* (обратитесь к главе 8, *Элементарные понятия статистики,* где обсуждается “статистическая значимость”). Поэтому вы можете заключить, что даже в столь раннем возрасте дети уже могут различать, синхронизированную с движением губ речь и несинхронизированную; слова “читай по моим губам”, оказывается, содержат большую долю правды о том, как учатся понимать язык в раннем возрасте.

По умолчанию для этого критерия принят график *Диаграмма размаха*. Вы можете построить его двумя способами: нажать на кнопку *Диаграмма размаха* в окне *Критерий знаков* или щелкнуть на таблицу результатов правой кнопкой мыши и выбрать в меню *Быстрые статистические графики* опцию *Диаграмма размаха*. Далее программа попросит определить переменные для отображения на графике. В этом примере выберите обе переменные. Затем определите тип статистики для графика в окне *Диаграмма размаха:* (см. ниже). Выберите опцию *Медиана/кварт./размах* и нажмите *OK*.



На диаграмме размаха для каждой переменной показаны: медиана, квартильный размах (25%, 75% процентили), размах (минимум, максимум).



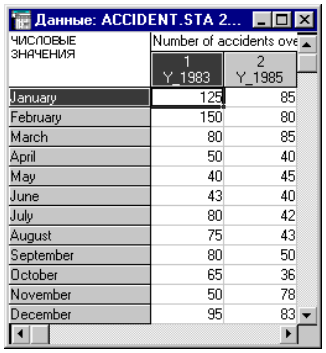
StatSoft

Copyright © StatSoft, 1995

График, очевидно, показывает, что переменная *In_sync* (измеряющее внимание к несинхронной речи) имеет более высокую медиану, а также больший разброс, чем переменная *Out_sync*.

Пример 3: Наблюдаемые значения в сравнении с ожидаемыми: *хи-квадрат*

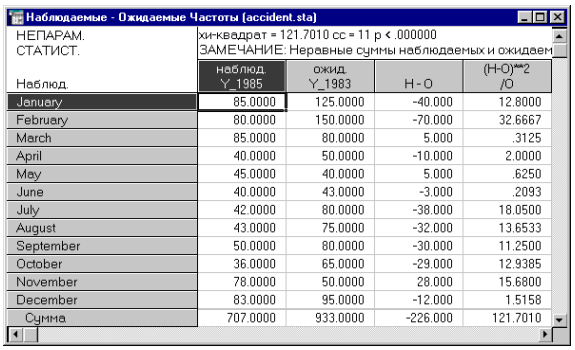
Следующий пример основан на данных (искусственных) об авариях на шоссе (данные содержатся в файле *Accident.sta*). Данные записывались каждый месяц в 1983 и 1985 году.



Представьте, что были потрачены значительные средства в 1984 году, чтобы улучшить безопасность движения на этом шоссе. Если затраченные средства ни к чему не привели (нулевая гипотеза), то число несчастных случаев в 1985 году могло бы прогнозироваться на том же уровне, что и в 1983 (при условии, что общее число машин на трассе и интенсивность движения не менялись). Таким образом, данные за 1985 год рассматриваются здесь как ожидаемые значения, данные за 1983 год, как наблюдаемые.

Задание анализа. После запуска модуля *Непараметрическая статистика и распределения* откройте файл *Accident.sta* и выберите в стартовой панели опцию *Наблюдаемые в сравнении ожидаемыми*. В появившемся диалоговом окне *Наблюдаемые и ожидаемые частоты* нажмите кнопку *Переменные* и выберите *Y_1983* - переменную с наблюдаемыми частотами и *Y_1985* – переменную с ожидаемыми частотами.

После нажатия *OK,* таблица с результатами появится на экране.



Из таблицы ясно видно, что снижение числа аварий в 1985 по сравнению с 1983 высоко значимо.

Заметим, что в нижней части таблицы результатов показано общее число аварий за каждый год (*Сумма*); разности между наблюдаемыми и ожидаемыми значениями даны в третьем столбце, квадраты разностей, деленные на ожидаемые значения (слагаемые *хи-квадрат*) – в четвертом столбце.

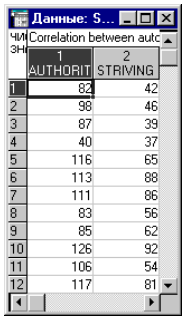
Пример 4: Корреляции

Обсуждение и сравнение различных непараметрических корреляций (Спирмена *R*, *Кендалла tau* и *Гамма*), доступных в модуле *Непараметрическая статистика и распределения,* дано на стр. 1624. Следующий

Copyright © StatSoft, 1995

пример основывается на данных (файл *Striving.sta*), представленных Siegel and Castellan (1988, стр. 238).

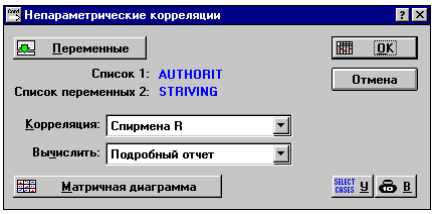
Двенадцать студентов ответили на вопрос анкеты, чтобы измерить две переменные (1) авторитарность и (2) борьба за социальный статус. Авторитарность (Adorno и др., 1950) - психологическая концепция, состоящая, грубо говоря, в том, что властные люди имеют тенденцию считать, что власть должна быть жесткой, и ей следует подчиняться (иными словами придерживаются принципа: “закон и порядок”).



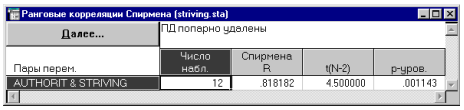
Цель исследования состояла в том, чтобы выяснить, зависимы, в действительности, эти две переменные или нет.

Задание анализа. После запуска модуля *Непараметрическая статистика и распределения,* откройте файл *Accident.sta* и выберите в стартовой панели опцию *Корреляции (Спирмена, тау Кендалла, гамма*). В появившемся диалоговом окне нажмите кнопку переменные и выберите *Authorit* как первую переменную, *Striving* как вторую переменную.

Модуль *Непараметрическая статистика и распределения* вычисляет также корреляционные матрицы. В этом примере выберите просто *Спирмена R* и *Подробный отчет*.



После нажатия *OK* таблица с результатами появится на экране.

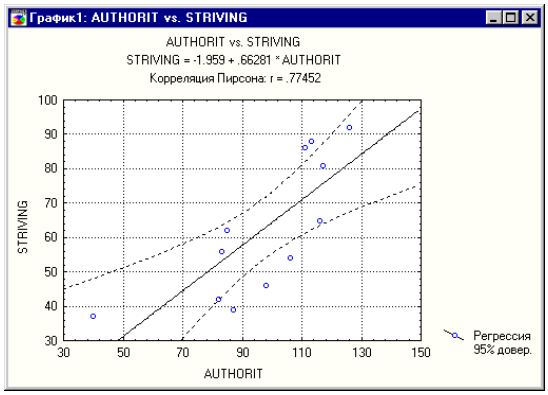


Вы видите, что корреляция между этими двумя шкалами высоко значима (обратитесь к главе 8, *Элементарные понятия статистики,* где обсуждается “статистическая значимость”) и вы можете сделать вывод, что индивидуумы, имеющие внутреннюю установку на авторитаризм, вероятно, также стремятся к борьбе за свой статус в обществе (при условии, что анкета адекватно данному исследованию), тем самым, подтверждается концепция Адорно. От себя добавим, что авторитаризм - внутренняя установка (ее трудно непосредственно измерить), тогда как борьба за положение в обществе наблюдается отчетливо.

Вы можете визуализировать найденную зависимость двумя способами. Либо нажав кнопку *Матричная диаграмма* в диалоговом окне *Непараметрические корреляции* (после того как выбрали переменные), либо щелкнув правой кнопкой мыши на таблице результатов и выбрав опцию *Диаграмма рассеяния/довер* из меню *Быстрые статистические графики.*

Параметрическая корреляция (*r* Пирсона) между шкалами (*r* = *.77*) показана в заголовке графика (см. ниже). Интересно, что эта корреляция меньше ранговой корреляции Спирмена (Спирмена *R* равно *.82*).

StatSoft
Copyright © StatSoft, 1995



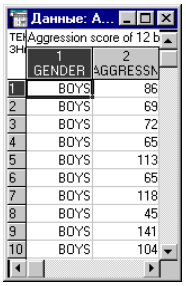
Если бы в этом примере размер выборки был больше, то можно было бы сделать вывод, что рассмотрение рангов (а не самих наблюдений) в действительности улучшает оценку зависимости между переменными, так как “подавляет” случайную изменчивость и уменьшает воздействия выбросов.

Статистики Кендалла *tau* и *Гамма*. Для сравнения, вернитесь в окно *Непараметрические корреляции* и выберите опцию *Статистика tau Кендалла* а также опцию *Гамма*. Обе статистики *Кендалла tau* и *Гамма* будут вычислены и окажутся равными *.67*.

Как объяснялось ранее, (см. описание стартовой панели на стр. 1624), эти статистики тесно связаны между собой, но отличаются от статистики Спирмена. Статистику Спирмена *R* можно представлять себе как вычисленную по рангам корреляцию Пирсона, т.е. в терминах доли изменения одной величины, связанной с изменением другой. Статистики *Кендалла tau* и *Гамма* скорее оценивают *вероятности,* точнее, разность между вероятностью того, что наблюдаемые значения переменных имеют один и тот же порядок и вероятностью того, что порядок различный.

Пример 5: Критерий серий Вальда-Вольфовица, Манна-Уитни *U* критерий, двухвыборочный критерий Колмогорова-Смирнова.

Все эти критерии представляют собой альтернативы *t-критерию* для независимых выборок (см. *Основные статистики и таблицы*, глава 9). Обратитесь к стр. 1614 и 1615 за обсуждением логики и предположений этих критериев. Пример основан на исследовании агрессивности четырехлетних мальчиков и девочек (Siegel, 1956, стр. 138). Данные содержатся в файле *Aggressn.sta*.



Двенадцать мальчиков и двенадцать девочек наблюдались в течение 15-минутной игры; агрессивность каждого ребенка оценивалась в баллах (в терминах частоты и степени проявления агрессивности) и суммировалась в один индекс агрессивности, который вычислялся для каждого ребенка.

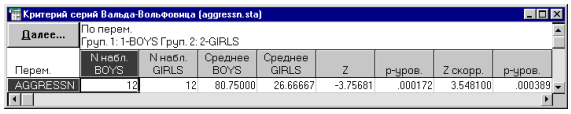
Задание анализа. После запуска модуля *Непараметрическая статистика и распределения* и выберите в стартовой панели или опцию *Критерий серий Вальда-Вольфовица,* или *U-критерий Манна-Уитни,* или *Двухвыборочный критерий Колмогорова-Смирнова* (диалоговые окна по существу одни и те же для этих трех тестов). Далее нажмите

Copyright © StatSoft, 1995

StatSoft®

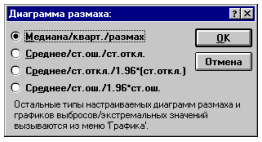
кнопку *Переменные* и выберите переменную *Пол - Gender* как независимую и переменную *Aggressn* как зависимую. Коды для однозначного отнесения каждого наблюдения к определенному полу будут автоматически выбраны программой.

Далее нажмите *OK,* чтобы выполнить анализ.

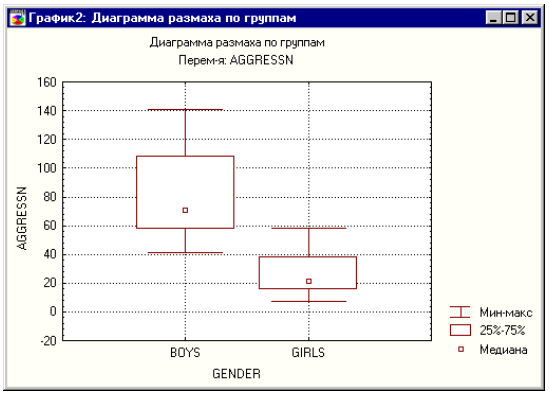


Как видно из таблицы результатов, разница агрессивности мальчиков и девочек в этом исследовании высоко значима (обратитесь к главе 8, *Элементарные понятия статистики* где обсуждается понятие “статистическая значимость”), независимо от того, какой критерий используется.

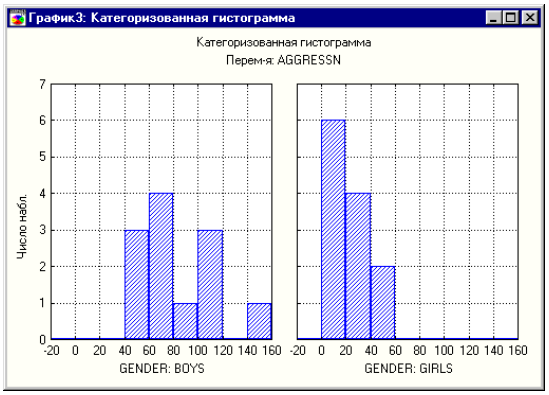
График по умолчанию для этих тестов – диаграмма размаха. Вы можете построить его двумя способами: нажав кнопку *Диаграмма размаха* в окне *Критерий знаков* или щелкнув на таблице результатов правой кнопкой мыши и выбрав затем опцию *Диаграмма размаха* в меню *Быстрые статистические графики*. Далее программа попросит выбрать переменные. В этом примере выберите обе переменные. Затем выберите тип графика в окне *Диаграмма размаха:* (см. ниже). Выберите *Медиана/кварт./размах* и нажмите *OK*.



На диаграмме размаха для каждой переменной показаны: медиана, квартильный размах (25%, 75% процентили), размах (минимум, максимум).



Из этого графика (а также следующего графика) отчетливо видно, что мальчики более агрессивны, чем девочки. Для того чтобы увидеть распределение зависимой переменной, разбитой на группы, нажмите кнопку *Категоризованная гистограмма*.

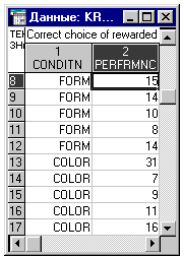


Пример 6: Ранговый дисперсионный анализ Краскела-Уоллиса и медианный тест

Эти тесты - альтернативны однофакторной межгрупповой ANOVA (см. *ANOVA/MANOVA*,

StatSoft®
Copyright © StatSoft, 1995

глава 13). Обратитесь к стр. 1615 за обсуждением логики и предположений этих тестов. Пример основан на (искусственных) данных, представленных в Hays (1981, стр. 592). Данные содержатся в файле *Kruskal.sta* и показаны ниже.



Эти данные получены в исследовании маленьких детей, которые случайным образом приписывались к одной из трех экспериментальных групп. Каждому ребенку предлагалась серия парных тестов. Задача ребенка состояла в том, чтобы сделать правильный выбор и получить вознаграждение. В первой группе тестом была форма (группа *1-Форма - 1-Form*), во второй - цвет (группа *2-Цвет - 2-Color*), в третьей - размер (*3-Размер - 3-Size*) предмета. Зависимая переменная – число испытаний, которые требовались каждому ребенку, чтобы получить вознаграждение

Задание анализа. После запуска модуля *Непараметрическая статистика и распределения* и выбора файла *Kruskal.sta,* выберите опцию *ANOVA Краскела-Уоллиса и медианный тест*, чтобы открыть диалоговое окно *Дисперсионный анализ Краскела-Уоллиса и медианный тест*.

Далее нажмите кнопку *Переменные* и выберите переменную *Conditn* как независимую и переменную *Perfrmnc* как зависимую. Нажмите кнопку *Коды* и выберите все коды для независимой переменной (нажмите кнопку *Все*). Диалоговое окно *Дисперсионный анализ Краскела-Уоллиса и медианный тест* появится на экране:

Результаты критерия Краскела-Уоллиса. В диалоговом окне нажмите *OK* и начните анализ. Результаты ранговой ANOVA Краскела-Уоллиса будут показаны в первой таблице результатов, результаты медианного теста – во второй.

Вы видите, что критерий Краскела-Уоллиса высоко значим. Таким образом, характеристики различных экспериментальных групп значимо отличаются друг от друга. Напомним, что процедура Краскела-Уоллиса, по существу, является дисперсионным анализом, основанным на рангах. Суммы рангов (для каждой группы) показаны в правом столбце таблицы результатов. Наибольшая ранговая сумма (самое худшее выполнение теста) относится к *Размеру - Size* (это тот параметр, который надо различить, чтобы получить вознаграждение). Наименьшая ранговая сумма (лучшее выполнение) относится к *Форме - Form*.

Результаты медианного теста. Медианный тест появится сразу, как только вы закроете таблицу результатов критерия Краскела-Уоллиса.

Copyright © StatSoft, 1995

StatSoft

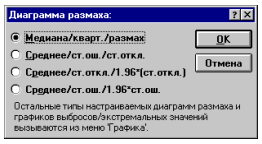
Медианный тест, общ. медиана= 19.50000 (kruskal.sta)

НЕПАРАМ. СТАТИСТ.
Зависимые: PERFRMNC
Независимая переменная: CONDITN
хи-квадрат = 8.666667, сс = 2, p = .0131

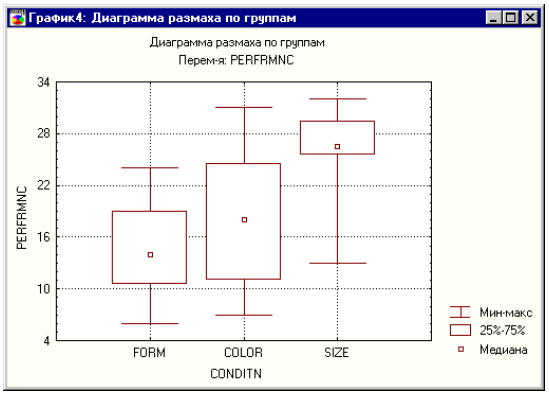
	FORM	COLOR	SIZE	Общий
<= Медиана: набл.	9.00000	7.00000	2.00000	18.00000
ожидаемые	6.00000	6.00000	6.00000	
набл.-ожид.	3.00000	1.00000	-4.00000	
> Медиана: набл.	3.00000	5.00000	10.00000	18.00000
ожидаемые	6.00000	6.00000	6.00000	
набл.-ожид.	-3.00000	-1.00000	4.00000	
Сумма: набл.	12.00000	12.00000	12.00000	36.00000

Медианный критерий также значим, однако, в меньшей степени. Напомним, что медианный критерий более “грубый” и менее чувствительный, чем критерий Краскела-Уоллиса (см. стр. 1615). В таблице результатов показано число наблюдений (детей) в каждой экспериментальной группе, которые лежат ниже (или равны) общей медианы и число наблюдений, лежащих выше общей медианы. Снова, наибольшее число испытуемых с числом попыток (до получения вознаграждения) выше общей медианы относятся к группе *Размер - Size.* Больше всего испытуемых с числом попыток ниже медианы относятся к группе *Форма - Form*. Таким образом, медианный тест подтверждает, что форма предмета наиболее легко различается детьми, тогда как размер различается хуже всего.

Графическое представление результатов. График по умолчанию для этих тестов – диаграмма размаха. Его можно построить двумя способами: нажав кнопку *Диаграмма размаха* в окне *Дисперсионный анализ Краскела-Уоллиса и медианный тест* или щелкнув на таблице результатов правой кнопкой мыши и выбрав опцию *Диаграмма размаха* в меню *Быстрые статистические графики*. Далее программа попросит выбрать переменные для графика. В этом примере выберите обе переменные. Затем выберите тип статистики для графика в окне *Диаграмма размаха:* (см. ниже). Выберите опцию *Медиана/кварт./размах* и нажмите *OK*.

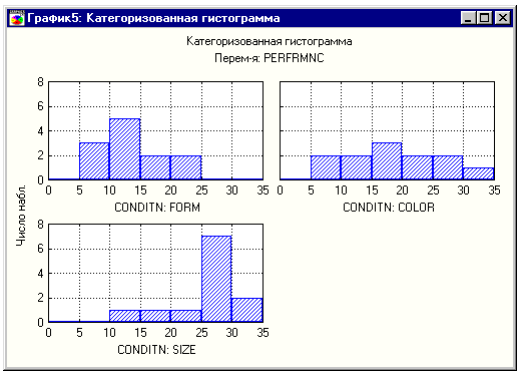


На диаграмме размаха для каждой переменной показаны: медиана, квартильный размах (25%, 75% процентиль), размах (минимум, максимум).



Снова ясно видно, выполнение теста *Форма - Form* было лучше любого другого; медиана числа испытаний при этом условии ниже, чем при любом другом.

Для того чтобы увидеть распределение зависимой переменной, разбитой на группы, нажмите кнопку *Категоризованная гистограмма*. Этот график снова подтверждает, что в группе *Форма - Form* выполнение “лучше” (распределение слегка скошено влево), чем при других условиях. Самое худшее выполнение, как отчетливо видно из графиков, для группы *Размер - Size*.



Отсюда также можно заключить, что наиболее легко дети различают *Форму - Form*.

StatSoft®

Copyright © StatSoft, 1995

Пример 7: Дисперсионный анализ Фридмана и конкордация Кендалла

Фридмана ранговая ANOVA альтернативна дисперсионному анализу с одним внутригрупповым фактором с повторными измерениями (см. *ANOVA/MANOVA*, глава 13). Этот критерий сравнивает переменные в зависимых выборках. Коэффициент конкордации Кендалла, по существу, представляет усредненную ранговую корреляцию (см. стр. 1617).

Данные: MOTHERS.STA 20п * 13н

ЧИСЛОВЫЕ ЗНАЧЕНИЯ	Ranking of mothers by 13 judges					
	1 MOTH_1	2 MOTH_2	3 MOTH_3	4 MOTH_4	5 MOTH_5	6 MOTH_6
H. Smith	1.000	2.000	3.000	4.000	5.000	6.000
B. Jones	5.000	1.000	16.000	8.000	9.000	2.000
T. Hill	3.000	2.000	7.000	5.000	14.000	9.000
W. Miller	8.000	3.000	10.000	11.000	4.000	2.000
B. Robertson	2.000	1.000	15.000	8.000	14.000	4.000
C. Hansson	16.000	17.000	5.000	13.000	15.000	11.000
S. Mann	12.000	9.000	14.000	6.000	7.000	2.000
F. Jordan	11.000	2.000	13.000	10.000	7.000	3.000
B. Frank	9.000	2.000	15.000	6.000	5.000	7.000
W. Weiner	2.000	4.000	16.000	3.000	10.000	6.000
W. Kink	11.000	14.000	12.000	8.000	7.000	2.000
R. Clark	8.000	1.000	13.000	3.000	5.000	2.000
L. Evans	5.000	3.000	13.000	2.000	8.000	1.000

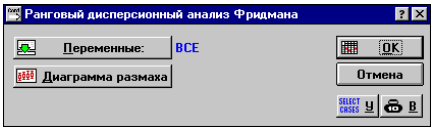
Следующий пример основан на данных, представленных Siegel (1956, стр. 233). 20 матерей и их глухих детей присутствовали на семинаре, обучающему уходу за детьми с данным видом дефекта. В конце семинара 13 экспертов ранжировали 20 матерей в соответствии с тем, насколько, по их мнению, они будут способствовать развитию своих детей. Данные содержатся в файле *Mothers.sta*.

Цель анализа. Анализ преследовал следующие цели:

(1) Определить, можно ли на основании оценок экспертов сделать вывод о значимых различиях между матерями относительно их способностей воспитывать детей с данным дефектом. Этот вопрос может быть решен с помощью рангового дисперсионного анализа (ANOVA) Фридмана.

(2) Можно ли доверять экспертам? Иными словами, согласованы ли их рейтинги, выставленные матерям (зависимы ли они)? Если нет, то вы, очевидно, не можете доверять их оценкам относительно способностей матерей. Эту гипотезу (эксперты согласованы в большей степени, чем можно бы ожидать из-за чисто случайных совпадений их мнений) можно проверить с помощью коэффициента конкордации Кендалла.

Задание анализа. После запуска модуля *Непараметрическая статистика и распределения* и выбора файла *Kruskal.sta,* выберите опцию *ANOVA Фридмана и конкордация Кендалла*. Нажмите кнопку *Переменные* и примите к обработке все 20 переменных (матерей).



Теперь нажмите *OK,* таблица с результатами появится на экране. Как вы видите, между матерями имеются высоко значимые различия. Дополнительно также видно, что эксперты, выставившие оценки, согласованы друг с другом - конкордация Кендалла равна *.57* (среднее ранговых корреляций равно *.53*).

Фридмана ANOVA и конкордация Кендалла (mothers.sta)

Далее...

ANOVA хи-квад. (N = 13, cc = 19) = 139.9814 p < .00000
Коэффиц. конкордации = .56673 Сред. ранг r = .53062

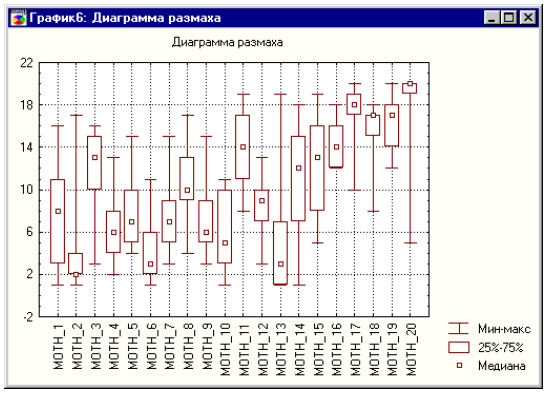
Перем.	Средний ранг	Сумма рангов	Среднее	Ст. откл.
MOTH_1	7.26923	94.5000	7.15385	4.597937
MOTH_2	4.69231	61.0000	4.69231	5.266001
MOTH_3	11.92308	155.0000	11.69231	4.230536
MOTH_4	6.69231	87.0000	6.69231	3.351234
MOTH_5	8.57692	111.5000	8.46154	3.755338
MOTH_6	4.38462	57.0000	4.38462	3.150092
MOTH_7	8.00000	104.0000	7.92308	3.988766
MOTH_8	11.07692	144.0000	10.92308	3.475187

График по умолчанию для этих таблиц результатов – диаграмма размаха. Его можно построить двумя способами: нажав кнопку *Диаграмма размаха* в окне *Ранговый дисперсионный анализ Фридмана* или щелкнув на таблице результатов правой кнопкой мыши и

Copyright © StatSoft, 1995

StatSoft

выбрав опцию *Диаграмма размаха* в меню *Быстрые статистические графики*. Далее программа попросит выбрать переменные для графика. В этом примере выберите все 20 переменных. Затем выберите тип статистики для графика в окне *Диаграмма размаха:* (см. ниже). Выберите опцию *Медиана/кварт./размах* и нажмите *OK*.



Если бы вы знали имена матерей, принявших участие в семинаре, то, используя график, могли бы сказать, кто, по мнению экспертов, наиболее успешно будет воспитывать своего глухого ребенка.

Пример 8: Критерий Вилкоксона парных сравнений

Критерий Вилкоксона парных сравнений представляет собой непараметрическую альтернативу *t*-критерию для зависимых выборок (см. *Основные статистики и таблицы*, глава 9). Он тестирует гипотезу о том, что значения двух переменных имеют одно и то же распределение. Обсуждение критерия и предположений дано на стр. 1616.

В примере использованы те же данные, что и для критерия знаков, основанные на исследовании Додда (1979). При произнесении речи, важную роль (для привлечения большого внимания) играют визуальные сигналы. Точнее, произносимые слова понимаются гораздо легче, когда видно лицо говорящего. Иными словами, между движением губ и словами существует, пусть мало осознанная (или забытая) связь. Поэтому можно сказать, что все люди “читают по губам”, по крайней мере, за небольшим исключением. Додд пытался выяснить, осознают ли младенцы в возрасте 10 - 16 недель зависимость между произнесенными словами и соответствующими движениями губ. Для этой цели Додд поместил младенцев в комнату, так что они могли наблюдать через окно за лицом человека, читающего речь. Эта речь передавалась в комнату двумя способами: или без задержки (*синхронно*) или с задержкой 400 миллисекунд (*асинхронно*). Зависимая переменная фиксировала время, которое ребенок наблюдал лицо за окном (время внимания). Не формулировалось никаких определенных гипотез относительно условия, вызывающего наибольшее внимание (асинхронная речь могла быть более интересной, или внимание могло привлечь что-то другое, потому что, по самой постановке опыта, лицо не выглядело источником речи).

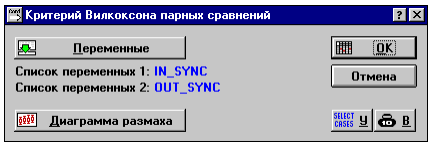
Результаты исследования содержатся в файле *Synchron.sta* (две буквы, идентифицирующие ребенка, введены в имя наблюдения).

Задание анализа. Критерий Вилкоксона парных сравнений - это непараметрическая альтернатива *t*-критерия для зависимых выборок. Так как один и тот же ребенок слушал речь в двух разных условиях (синхронно и асинхронно), то этот тест вполне подходит здесь. Детали, касающиеся предположений, и альтернативные процедуры даны на стр.1616, 1616 и 1644. После запуска модуля *Непараметрическая статистика и распределения* откройте файл *Synchron.sta*. Далее выберите опцию *Критерий Вилкоксона*. В принципе, можно выполнить тест для списка переменных (точнее, списка пар); однако в данном примере мы имеем только 2 переменные.

StatSoft®

Copyright © StatSoft, 1995

Нажмите кнопку *Переменные* и выберите переменные *In_sync* и *Out_sync*.



Далее нажмите *OK*, таблица с результатами будет показана на экране.

Результаты. Из таблицы вы видите, что критерий Вилкоксона значим на уровне *.01* (обратитесь к главе 8, *Элементарные понятия статистики* где обсуждается концепция "статистическая значимость"). Это больше, чем значимость критерия знаков для тех же данных (см. стр. 1602 для сравнения этих тестов), что, конечно, объясняется тем, что критерий знаков использует информацию только о знаках разностей, тогда как критерий Вилкоксона использует также относительную величину разностей и, таким образом, более чувствителен. Из результатов можно сделать вывод, что уже в столь раннем возрасте дети могут различать речь, синхронизированную с движением губ и несинхронизированную; слова "читай по моим губам", оказывается, содержат большую долю истины о том, как дети учатся понимать язык

График по умолчанию для этого теста – диаграмма размаха. Его можно построить двумя способами: нажав кнопку *Диаграмма размаха* в окне *Критерий Вилкоксона* или щелкнув на таблице результатов правой кнопкой мыши и выбрав опцию *Диаграмма размаха* в меню *Быстрые статистические графики*. Далее программа попросит выбрать переменные для графика. В этом примере выберите обе переменные. Затем выберите тип статистики для графика в окне *Диаграмма размаха:* (см. ниже). Выберите опцию *Медиана/кварт./размах* и нажмите *OK*.

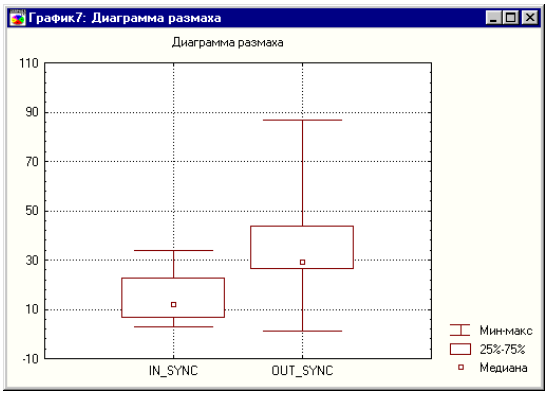


График ясно показывает, что переменная Out_sync (измеряющее внимание к несинхронной речи) имеет большее среднее, а также большую изменчивость, чем переменная In_sync.

Пример 9: Q критерий Кохрена

Критерий Кохрена *Q* - это развитие *хи-квадрат* теста Макнемара (см. стр. 1623) на случай более двух зависимых выборок. Критерий также может рассматриваться как альтернатива однофакторной ANOVA с повторными измерениями (см. *ANOVA/MANOVA*, глава 13) в ситуации, когда зависимая переменная является дихотомией (принимает два значения). Критерий также применим, когда хотят тестировать изменения частот в различные моменты времени в одной и той же выборке. Логика критерия и предположения дана на стр. 1630.

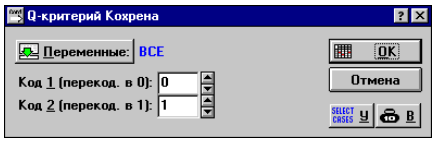
Следующий пример основан на (искусственных) данных Siegel and Castellan (1988, стр. 173). Положим, вы интересуетесь эффектом, какое влияние оказывает стиль интервьюера на число респондентов, согласившихся дать ответы на деликатные вопросы в очном опросе. Заметим, что всегда трудно получить ответы на такие деликатные вопросы, как, например, финансовое

Copyright © StatSoft, 1995

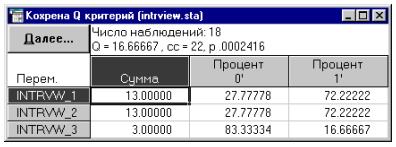
StatSoft

положение или здоровье в таких опросах, и возможно, вы захотите выяснить, повышает или нет, стиль задания вопросов интенсивность ответов (т.е. уменьшается ли количество отказавшихся отвечать). Для этой цели, можно научить интервьюера проводить интервью в одном из следующих стилей (1) восторженно, дружелюбно и заинтересовано (*Interview 1*), (2) очень сдержанно и формально (*Interview 2*), (3) беспристрастно и резко (*Interview 3*). Затем вы могли бы выбрать 18 троек семей, которые были ранее осторожны в своих ответах, и случайным образом приписать три семейства в каждом наборе к одному из трех стилей, в каком с ними будет проводиться интервью. Зависимая переменная определяет удачу (*1-Да - 1-Yes*) или неудачу интервью (*0-Нет - 0-No*). Результаты исследования записаны в файл *Interview.sta*.

Задание анализа. После запуска модуля *Непараметрическая статистика и распределения*, откройте файл *Interview.sta* и затем выберите опцию *Q-критерий Кохрена*. В появившемся диалоговом окне задайте сравниваемые переменные (нажмите кнопку *Переменные*). В данном случае выберите все 3 переменные (*Intrvw_1*, *Intrvw_2*, and *Intrvw_3*).

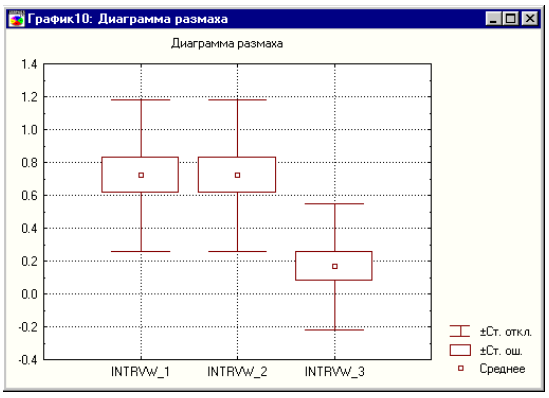


Нажмите *OK*, чтобы выполнить анализ.



Перем.	Сумма	Процент 0'	Процент 1'
INTRVW_1	13.00000	27.77778	72.22222
INTRVW_2	13.00000	27.77778	72.22222
INTRVW_3	3.00000	83.33334	16.66667

Результаты. Как видно из таблицы, *Q* критерий высоко значим. Самая правая колонка показывает процент семей, которые раскрыли свою личную информацию (*Процент 1'*). Интересно отметить, что не обнаружено различия в эффективности между первым стилем: восторженно, дружелюбно и заинтересованно (переменная *Intrvw_1*) и вторым: (2) очень сдержанно и формально (*Intrvw_2*), таким образом, выражение интереса интервьюером, кажется, не имеет значения. Однако число отказавшихся отвечать увеличилось, когда использовалось интервью третьего типа: беспристрастно и резко (*Intrvw_3*). Быстрая визуализация результатов может быть получена с помощью графика, принятого по умолчанию для этой таблицы результатов. Выберите из меню правой кнопки мыши *Быстрые статистические графики* опцию *Диаграмма размаха*. После выбора переменных для графика, выберите *Среднее/ст.ош./ст.откл.* из диалогового окна *Диаграмма размаха:*, чтобы построить следующий график.



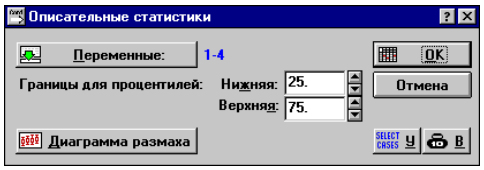
Пример 10: Описательные статистики

Опция предоставляет широкий набор описательных статистик, которые могут быть особенно интересны, если подозревается, что некоторые переменные распределены не нормально, и возможно, измерены в порядковой шкале. Обсуждение и вычисление этих статистик можно найти на стр. 1618.

StatSoft®
Copyright © StatSoft, 1995

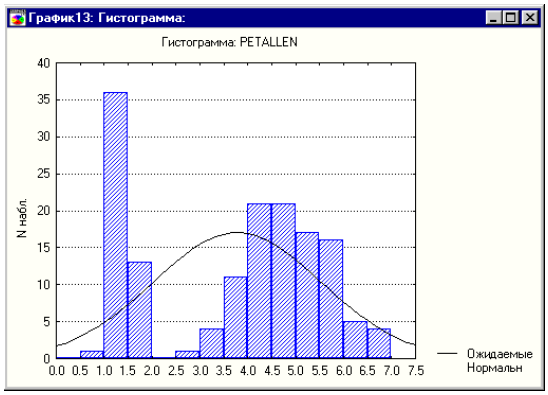
В этом примере используется файл *Irisdat.sta*. Файл содержит данные, описанные Фишером (1936). Он содержит длину и ширину чашелистиков (*Sepallen, Sepalwid*) и лепестков *(Petallen, Petalwid*) 50 цветков трех типов ириса.

Задание анализа. После запуска модуля *Непараметрическая статистика и распределения*, откройте файл *Irisdat.sta*. Затем выберите опцию *Обычные описательные статистики* в стартовой панели и в появившемся окне *Описательные статистики* нажмите кнопку *Переменные*. Переменная *5 - Iristype* - это кодовая переменная для определения типа ириса, поэтому выберите только переменные *1-4.*



Нажмите *OK,* чтобы получить таблицу с результатами.

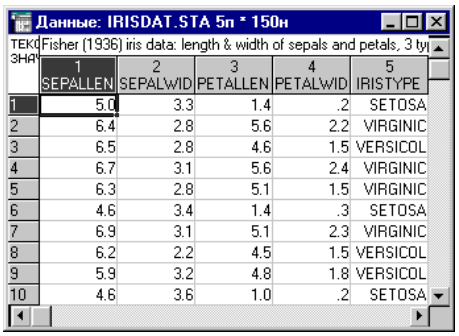
График по умолчанию для этой таблицы результатов - гистограмма частот с ожидаемой нормальной кривой.



Этот график очень полезен для определения отклонений от нормального распределения. Для того чтобы построить график для переменной *Petallen*, щелкните на таблице результатов правой кнопкой мыши и выберите опцию *Гистограмма/нормальн.* Гистограмма показывает, что, на самом деле, переменная *Petallen* является мультимодальной (как отмечено в таблице результатов); это возможно, если одна из наблюдаемых групп образует кластер в левой части графика.

Пример 11: Подгонка распределения

Опция подгонка распределения позволяет оценить степень согласия наблюдаемых данных с некоторым теоретическим распределением. Обратитесь к стр. 1620 за описанием доступных распределений. Заметим, модуль *Анализ выживаемости* содержит специальные программы для подгонки цензурированных (неполных) данных типа времен выживания и отказов к распределениям Вейбулла и Гомпертца. В этом примере используется файл *Irisdat.sta* (показан ниже). Файл содержит данные, представленные Фишером (1936), включает длину и ширину чашелистиков (*Sepallen, Sepalwid*) и лепестков *(Petallen, Petalwid*) 50 цветков трех типов ириса. Дискриминантный анализ этих данных описан также в главе *Дискриминантный анализ* (том III).

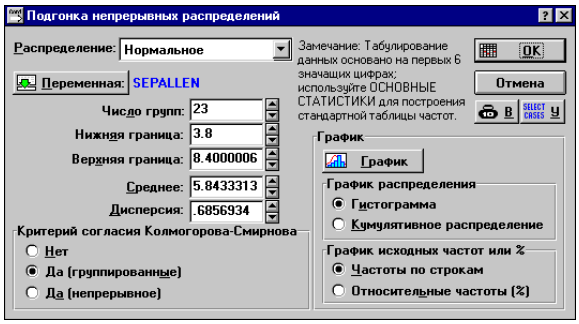


Copyright © StatSoft, 1995

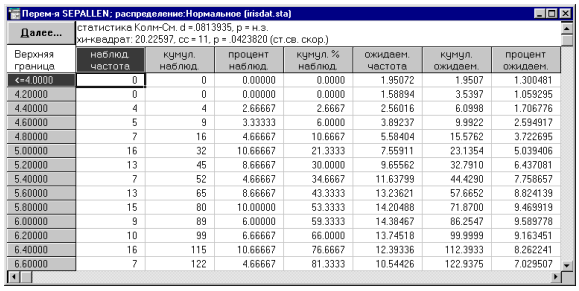
StatSoft®

Далее оцениваются распределения четырех переменных, описывающих длину и ширину лепестков и чашелистиков. Ожидается, что распределения будут нормальными.

Задание анализа. После запуска модуля *Непараметрическая статистика и распределения*, откройте файл *Irisdat.sta*. и выберите опцию *Непрерывные Распределения: Нормальное*. В появившемся диалоговом окне нажмите кнопку *Переменные* и выберите переменную *Sepallen*. В этот момент данные будут обработаны и в диалоговом окне будут показаны оценки среднего и дисперсии. Можно задать *Число групп* и *Нижнюю границу* и *Верхнюю границу* для вычисления частот. Диалоговое окно *Подгонка непрерывных распределений* теперь выглядит следующим образом.

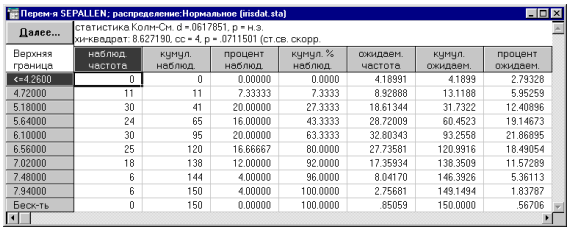


Сделайте все показанные в окне установки и нажмите *OK,* чтобы вычислить частоты.



Верхняя граница	наблюд. частота	кумул. наблюд.	процент наблюд.	кумул. % наблюд.	ожидаем. частота	кумул. ожидаем.	процент ожидаем.
<=4.0000	0	0	0.00000	0.0000	1.95072	1.9507	1.300481
4.20000	0	0	0.00000	0.0000	1.58894	3.5397	1.059295
4.40000	4	4	2.66667	2.6667	2.56016	6.0998	1.706776
4.60000	5	9	3.33333	6.0000	3.89237	9.9922	2.594917
4.80000	7	16	4.66667	10.6667	5.58404	15.5762	3.722695
5.00000	16	32	10.66667	21.3333	7.55911	23.1354	5.039406
5.20000	13	45	8.66667	30.0000	9.65562	32.7910	6.437081
5.40000	7	52	4.66667	34.6667	11.63799	44.4290	7.758657
5.60000	13	65	8.66667	43.3333	13.23621	57.6652	8.824139
5.80000	15	80	10.00000	53.3333	14.20488	71.8700	9.469919
6.00000	9	89	6.00000	59.3333	14.38467	86.2547	9.589778
6.20000	10	99	6.66667	66.0000	13.74518	99.9999	9.163451
6.40000	16	115	10.66667	76.6667	12.39336	112.3933	8.262241
6.60000	7	122	4.66667	81.3333	10.54426	122.9375	7.029507

Значения статистик. *Хи-квадрат* значим на уровне *.05* (*p = .042*). Таким образом, основываясь на критерии *хи-квадрат*, можно заключить, что распределение значимо отклоняется от нормального распределения. Однако критерий Колмогорова-Смирнова не значим (тем не менее, см. замечание в разделе *Диалоговые окна, опции, статистики* о вероятностях Лиллиефорса, стр. 1633). Результат этого примера не является необычным, потому что критерий Колмогорова-Смирнова не является *точной* процедурой, а, скорее, методом обнаружения больших (грубых) отклонений от гипотетического распределения. Часто значение статистики *хи-квадрат* сильно зависит от способа группировки, иными словами, от числа групп, минимального и максимального значения, которые задает пользователь. Например, если вы разобьете распределение *Sepallen* только на 10 групп (установите *Число групп 10*), а не на *23*, то получите значение *хи-квадрат,* значимое только на уровне *.07*.

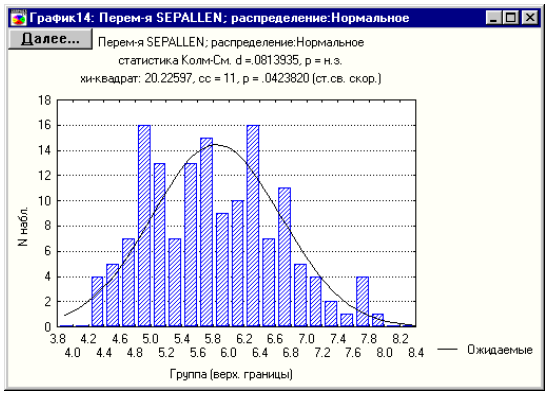


Верхняя граница	наблюд. частота	кумул. наблюд.	процент наблюд.	кумул. % наблюд.	ожидаем. частота	кумул. ожидаем.	процент ожидаем.
<=4.2600	0	0	0.00000	0.0000	4.18991	4.1899	2.79328
4.72000	11	11	7.33333	7.3333	8.92888	13.1188	5.95259
5.18000	30	41	20.00000	27.3333	18.61344	31.7322	12.40896
5.64000	24	65	16.00000	43.3333	28.72009	60.4523	19.14673
6.10000	30	95	20.00000	63.3333	32.80343	93.2558	21.86895
6.56000	25	120	16.66667	80.0000	27.73581	120.9916	18.49054
7.02000	18	138	12.00000	92.0000	17.35934	138.3509	11.57289
7.48000	6	144	4.00000	96.0000	8.04170	146.3926	5.36113
7.94000	6	150	4.00000	100.0000	2.75681	149.1494	1.83787
Беск-ть	0	150	0.00000	100.0000	.85059	150.0000	.56706

Самое важное понять, как общая форма распределения построенного по данным (т.е. наблюдаемого или эмпирического распределения) аппроксимирует гипотетическое нормальное распределение. Теперь вернемся в окно *Подгонка непрерывных распределений*. В поле *График распределения* выберите *Гистограмму* или *Кумулятивное распределение* с *По строкам* или *Относительными* частотами. По умолчанию, число групп (*23*), воспользуйтесь графиком по умолчанию, нажав кнопку *График,* чтобы построить гистограмму частот для этой переменной.

StatSoft®

Copyright © StatSoft, 1995



Похоже, что распределение *Sepallen* бимодально, иными словами, имеет два “пика”. Также видно, что подгонка наиболее удачна в левой части графика, где находится пик. Таким образом, можно заключить, что непрерывное нормальное распределение, по-видимому, не вполне адекватная модель для наблюдаемых данных.

Copyright © StatSoft, 1995

ТЕХНИЧЕСКИЕ ЗАМЕЧАНИЯ

Совпадающие ранги

Во всех ранговых методах делаются поправки на совпадающие ранги как указано в работах Siegel (1956) и Siegel and Castellan (1988).

Критерии значимости для выборок малого объема

Если размер выборки очень мал (например, n = 10 или меньше), то уровни значимости для тех критериев, которые используют z приближения, можно рассматривать только как грубые оценки (см. Hays (1981), Siegel (1956), and Siegel and Castellan (1988) для детального обсуждения).

Подгонка распределения Вейбулла и Гомпертца для цензурированных данных

Эти распределения особенно используются в биомедицине, а также в инженерных исследованиях. Они являются теоретическими распределениями времен отказов или времени жизни. Дополнительная трудность, возникающая в этих исследованиях, связана с наличием неполных или цензурированных наблюдений, например, пациенты, дожившие до того момента, когда исследование было закончено, или элементы (технической системы), которые не отказали вообще. Эти наблюдения, очевидно, содержат значимую информацию (пациент *X* жил более 3 лет), но они не полны (так как мы не можем сказать, как долго жил пациент *X*). Поскольку эти специальные задачи связаны именно с распределением Вейбулла и Гомпертца, то его подгонка полных и неполных (цензурированных) данных помещены в модуль *Анализ выживаемости*.

Подгонка распределения с помощью *2M графиков* и в модуле *Анализ процессов*

Имеется несколько альтернативных процедур доступных в *STATISTICA* для подгонки непрерывных распределений к данным. Например, вы можете выбрать (категоризованные или некатегоризованные) графики *Квантиль-квантиль*, *Вероятность-вероятность*, или *2M гистограммы* (с кривой, показывающей ожидаемые частоты) из меню *Галерея графиков* или *Графика,* а затем выбрать графики, основанные на распределениях: *Нормальное*, *Бета*, *Экстремального значения*, *Гамма*, *Геометрическое*, *Лапласа*, *Логистическое*, *Логнормальное*, *Пуассона*, *Релея*, *Вейбулла*. Программа вычисляет оценки максимального правдоподобия для параметров соответствующих распределений и далее осуществляет подгонку (см. том II относительно деталей).

Модуль *Анализ процессов* (см. том IV) также содержит несколько опций для подгонки этих распределений, а также для вычисления таблиц с ожидаемыми и наблюдаемыми частотами. Этот модуль имеет опцию подгонки всех этих непрерывных распределений к данным. Таблица результатов включает значения статистики *d-max* Колмогорова-Смирнова, упорядоченные по возрастанию (от меньшего значения к большему), что позволяет выбрать распределение, лучше всего подходящее к данным.

StatSoft®

Copyright © StatSoft, 1995

ПРЕДМЕТНЫЙ УКАЗАТЕЛЬ

Т

t-критерий для зависимых выборок, 1610, 1644
t-критерий для независимых выборок, 1610, 1639

А

Альтернативы ANOVA с повторными измерениями, 1610
Альтернативы t-критерию для зависимых выборок, 1610
Альтернативы t-критерию для независимых выборок, 1610
Альтернативы межгрупповой ANOVA, 1610
Асимметрия, 1619

Б

Бернулли распределение, 1622
Бета распределение, 1609, 1611, 1631, 1650
Биномиальное распределение, 1622
Большие массивы данных и непараметрические методы, 1608

В

Вальда-Вольфовица критерий серий, 1614, 1626, 1639
Вейбулла распределение, 1609, 1611, 1631, 1650
Вероятности Лиллиефорса, 1633
Вилкоксона критерий парных сравнений, 1616, 1628, 1644
Внутригрупповые (повторно измеренные) факторы ANOVA, 1610, 1643

Г

Гамма коэффициент, 1613, 1625, 1637, 1639
Гамма распределение, 1609, 1611, 1621, 1631, 1650
Гармоническое среднее, 1619
Геометрическое распределение, 1609, 1611, 1622, 1631, 1650
Геометрическое среднее, 1618
Гипотеза нормальности, 1608
Гомпертца распределение, 1611, 1631, 1650

Д

Диаграмма размаха, 1627, 1628, 1629
Дисперсионный анализ (ANOVA) Фридмана, 1617, 1629, 1643
Дисперсия, 1619

З

Зависимости между переменными, 1607

К

Категоризованная гистограмма, 1627, 1628
Квартильный размах, 1619
Колмогорова-Смирнова d-max статистика, 1633
Колмогорова-Смирнова двухвыборочный критерий, 1615, 1639
Колмогорова-Смирнова одновыборочный критерий, 1632
Конкордация Кендалла, 1617, 1629, 1643
Корреляции
  гамма, 1613, 1625, 1637, 1639
  Кендалла тау, 1613, 1624, 1637, 1639
  Спирмена R, 1613, 1624, 1637
Корреляция Кендалла, 1611, 1613, 1624, 1637, 1639
Кохрена Q критерий, 1617, 1630, 1645
Коэффициент корреляции, 1624
Краскела-Уоллиса ранговая ANOVA, 1615, 1627, 1640
Критерии значимости для выборок малого объема, 1650
Критерий V-квадрат, 1613, 1623, 1634
Критерий знаков, 1616, 1628, 1635
Критерий хи-квадрат
  поправка Йетса, 1623

Л

Лапласа распределение, 1609, 1611, 1631, 1650
Логистическое распределение, 1609, 1611, 1631, 1650
Логнормальное распределение, 1609, 1611, 1621, 1631, 1650

М

Макнемара хи-квадрат, 1613, 1623, 1635
Манна-Уитни U критерий, 1614, 1626, 1639
Медиана, 1618
Медианный тест, 1615, 1627, 1640
Межгрупповая ANOVA, 1610, 1640
Ментела-Хенцела критерий, 1611, 1624, 1635
Мода, 1618

Н

Наблюдаемые частоты в сравнении с ожидаемыми частотами хи-квадрат, 1613, 1624, 1637
Непараметрическая статистика, 1605, 1606
  альтернативы ANOVA с повторными измерениями, 1610
  альтернативы t-критерию для зависимых выборок, 1610
  альтернативы t-критерию для независимых выборок, 1610
  альтернативы межгрупповой ANOVA, 1610
  асимметрия, 1619
  большие массивы данных и непараметрические методы, 1608
  Вальда-Вольфовица критерий серий, 1614, 1626, 1639
  вероятности Лиллиефорса, 1633
  Вилкоксона критерий парных сравнений, 1616, 1628, 1644
  гамма, 1625
  гармоническое среднее, 1619
  геометрическое среднее, 1618
  дисперсионный анализ (ANOVA) Фридмана, 1617, 1629, 1643
  дисперсия, 1619

Copyright © StatSoft, 1995

StatSoft®

зависимости между переменными, 1607
квартильный размах, 1619
Колмогорова-Смирнова двухвыборочный критерий, 1615, 1639
конкордация Кендалла, 1617, 1629, 1643
корреляции, 1613, 1624, 1637
корреляция Кендалла, 1624
Кохрена Q критерий, 1617, 1630, 1645
Краскела-Уоллиса ранговая ANOVA, 1615, 1627, 1640
критерий V-квадрат, 1613, 1623
критерий знаков, 1616, 1628, 1635
критерий Ментела-Хенцела, 1611, 1624, 1635
Макнемара хи-квадрат, 1613
Манна-Уитни U критерий, 1614, 1626, 1639
медиана, 1618
медианный тест, 1615, 1627, 1640
мода, 1618
наблюдаемые частоты в сравнении с ожидаемыми частотами хи-квадрат, 1613, 1624, 1637
описательные статистики, 1607, 1610, 1630, 1646
процедуры для данных в виде частот, 1610
различия между зависимыми группами, 1606
различия между независимыми группами, 1606
размах, 1619
Спирмена R корреляция, 1624
сравнение двух групп, 1626
сравнение двух переменных, 1628
среднее отклонение, 1619
стандартное отклонение, 1619
таблицы частот два на два, 1613, 1623, 1634
фи-квадрат критерий, 1613, 1623, 1634
Фишера точный критерий, 1613
хи-квадрат критерий, 1613, 1623
эксцесс, 1620
Нормальное распределение, 1609, 1611, 1620, 1631, 1650

О

Объем выборки, 1605
Описательные статистики, 1607, 1618, 1630, 1646
асимметрия, 1619
гармоническое среднее, 1619
геометрическое среднее, 1618
дисперсия, 1619
квартильный размах, 1619
медиана, 1618
мода, 1618
размах, 1619
среднее отклонение, 1619
стандартное отклонение, 1619
эксцесс, 1620

П

Параметрические и непараметрические методы, 1606
Повторные измерения в ANOVA, 1610, 1643
Подгонка распределения, 1608, 1610, 1620, 1631, 1647
Поправка Йетса для критерия хи-квадрат, 1623
Пуассона распределение, 1609, 1611, 1622, 1631, 1650

Р

Равномерное распределение, 1621
Различия между зависимыми группами, 1606
Различия между независимыми группами, 1606
Размах, 1619
Ранги, совпадающие ранги, 1650
Распределения
Бернулли, 1622
бета, 1609, 1611, 1631, 1650
биномиальное, 1622
Вейбулла, 1609, 1611, 1631, 1650
гамма, 1609, 1611, 1621, 1631, 1650
геометрическое, 1609, 1611, 1622, 1631, 1650
Лапласа, 1609, 1611, 1631, 1650
логистическое, 1609, 1611, 1631, 1650
логнормальное, 1609, 1611, 1621, 1631, 1650
нормальное, 1609, 1611, 1620, 1631, 1650
Пуассона, 1609, 1611, 1622, 1631, 1650
равномерное, 1621
Релея, 1609, 1611, 1631, 1650
хи-квадрат, 1622
экспоненциальное, 1621
экстремального значения, 1609, 1611, 1631, 1650
Релея распределение, 1609, 1611, 1631, 1650

С

Согласие, 1632
Спирмена R корреляция, 1607, 1613, 1624, 1637
Сравнение двух групп, 1626
Сравнение двух переменных, 1628
Среднее отклонение, 1619
Стандартное отклонение, 1619
Стартовая панель, 1620
Статистика d-max, 1632
Статистика фи-квадрат, 1623

Т

Таблицы сопряженности, 1634
Таблицы частот, 1623
Таблицы частот два на два, 1613, 1623, 1634

Ф

Фи-квадрат критерий, 1613, 1623, 1634
Фишера точный критерий, 1613, 1623, 1635

Х

Хи-квадрат критерий, 1613, 1623, 1632, 1634
Макнемара хи-квадрат, 1623, 1635
Хи-квадрат распределение, 1622

Э

Экспоненциальное распределение, 1621
Экстремального значения распределение, 1609, 1611, 1631, 1650
Эксцесс, 1620

StatSoft®

Copyright © StatSoft, 1995

Глава 12:
МНОЖЕСТВЕННАЯ РЕГРЕССИЯ

Оглавление

Вводный обзор 1657
Обзор программы 1663
Примеры 1665
Пример 1: Стандартная регрессия 1665
Пример 2: Пошаговая регрессия 1673
Диалоговые окна, опции, статистики 1677
Стартовая панель 1677
Фиксированная нелинейная регрессия 1679
Просмотр описательных статистик 1680
Определение модели 1682
Результаты множественной регрессии 1685
Предсказание значений зависимой переменной 1689
Анализ остатков 1689
Примечания 1699
Техническая информация 1704
Предметный указатель 1706

Подробное оглавление

ВВОДНЫЙ ОБЗОР 1657
Общее назначение 1657
Вычислительные аспекты 1658
Предположения, ограничения и обсуждение практических вопросов 1660

ОБЗОР ПРОГРАММЫ 1663

ПРИМЕРЫ 1665
Пример 1: Стандартная регрессия 1665
Файл данных 1665
Цель исследования 1665
Начальный анализ 1665
Задание множественной регрессии 1668
Просмотр результатов 1668
Анализ остатков 1669
Пример 2: Пошаговая регрессия 1673
Файл данных 1673

Copyright © StatSoft, 1995

StatSoft®

Цель исследования ... 1673
Начало анализа ... 1674
Определение модели для пошаговой регрессии ... 1674
Шаг 0 ... 1674
Шаг 1 ... 1675
Шаг 2 ... 1675
Шаг 3 (Окончательное решение) ... 1676

ДИАЛОГОВЫЕ ОКНА, ОПЦИИ, СТАТИСТИКИ ... 1677

Стартовая панель ... 1677
Переменные ... 1677
Файл ввода ... 1677
Удаление пропущенных данных ... 1677
Тип регрессии ... 1677
Провести анализ по умолчанию ... 1678
Показывать описательные статистики и корреляционную матрицу ... 1678
Вычисления с повышенной точностью ... 1678
Пакетная обработка и печать ... 1679
Печатать результаты анализа остатков ... 1679
Фиксированная нелинейная регрессия ... 1679
Просмотр описательных статистик ... 1680
Средние и стандартные отклонения ... 1680
Диаграмма размаха ... 1680
Корреляции ... 1681
Матричный график ... 1681
Ковариации ... 1681
Сохранить корреляции ... 1681
Число пар ... 1681
Стд.откл. = *Сумма квадратов*/*n* ... 1681
Определение модели ... 1682
Переменные ... 1682
Процедура ... 1682
Пошаговая регрессия по блокам (иерархическая регрессия) ... 1682
Свободный член ... 1683
Толерантность ... 1683
F-включить, F-исключить ... 1684
Число шагов ... 1684
Отображение результатов ... 1684
Гребневая регрессия ... 1684
Пакетная обработка и печать ... 1685
Печатать результаты анализа остатков ... 1685
Просмотреть описательные статистики ... 1685
Результаты множественной регрессии ... 1685
Информационное поле ... 1686

StatSoft®

Copyright © StatSoft, 1995

Поле параметров модели ... 1686
Итоговая таблица регрессии ... 1686
Дисперсионный анализ ... 1687
Дисперсионный анализ, скорректированный на среднее ... 1687
Ковариации коэффициентов ... 1687
Текущая матрица выметания ... 1687
Частные корреляции ... 1688
Итоги по шагам ... 1688
Предсказать зависимую переменную ... 1688
Избыточность ... 1688
Корреляции и описательные статистики ... 1689
Выделяемый уровень значимости ... 1689
Предсказание значений зависимой переменной ... 1689
Анализ остатков ... 1689
Корреляции и описательные статистики ... 1690
Итоги регрессии ... 1690
Остатки и предсказанные ... 1690
Статистика Дарбина-Уотсона ... 1691
Сохранить остатки и предсказанные ... 1691
Графики остатков ... 1692
Графики выбросов ... 1692
Графики предсказанных значений ... 1693
Предсказанные и остатки ... 1693
Предсказанные и квадраты остатков ... 1693
Предсказанные и наблюдаемые ... 1694
Наблюдаемые и остатки, наблюдаемые и квадраты остатков ... 1694
Остатки и удаленные остатки ... 1694
График наблюдаемых ... 1695
График предсказанных ... 1695
График остатков ... 1695
Нормальный график остатков ... 1695
Полунормальный график остатков ... 1696
Нормальный график остатков без тренда ... 1696
Две переменные ... 1697
Остатки и переменная ... 1697
Предсказанные и переменная ... 1697
Частные остатки ... 1698

ПРИМЕЧАНИЯ ... 1699
Исходные данные на входе, удаление пропущенных данных ... 1699
Матрица на входе ... 1699
Опция пакетной обработки ... 1699
Использование попарного удаления пропущенных данных ... 1700
Задание значений *F-включить /исключить* для пошаговой регрессии ... 1700

Copyright © StatSoft, 1995

Включение переменных по блокам (иерархический анализ)..... 1701
Плохо обусловленная матрица 1701
Гребневая регрессия 1702
Некоторые полезные программы на языке *STATISTICA BASIC* 1702
Исключительно большие задачи множественной регрессии 1702
Оценивание методом взвешенных наименьших квадратов 1702
Оценивание двухступенчатым методом наименьших квадратов 1702
Преобразования Бокса-Кокса и Бокса-Тидвелла 1702

ТЕХНИЧЕСКАЯ ИНФОРМАЦИЯ..... 1704
Общий обзор 1704
Вычисления при проведении множественной регрессии, обращение матрицы 1704
Проверка статистической значимости 1704
Остатки 1704
Вычисления для нормальных вероятностных графиков 1705

ПРЕДМЕТНЫЙ УКАЗАТЕЛЬ..... 1706

StatSoft®

Copyright © StatSoft, 1995

Глава 12:

МНОЖЕСТВЕННАЯ РЕГРЕССИЯ

ВВОДНЫЙ ОБЗОР

Общее назначение

Общее назначение множественной регрессии состоит в анализе связи между несколькими независимыми переменными (называемыми также регрессорами или предикторами) и зависимой переменной. Например, агент по продаже недвижимости мог бы вносить в каждый элемент реестра размер дома (в квадратных футах), число спален, средний доход населения в этом районе в соответствии с данными переписи и субъективную оценку привлекательности дома.

Как только эта информация собрана для различных домов, было бы интересно посмотреть, связаны ли и каким образом эти характеристики дома с ценой, по которой он был продан. Например, могло бы оказаться, что число спальных комнат является лучшим предсказывающим фактором (предиктором) для цены продажи дома в некотором специфическом районе, чем "привлекательность" дома (субъективная оценка). Могли бы также обнаружиться и "выбросы", т.е. дома, которые могли бы быть проданы дороже, учитывая их расположение и характеристики.

Специалисты по кадрам обычно используют процедуры множественной регрессии для определения вознаграждения адекватного выполненной работе. Можно определить некоторое количество факторов или параметров, таких, как "размер ответственности" (*Resp*) или "число подчиненных" (*No_Super*), которые, как ожидается, оказывают влияние на стоимость работы. Кадровый аналитик затем проводит исследование размеров окладов (*Salary*) среди сравнимых компаний на рынке, записывая размер жалования и соответствующие характеристики (т.е. значения параметров) по различным позициям. Эта информация может быть использована при анализе с помощью множественной регрессии для построения регрессионного уравнения в следующем виде:

```
Salary = .5*Resp + .8*No_Super
```

Как только эта так называемая линия регрессии определена, аналитик оказывается в состоянии построить график ожидаемой (предсказанной) оплаты труда и реальных обязательств компании по выплате жалования. Таким образом, аналитик может определить, какие позиции недооценены (лежат ниже линии регрессии), какие оплачиваются слишком высоко (лежат выше линии регрессии), а какие оплачены адекватно.

В общественных и естественных науках процедуры множественной регрессии чрезвычайно широко используются в исследованиях. В общем, множественная регрессия позволяет исследователю задать вопрос (и, вероятно, получить ответ) о том, “что является лучшим предиктором для...”.

Например, исследователь в области образования мог бы пожелать узнать, какие факторы являются лучшими предикторами успешной учебы в средней школе. А психолога мог быть заинтересовать вопрос, какие индивидуальные качества позволяют лучше предсказать степень социальной адаптации индивида. Социологи, вероятно, хотели бы найти те социальные индикаторы, которые лучше других предсказывают результат адаптации новой иммигрантской группы и степень ее слияния с обществом. [Заметим, что термин множественная указывает на наличие нескольких

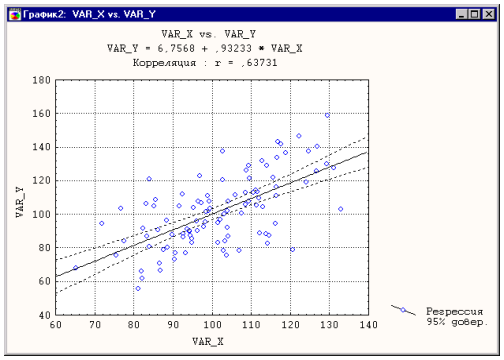
Copyright © StatSoft, 1995

StatSoft®

предикторов или регрессоров, которые используются в модели.]

Вычислительные аспекты

Общая вычислительная задача, которую требуется решать при анализе методом множественной регрессии, состоит в подгонке прямой линии к некоторому набору точек.



В простейшем случае, когда имеется одна зависимая и одна независимая переменная, это можно увидеть на диаграмме рассеяния.

Линейное оценивание методом наименьших квадратов. На диаграмме рассеяния имеется независимая переменная или переменная *X* и зависимая переменная — *Y*. Эти переменные могут, например, представлять коэффициент IQ (уровень интеллекта, оцененный с помощью теста) и достижения в учебе (средний балл успеваемости - grade point average; GPA) соответственно. Каждая точка на диаграмме представляет данные одного студента, т.е. его соответствующие показатели IQ и GPA.

Целью процедур линейной регрессии является подгонка прямой линии по точкам. А именно, программа строит линию регрессии так, чтобы минимизировать квадраты отклонений этой линии от наблюдаемых точек. Поэтому на эту общую процедуру иногда ссылаются как на оценивание методом *наименьших квадратов*.

Уравнение регрессии. Прямая линия в двумерном пространстве или пространстве двух переменных, задается уравнением $Y=a+b*X$; более подробно: переменная *Y* может быть выражена через константу (*a*) и угловой коэффициент (*b*), помноженный на переменную *X*. Константу иногда называют также свободным членом, а угловой коэффициент — *регрессионным коэффициентом* или *B-коэффициентом*. Например, значение GPA можно лучше всего предсказать по формуле *1+.02*IQ*. Таким образом, зная, что коэффициент *IQ* у студента равен 130, вы могли бы предсказать его показатель успеваемости GPA, скорее всего, он близок к 3.6 (поскольку 1+.02*130=3.6).

В многомерном случае, когда имеется более одной независимой переменной, линия регрессии не может быть отображена в двумерном пространстве, однако она также может быть легко оценена (с помощью модуля *Множественной регрессии*; хотя вычисления являются в действительности довольно сложными).

Например, если в дополнение к *IQ* вы имеете другие предикторы успеваемости (например, *Мотивация*, *Самодисциплина*), вы можете построить линейное уравнение, содержащее все эти переменные. Тогда, в общем случае, процедуры множественной регрессии будут оценивать параметры линейного уравнения вида:

$$Y = a + b_1*X_1 + b_2*X_2 + ... + b_p*X_p$$

Однозначный прогноз и частная корреляция. Заметим, что в этом уравнении, регрессионные коэффициенты (или *B*-коэффициенты) представляют *независимые* вклады каждой независимой переменной в предсказание зависимой переменной. Другими словами, переменная X_1, к примеру, коррелирует с переменной *Y* после учета влияния всех других независимых переменных. Этот тип корреляции упоминается также под названием *частной*

StatSoft
Copyright © StatSoft, 1995

корреляции. Вероятно, следующий пример пояснит это понятие.

Кто-то мог бы, вероятно, обнаружить значимую отрицательную корреляцию в популяции между длиной волос и ростом (невысокие люди обладают более длинными волосами). На первый взгляд это может показаться странным; однако, если добавить переменную *Пол - Gender* в уравнение множественной регрессии, эта корреляция, скорее всего, исчезнет. Это произойдет из-за того, что женщины, в среднем, имеют более длинные волосы, чем мужчины; при этом они также в среднем ниже мужчин. Таким образом, после удаления разницы по полу посредством ввода предиктора *Пол - Gender* в уравнение, связь между длиной волос и ростом исчезает, поскольку длина волос не дает какого-либо самостоятельного вклада в предсказание роста помимо того, который она разделяет с переменной *Пол - Gender*. Другими словами, после учета переменной *Пол - Gender* частная корреляция между длиной волос и ростом нулевая. [Иными словами, если одна величина коррелирована с другой, то это может быть отражением того факта, что они обе коррелированы с третьей величиной или с совокупностью величин.]

Предсказанные значения и остатки. Линия регрессии выражает наилучшее предсказание зависимой переменной (*Y*) по независимым переменным (*X*). Однако, природа редко (если вообще когда-нибудь) бывает полностью предсказуемой и обычно имеется существенный разброс наблюдаемых точек около подогнанной линии регрессии (как это было показано ранее на диаграмме рассеяния). Отклонение отдельной точки от линии регрессии (от предсказанного значения) называется остатком.

Остаточная дисперсия и коэффициент детерминации *R-квадрат*. Чем меньше разброс значений остатков около линии регрессии по отношению к общему разбросу значений, тем, очевидно, лучше прогноз.

Например, если связь между переменными *X* и *Y* отсутствует, то отношение остаточной изменчивости переменной *Y* к исходной дисперсии равно 1.0. Если *X* и *Y* жестко связаны, то остаточная изменчивость отсутствует, и отношение дисперсий будет равно 0.0. В большинстве случаев отношение будет лежать где-то между этими экстремальными значениями, т.е. между 0.0 и 1.0. 1.0 минус это отношение называется *R-квадратом* или *коэффициентом детерминации*.

Это значение непосредственно интерпретируется следующим образом. Если имеется *R-квадрат* равный 0.4, то изменчивость значений переменной *Y* около линии регрессии составляет 1-0.4 от исходной дисперсии; другими словами, 40% от исходной изменчивости могут быть объяснены, а 60% остаточной изменчивости остаются необъясненными. В идеале желательно иметь объяснение если не для всей, то хотя бы для большей части исходной изменчивости. Значение *R-квадрата* является индикатором степени подгонки модели к данным (значение *R-квадрата* близкое к 1.0 показывает, что модель объясняет почти всю изменчивость соответствующих переменных).

Интерпретация коэффициента множественной корреляции *R*. Обычно, степень зависимости двух или более предикторов (независимых переменных или переменных *X*) с зависимой переменной (*Y*) выражается с помощью коэффициента множественной корреляции *R*. По определению он равен корню квадратному из коэффициента детерминации. Это неотрицательная величина, принимающая значения между 0 и 1.

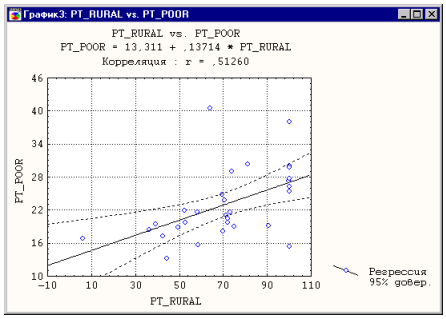
Для интерпретации направления связи между переменными смотрят на знаки (плюс или минус) регрессионных коэффициентов или *B*-коэффициентов. Если *B*-коэффициент

Copyright © StatSoft, 1995

положителен, то связь этой переменной с зависимой переменной положительна (например, чем больше IQ, тем выше средний показатель успеваемости оценки); если *B*-коэффициент отрицателен, то и связь носит отрицательный характер (например, чем меньше число учащихся в классе, тем выше средние оценки по тестам). Конечно, если *B*-коэффициент равен 0, связь между переменными отсутствует.

Предположения, ограничения и обсуждение практических вопросов

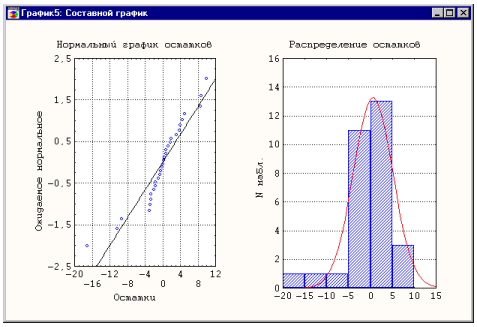
Предположение линейности. Прежде всего, как это видно уже из названия множественной *линейной* регрессии, предполагается, что связь между переменными является линейной. На практике это предположение, в сущности, никогда не может быть подтверждено; к счастью, процедуры множественного регрессионного анализы в незначительной степени подвержены воздействию малых отклонений от этого предположения. Однако *всегда* имеет смысл посмотреть на двумерные диаграммы рассеяния переменных, представляющих интерес.



В модуле *Множественная регрессия*, эти графики легко доступны из любого места программы просто с помощью опции электронных таблиц с корреляционной матрицей, позволяющей построить график (диаграмму рассеяния) для выбранных ячеек. Если нелинейность связи очевидна, то можно рассмотреть или преобразования переменных (с помощью программы на языке *STATISTICA BASIC* или *формул электронных таблиц*, см. главу 3), или явно допустить включение нелинейных членов.

Фиксированный нелинейный режим модуля *Множественная регрессия* позволяет вводить различные нелинейные члены в уравнение регрессии, т.е. явно проверять значимость нелинейного члена для взаимосвязи двух или более переменных (другие опции нелинейной регрессии доступны в модуле *Нелинейное оценивание*; см. том III).

Предположение нормальности. В множественной регрессии предполагается, что остатки (предсказанные значения минус наблюдаемые) распределены нормально (т.е. подчиняются закону нормального распределения). И снова, хотя большинство тестов (в особенности *F*-тест) довольно робастны (устойчивы) по отношению к отклонениям от этого предположения, всегда, прежде чем сделать окончательные выводы, стоит рассмотреть распределения большинства представляющих интерес переменных. В модуле *Множественная регрессия* можно построить гистограмму (с наложенной на нее нормальной плотностью) с помощью электронных таблиц с описательными статистками, которые доступны в любой момент анализа. Пользователь также может получить гистограммы и нормальный вероятностный график остатков для визуального анализа их распределения.



StatSoft
Copyright © StatSoft, 1995

Ограничения. Основное концептуальное ограничение всех методов регрессионного анализа состоит в том, что они позволяют обнаружить только числовые зависимости, а не лежащие в их основе причинные связи. Например, можно обнаружить сильную положительную связь (корреляцию) между разрушениями, вызванными пожаром, и числом пожарных, участвующих в борьбе с огнем. Следует ли заключить, что пожарные вызывают разрушения? Конечно, наиболее вероятное объяснение этой корреляции состоит в том, что размер пожара (внешняя переменная, которую забыли включить в исследование) оказывает влияние, как на масштаб разрушений, так и на привлечение определенного числа пожарных (т.е. чем больше пожар, тем большее количество пожарных вызывается на его тушение). Хотя этот пример довольно прозрачен, в реальности при исследовании корреляций альтернативные причинные объяснения часто даже не рассматриваются.

Выбор числа переменных. Множественная регрессия - "соблазнительный" метод: “включайте” в качестве предикторов все переменные, какие только можете, и, по крайней мере, некоторые из них окажутся значимыми. Это происходит из-за того, что извлекается выгода из случайностей, возникающих при простом включении возможно большего числа переменных, рассматриваемых в качестве предикторов другой, представляющей интерес переменной. Эта проблема возникает тогда, когда к тому же и число наблюдений относительно мало. Интуитивно ясно, что едва ли можно делать выводы из анализа вопросника со 100 пунктами на основе ответов 10 респондентов. Большинство авторов советуют использовать, по крайней мере, от 10 до 20 наблюдений (респондентов) на одну переменную, в противном случае оценки регрессионной линии будут, вероятно, очень ненадежными и, скорее всего, невоспроизводимыми для желающих повторить это исследование.

Мультиколлинеарность и плохая обусловленность матрицы. Это является общей проблемой для многих методов корреляционного анализа. Представим, что имеется два предиктора (переменные *X*) для роста субъекта: (1) вес в фунтах и (2) вес в унциях. Очевидно, иметь оба предиктора совершенно излишне; вес является одной и той же переменной, меряется ли он в фунтах или унциях. Попытка определить, какая из двух мер является лучшим предиктором, выглядит довольно глупо; однако, в точности это происходит при попытке выполнить множественный регрессионный анализ с ростом в качестве зависимой переменной (*Y*) и двумя мерами веса, как независимыми переменными (*X*). *STATISTICA* в этом случае выдаст сообщение о *плохой обусловленности матрицы* для извещения пользователя о том, что он или она пытаются сделать невозможное. Если в анализ включено много переменных, то часто не сразу очевидно существование этой проблемы, и она может возникнуть только после того, как некоторые переменные будут уже включены в регрессионное уравнение. Тем не менее, если такая проблема возникает, это означает, что, по крайней мере, одна из зависимых переменных (предикторов) является совершенно лишней при наличии остальных предикторов. Модуль *Множественной регрессии* содержит много статистических индикаторов такого рода избыточности (*толерантность*, *получастное R* и т.д., см. раздел *Результаты множественной регрессии*, стр. 1685), а также некоторые средства борьбы с ней (например, *Гребневая регрессия*).

Важность анализа остатков. Хотя большинство предположений множественной регрессии нельзя проверить в точности, большие отклонения от этих предположений могут быть обнаружены, и должны быть соответствующим

Copyright © StatSoft, 1995

образом рассмотрены. В частности, выбросы (т.е. экстремальные наблюдения) могут вызвать серьезное смещение результатов, "сдвигая" линию регрессии в определенном направлении и тем самым, вызывая смещение регрессионных коэффициентов. Часто исключение всего одного экстремального наблюдения приводит к совершенно другому результату. Поэтому одна из задач при создании интерфейса модуля *Множественной регрессии* упрощении анализ остатков для пользователя, насколько это возможно, с тем чтобы отклонения от предположений, угрожающие обоснованности результатов анализа, могли быть легко обнаружены.

StatSoft®

Copyright © StatSoft, 1995

ОБЗОР ПРОГРАММЫ

С помощью этого модуля, пользователь может выполнить множественный линейный регрессионный анализ, регрессию с пошаговым включением или исключением предикторов, включение или исключение предикторов блоками (иерархический анализ), выполнить гребневую или ридж регрессию, а также фиксированную нелинейную регрессию (например, подобрать полиномиальные уравнения). Модуль *Множественной регрессии* также позволит выполнить обширный анализ остатков. Пользователь может вычислить предсказанные значения и остатки (включая удаленные остатки, стандартизованные предсказанные значения и остатки, расстояния Махаланобиса, расстояния Кука и др.), вычислить статистику Дарбина-Уотсона и сериальную корреляцию остатков. Остатки и предсказанные значения можно сохранить для дальнейшего анализа в других модулях.

Любой вид анализа может быть выполнен с учетом и без учета равенства нулю свободного члена регрессионного уравнения (если свободный член равен нулю; то линия регрессии проходит через начало координат); скорректированные суммы квадратов и значения статистики *R-квадрат* могут быть вычислены для моделей без свободного члена (основываясь на изменчивости относительно среднего, а не относительно нуля). Можно использовать как файлы исходных данных, так и корреляционные матрицы. В дополнение к стандартному выводу множественной регрессии пользователь может просмотреть стандартизованные и нестандартизованные регрессионные коэффициенты, их стандартные ошибки и статистическую значимость, частные корреляции, получастные корреляции, телерантность, корреляции и ковариации регрессионных весов, обратную матрицу к корреляционной матрице независимых переменных и сводку результатов пошагового регрессионного анализа (включение, исключение или блочное включение-исключение предикторов).

Также после построения оценки для регрессионного уравнения пользователь может в интерактивном режиме вычислять предсказанные значения по любому набору предикторов. Корреляционная матрица может вычисляться с попарным или построчным удалением пропущенных данных. Пропущенные данные также могут быть автоматически заменены на соответствующие средние значения. Корреляционная матрица может быть сохранена для последующего анализа другими модулями *STATISTICA*.

В дополнение к стандартным опциям и опциям макрообработки, доступным во всех модулях *STATISTICA*, модуль *Множественной регрессии* содержит внутреннюю пакетную опцию, позволяющую пользователю задавать список зависимых переменных, подлежащих последовательной обработке с одинаковой регрессионной схемой анализа.

Графики. Многочисленные статистические графики доступны при вычислении регрессионной модели. Распределение остатков может быть отображено на нормальном, полунормальном и нормальном без тренда вероятностных графиках и гистограммах; предсказанные значения, остатки, удаленные остатки, расстояния Махаланобиса и др. могут быть одновременно отображены на графиках с пиктограммами; различные типы диаграмм рассеяния могут отображать остатки против предсказанных значений, квадраты остатков против предсказанных значений, удаленные остатки против остатков и т.д.

Copyright © StatSoft, 1995

Альтернативные процедуры. Модуль *Факторного анализа* (см. том III) также выполняет множественный регрессионный анализ для больших задач (до 300 переменных); регрессионные задачи практически неограниченного размера можно решить с помощью программы-примера *Regressn.stb* на языке *STATISTICA BASIC* (находящемся в подкаталоге *STBASIC*). Общая нелинейная процедура представлена в модуле *Нелинейное оценивание*, с помощью которой пользователь может оценить произвольную регрессионную функцию, используя оценивание методом наименьших квадратов, или используя другую функцию потерь (например, модули отклонений, взвешенные наименьшие квадраты и т.д.). Если имеется несколько как зависимых, так и независимых переменных, то подходящим будет канонический анализ (см. *Каноническая корреляция*, том III). Для проверки особых априорных гипотез о структуре корреляционной матрицы и связи между переменными или лежащих в их основе скрытых соотношений (например, для выполнения путевого анализа), используйте модуль *Моделирование структурными уравнениями и анализ путей* (см. том III). Если зависимая переменная измерена в категориальных терминах (например, претендент сдал или не сдал экзамен), то возможно использование модуля *Дискриминантного анализа* (для нахождения наилучших предикторов, которые разделяют людей, сдавших экзамен и не прошедших его). Если зависимая переменная по своей природе является действительно категориальной (например, фенотип может быть либо мужским, либо женским; вы хотите предсказывать *лежащую в основе вероятность* получения того или иного фенотипа), то пробит или логит регрессионная модель может оказаться более подходящей (см. *Нелинейное оценивание*, том III). Если набор данных включает неполные (или цензурированные) наблюдения (например, при анализе времени до отказа или времени дожития), используйте регрессионные модели, представленные в модуле *Анализ выживаемости*. Модуль *Временные ряды* содержит процедуры для вычисления автокорреляций, а также содержит обширный набор процедур анализа корреляций с запаздыванием. Наконец, имеется несколько примеров программ на языке *STATISTICA BASIC* (в подкаталоге *STBASIC*), предназначенных для решения специфических регрессионных задач (например, оценивания по методу взвешенных наименьших квадратов, по двухступенчатому методу наименьших квадратов, проведения преобразования Бокса-Кокса и Бокса-Тидвелла и др.; более подробное описание этих программ смотрите на стр. 1702).

StatSoft®

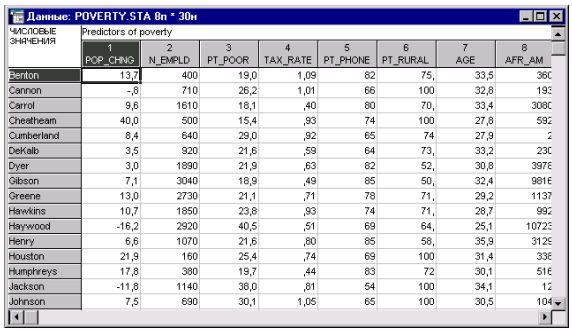
Copyright © StatSoft, 1995

ПРИМЕРЫ

Пример 1:
Стандартная регрессия

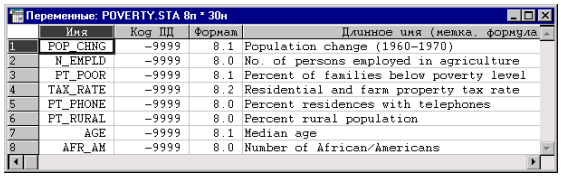
Файл данных

Следующий пример использует файл данных *Poverty.sta*, включенный в *STATISTICA*. Ниже приведена часть этого файла:



Данные основаны на сравнении результатов переписи 1960 и 1970 годов для случайной выборки из 30 округов. Имена округов введены в качестве идентификаторов наблюдений.

Следующая информация по каждой переменной приводится в электронной таблице *Просмотр спецификаций всех переменных* (открывающемся при щелчке мышью на кнопке *Спецификации всех переменных* панели инструментов электронных таблиц или выбором *Переменные - Спецификации всех переменных* из выпадающего меню для элемента меню *Правка* в окне электронной таблицы; см. главу 3, *Окно электронной таблицы*):

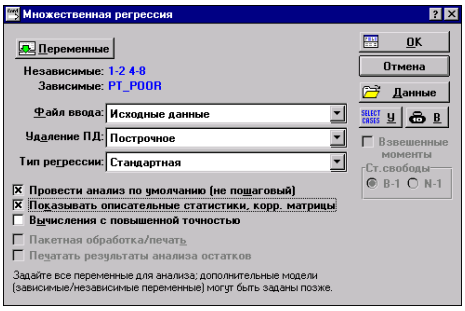


Цель исследования

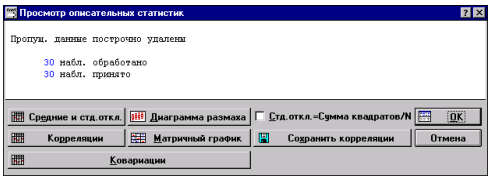
Мы проанализируем корреляты бедности (т.е. предикторы, "сильно" коррелирующие с процентом семей, живущих за чертой бедности). Таким образом, будем рассматривать переменную *3* (*Pt_Poor*), как зависимую или критериальную переменную, а все остальные переменные — в качестве независимых переменных или предикторов.

Начальный анализ

Когда вы запускаете модуль *Множественной регрессии*, открывается стартовая панель *Множественной регрессии*. Вы можете задать регрессионное уравнение щелчком мыши по кнопке *Переменные* и выбрав *Pt_Poor* в качестве зависимой переменной, а все остальные переменные набора данных — в качестве независимых. Отметьте также опции *Провести анализ по умолчанию* и *Показывать описательные статистики, корр. матрицы*. При этом *STATISTICA* выполнит стандартный (т.е. не пошаговый) регрессионный анализ.



Теперь нажмите *OK* этого диалогового окна, после чего откроется диалоговое окно *Просмотр описательных статистик*.



Copyright © StatSoft, 1995

StatSoft®

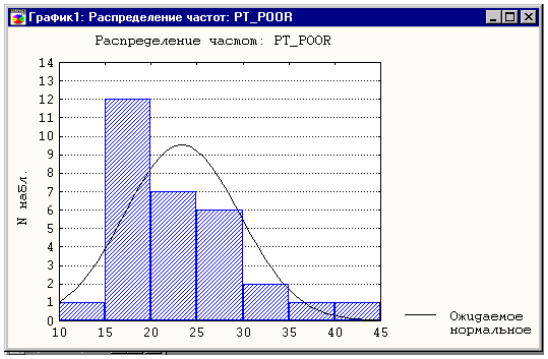
Здесь вы можете просмотреть средние и стандартные отклонения, корреляции и ковариации между переменными. Отметим, что это диалоговое окно доступно практически из всех последующих окон модуля *Множественная регрессия*, так что вы всегда сможете вернуться назад, чтобы посмотреть на описательные статистики определенных переменных.

Распределение переменных. Сначала изучим распределение зависимой переменной *Pt_Poor* по округам. Нажмите *Средние и стд.откл.* для показа таблицы результатов.

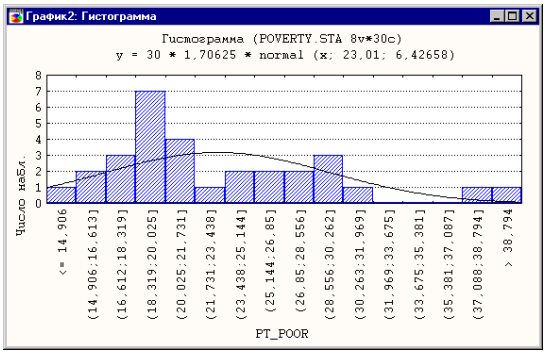
Средние и станд.отклонения (poverty.sta)

Далее...	Среднее	См.откл.	N
POP_CHNG	7,87	10,33	30
N_EMPLD	1548,67	2038,39	30
TAX_RATE	,72	,20	30
PT_PHONE	74,83	10,01	30
PT_RURAL	70,73	24,02	30
AGE	30,28	2,88	30
AFR_AM	11655,00	48289,92	30
PT_POOR	23,01	6,43	30

Теперь щелкните правой кнопкой мыши на переменной *Pt_Poor* и выберите опцию *Гистограмма/нормальное* из меню *Быстрые статистические графики*, после чего программа построит следующую гистограмму.



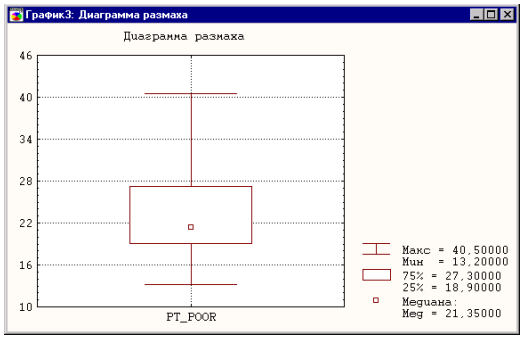
Используя опцию *Статистические 2М графики – Гистограммы...* из выпадающего меню для *Графика*, вы можете получить следующую гистограмму переменной *Pt_Poor* с заданным числом интервалов (в окне *2М гистограммы* установите опцию *Число категорий* в строке *Категории* равной *16*). Как видно ниже, распределение этой переменной чем-то отличается от нормального распределения. Коэффициенты корреляции могут оказаться существенно завышенными или заниженными при наличии в выборке существенных выбросов. Однако, хотя два округа (две самые правые колонки) имеют более высокий процент семей, проживающих за чертой бедности, чем это можно было бы ожидать в соответствии с нормальным распределением, они все еще, как нам кажется, находятся "в рамках допустимого".



Это решение является в определенной степени субъективным; эмпирическое правило состоит в том, что беспокойство требуется проявлять только тогда, когда наблюдение (или наблюдения) лежат вне интервала, заданного средним значением ± 3 стандартных отклонения. В этом случае будет разумно повторить критическую (с точки зрения влияния выбросов) часть анализа с выбросами и без них, с тем, чтобы удостовериться в отсутствии их влияния на характер взаимных корреляций. Вы также можете просмотреть распределение этой переменной, щелкнув мышкой на кнопке *Диаграмма размаха* диалогового окна *Просмотр описательных статистик*, выбрав переменную *Pt_Poor* и опцию *Медиана/квартили/размах*, а затем нажав *OK*.

StatSoft

Copyright © StatSoft, 1995



(Заметим, что определенный метод вычисления медианы и квартилей может быть выбран для всей "системы" в диалоговом окне *Сервис – Общие умолчания*; см. главу 3.)

Диаграммы рассеяния. Если имеются априорные гипотезы о связи между определенными переменными, на этом этапе может оказаться полезным вывести соответствующую диаграмму рассеяния.

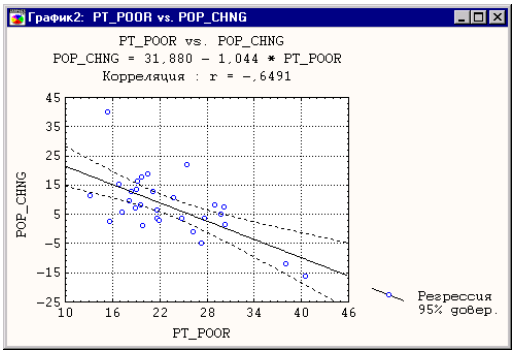
Например, посмотрим на связь между изменением популяции и процентом семей, проживающих за чертой бедности. Было бы естественно ожидать, что бедность приводит к миграции населения; таким образом, должна наблюдаться отрицательная корреляция между процентом семей, проживающих за чертой бедности, и изменением популяции.

Возвратимся к диалоговому окну *Просмотр описательных статистик* и щелкнем мышкой по кнопке *Корреляции* для отображения таблицы результатов с корреляционной матрицей.

Корреляции (poverty.sta)

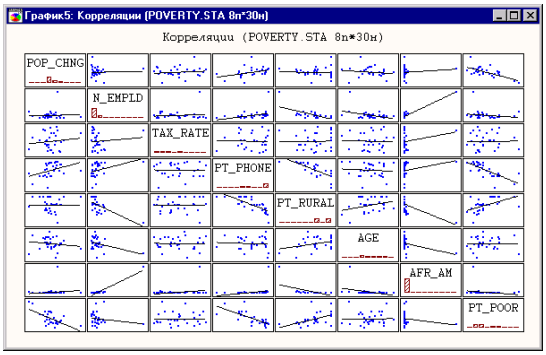
Далее...	POP_CHNG	N_EMPLD	TAX_RATE	PT_PHONE	PT_RURAL	AGE	AFR_AM	PT_POOR
POP_CHNG	1,0000	,0403	,1310	,3783	-,0188	-,1469	,1211	-,6491
N_EMPLD	,0403	1,0000	,1035	,3555	-,6577	-,3637	,9418	-,1677
TAX_RATE	,1310	,1035	1,0000	-,0380	,0233	-,0467	,2279	,0090
PT_PHONE	,3783	,3555	-,0380	1,0000	-,7486	-,0784	,2703	-,7335
PT_RURAL	-,0188	-,6577	,0233	-,7486	1,0000	,3144	-,5612	,5126
AGE	-,1469	-,3637	-,0467	-,0784	,3144	1,0000	-,3487	,0207
AFR_AM	,1211	,9418	,2279	,2703	-,5612	-,3487	1,0000	-,1921
PT_POOR	-,6491	-,1677	,0090	-,7335	,5126	,0207	-,1921	1,0000

Щелкните правой кнопкой мыши на корреляции между *Pop_Chng* и *Pt_Poor* и выберите опцию *Быстрые статистические графики* для получения диаграммы рассеяния с параметрами по умолчанию.



Эта диаграмма рассеяния иллюстрирует существенную отрицательную корреляцию (*-.65*) между двумя переменными. Она также показывает 95% доверительные интервалы для линии регрессии, т.е., вы можете на 95% быть уверенными в том, что действительная линия регрессии в популяции окажется внутри границ, заданных двумя пунктирными кривыми.

Корреляции между переменными могут быть отображены также и на матричной диаграмме рассеяния. Матричная диаграмма рассеяния для выбранных переменных может быть получена щелчком мыши по кнопке *Матричный график* диалогового окна *Просмотр описательных статистик* и последующим выбором интересующих переменных. Для матричной диаграммы рассеяния, показанной ниже, были выбраны все переменные.



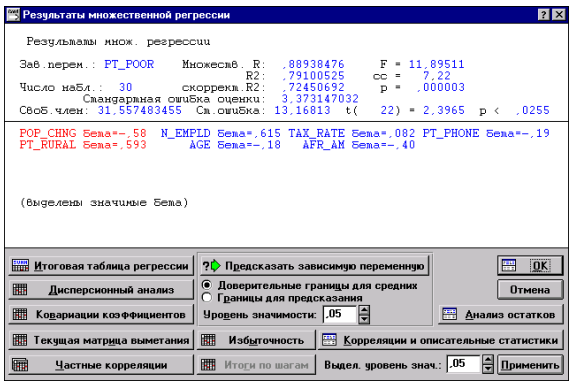
Copyright © StatSoft, 1995
StatSoft®

Задание множественной регрессии

Если опция *Выполнить анализ по умолчанию* на стартовой панели отмечена (установка по умолчанию), то для выполнения регрессионного анализа от вас требуется только щелкнуть по кнопке *OK* в окне *Просмотр описательных статистик* и перейти в диалоговое окно результатов. Стандартная регрессионный анализ (со свободным членом) будет выполнен автоматически.

Просмотр результатов

Ниже изображено диалоговое окно *Результаты множественной регрессии*. Общее уравнение множественной регрессии высоко значимо (см. главу *Элементарные понятия статистики*, глава 8, по поводу обсуждения проверки статистической значимости). Таким образом, зная значения независимых переменных, можно "предсказать" предиктор, связанный с бедностью, лучше, чем угадывая его чисто случайно.



Регрессионные коэффициенты. Чтобы узнать, какие из независимых переменных дают больший вклад в предсказание предиктора, связанного с бедностью, изучим регрессионные (или *B*) коэффициенты. Щелкните мышкой по кнопке *Итоговая таблица регрессии* для вывода таблицы результатов с этими коэффициентами.

	БЕТА	Ст. ош. БЕТА	B	Ст. ош. B	t(22)	p-уров.
Св. член			31,55748	13,16813	2,39651	,025487
POP_CHNG	-,578434	,136127	-,35978	,08467	-4,24924	,000328
N_EMPLD	,614602	,352958	,00194	,00111	1,74129	,095601
TAX_RATE	,082054	,106579	2,60150	3,37904	,76989	,449552
PT_PHONE	-,192963	,209013	-,12392	,13423	-,92321	,365914
PT_RURAL	,592775	,230528	,15859	,06167	2,57138	,017413
AGE	-,176649	,114227	-,39352	,25446	-1,54647	,136256
AFR_AM	-,396321	,341287	-,00005	,00005	-1,16125	,257984

Эта таблица показывает стандартизованные регрессионные коэффициенты (*Бета*) и обычные регрессионные коэффициенты (*B*). *Бета*-коэффициенты — это коэффициенты, которые получатся, если предварительно стандартизовать все переменные к среднему *0* и стандартному отклонению *1*. Таким образом, величина этих *Бета*-коэффициентов позволяет сравнивать относительный вклад каждой независимой переменной в предсказание зависимой переменной. Как видно из таблицы результатов, изображенной выше, переменные *Pop_Chng*, *Pt_Rural*, and *N_Empld* являются наиболее важными предикторами для бедности; из них только первые два статистически значимы. Регрессионный коэффициент для *Pop_Chng* отрицателен; т.е. чем меньше прирост популяция, тем большее число семей живут ниже уровня бедности в соответствующем округе. Вклад в регрессию для *Pt_Rural* положителен; т.е. чем больше процент сельского населения, тем выше уровень бедности.

Частные корреляции. Другой путь изучения вкладов каждой независимой переменной в предсказание зависимой переменной состоит в вычислении частных и получастных корреляций (щелкните на кнопке *Частные корреляции* в диалоговом окне результатов, см. ниже). Частные корреляции являются корреляциями между соответствующей независимой переменной и зависимой переменной, скорректированными относительно других переменных. Таким образом, это корреляция между остатками после корректировки

StatSoft®
Copyright © StatSoft, 1995

относительно независимых переменных. Частная корреляция представляет самостоятельный вклад соответствующей независимой переменной в предсказание зависимой переменной.

Переменные, входящие в уравнение; ЗП: PT_POOR

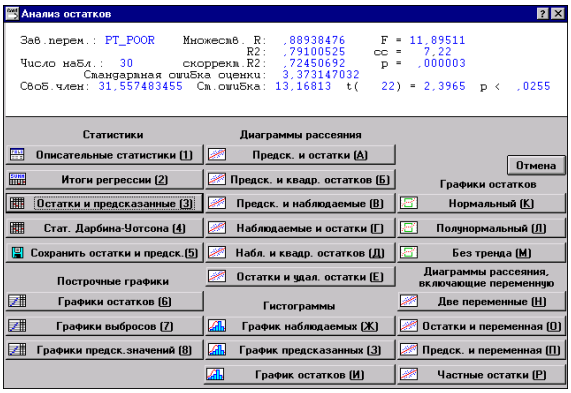
МНОЖЕСТВ РЕГРЕС	Бета (в)	Частная Корр	Получаст Корр	Толеран	R-квадр	t(22)	p-уров
POP_CHNG	-,578434	-,671393	-,414159	,512657	,487343	-4,24924	,000328
N_EMPLD	,614602	,340034	,169718	,076255	,923745	1,74129	,095601
TAX_RATE	,082054	,161974	,075039	,836319	,163681	,76989	,449552
PT_PHONE	-,192963	-,193123	-,089982	,217453	,782547	-,92321	,365914
PT_RURAL	,592775	,480720	,250624	,178758	,821242	2,57138	,017413
AGE	-,176649	-,313128	-,150730	,728071	,271929	-1,54647	,136256
AFR_AM	-,396321	-,240324	-,113184	,081559	,918441	-1,16125	,257984

Получастные корреляция являются корреляциями между соответствующей независимой переменной, скорректированной относительно других переменных, и исходной (нескорректированной) зависимой переменной. Таким образом, получастная корреляция является корреляцией соответствующей независимой переменной после корректировки относительно других переменных, и нескорректированными исходными значениями зависимой переменной. Иначе говоря, квадрат получастной корреляции является показателем процента *общей* дисперсии самостоятельно объясняемой соответствующей независимой переменной, в то время как квадрат частной корреляции является показателем процента остаточной дисперсии, учитываемой после корректировки зависимой переменной относительно независимых переменных.

В этом примере частные и получастные корреляции имеют близкие значения. Однако иногда их величины могут различаться значительно (получастная корреляция всегда меньше). Если получастная корреляция очень мала, в то время как частная корреляция относительно велика, то соответствующая переменная может иметь самостоятельную "часть" в объяснении изменчивости зависимой переменной (т.е. "часть", которая не объясняется другими переменными). Однако в смысле практической значимости, эта часть может быть мала, и представлять только небольшую долю от общей изменчивости (подробнее см., например, в работах Lindeman, Merenda, and Gold, 1980; Morrison, 1967; Neter, Wasserman, and Kutner, 1985; Pedhazur, 1973; или Stevens, 1986).

Анализ остатков

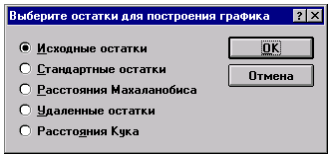
После подбора уравнения регрессии всегда полезно изучить полученные предсказанные значения и остатки. Например, экстремальные выбросы могут существенно сместить результаты и привести к ошибочным заключениям. В диалоговом окне *Результаты множественной регрессии* нажмите кнопку *Анализ остатков* для перехода в соответствующее диалоговое окно.



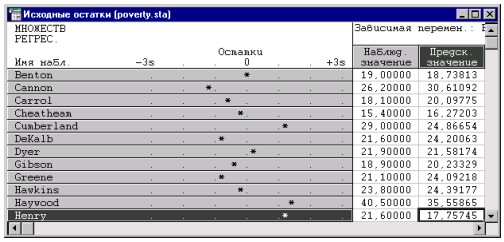
Построчный график остатков. Эта опция диалогового окна предоставляет вам возможность выбрать один из возможных типов остатков для построения построчного графика. Обычно, следует изучить характер исходных (нестандартизованных) или стандартизованных остатков для идентификации экстремальных наблюдений.

В нашем примере, щелкните мышкой по кнопке *Графики остатков* и выберите *Исходные остатки* в открывшемся диалоговом окне *Выберите остатки для построения графика*.

Copyright © StatSoft, 1995
StatSoft®

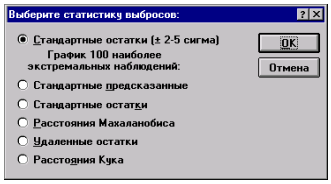


После нажатия кнопки *OK* откроется следующая таблица результатов с построчной диаграммой:



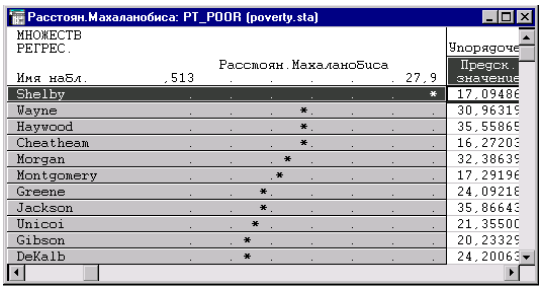
Масштаб, используемый в построчном графике в самой левой колонке, задается в терминах *сигмы*, т.е. стандартного отклонения остатков. Если один или несколько наблюдений попадают за границы ± 3 * *сигма*, то, вероятно, следует исключить соответствующие наблюдения (это легко достигается с помощью условий отбора) и выполнить анализ снова, чтобы убедиться в отсутствии смещения ключевых результатов, вызванного этими выбросами в данных.

Построчный график выбросов. Быстрый способ идентификации выбросов состоит в использовании опции *График выбросов*. Вы можете выбрать просмотр всех стандартных остатков, выпадающих за границы ± 2-5 *сигма*, или просмотр 100 наиболее выделяющихся наблюдений, упорядоченных относительно одной из статистик анализа.



При использовании опции *Стандартные остатки (± 2-5 сигма)* в нашем примере какие-либо выбросы не заметны.

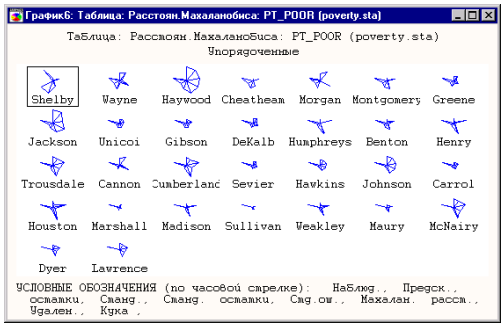
Расстояния Махаланобиса. Большинство учебников по статистике отводят определенное место для обсуждения темы выбросов и остатков для зависимой переменной. Однако роль выбросов для набора *независимых* переменных часто упускается из виду. Со стороны независимых переменных, имеется список переменных, участвующий с различными весами (регрессионные коэффициенты) в предсказании зависимой переменной. Независимые переменные можно представить себе в виде точек некоторого многомерного пространства, в котором может располагаться каждое наблюдение. Например, если вы имеете две независимые переменные с равными регрессионными коэффициентами, то можно построить диаграмму рассеяния этих двух переменных и расположить каждое наблюдение на этом графике. Вы можете затем нарисовать точку средних значений обоих переменных и вычислить расстояния от каждого наблюдения до этого среднего (называемого теперь *центроидом*) в этом двумерном пространстве; в этом состоит концептуальная идея, стоящая за вычислением *расстояний Махаланобиса*. Теперь посмотрим на эти расстояния, отсортированные по величине, с целью идентификации экстремальных наблюдений по независимым переменным. Щелкните мышкой по кнопке *График выбросов* и выберите опцию *Расстояния Махаланобиса* диалогового окна *Выберите статистику выбросов*. Полученный график показывает расстояния Махаланобиса, отсортированные в порядке убывания.



StatSoft

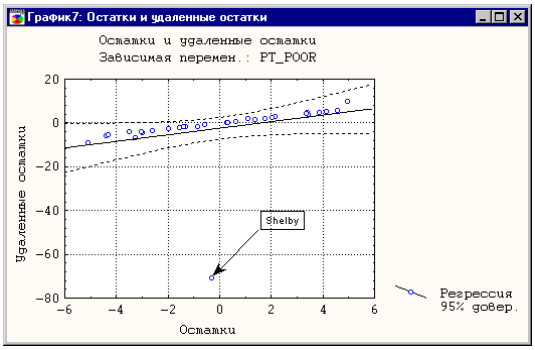
Copyright © StatSoft, 1995

Отметим, что округ Shelby (в первой строчке) оказывается в чем-то выделяющимся по сравнению с другими округами на графике. Если посмотреть на исходные данные, можно обнаружить, что в действительности округ Shelby — значительно больший по размеру округ с большим числом людей, занятых сельским хозяйством (переменная *N_Empld*), и намного более весомой популяцией афроамериканцев. Вероятно, было бы разумно выражать эти числа в процентах, а не в абсолютных значениях, в этом случае расстояние Махаланобиса округа Shelby от других округов в данном примере не было бы столь велико. Однако мы получили, что округ Shelby оказывается явным выбросом. Теперь, щелкнем на таблице результатов с расстояниями Махаланобиса правой кнопкой мыши и выберем опцию *Пиктографик наблюдений* из меню *Быстрые статистические графики*.



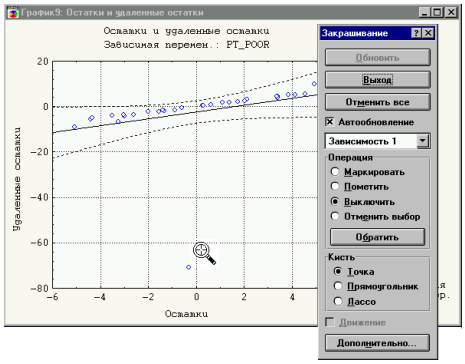
Результирующий пиктографик содержит по одной пиктограмме для каждого изучаемого наблюдения. Индивидуальные компоненты каждой пиктограммы представляют определенными параметры, полученные при регрессионном анализе соответствующего наблюдения (описание этих параметров приводится в нижней части графика в разделе *Условные обозначения*, например, наблюдаемое значение, предсказанное значение, остаток, и др.). Пиктографик помогает графически изобразить расхождения между различными наблюдениями (в нашем случае, округами).

Удаленные остатки. Другой очень важной статистикой, позволяющей оценить масштаб проблемы выбросов, являются *удаленные остатки*. Они определяются как стандартизованные остатки для соответствующих наблюдений, которые получились бы при исключении соответствующих наблюдений из анализа. Напомним, что процедура множественной регрессии подбирает прямую линию для выражения взаимосвязи между зависимой и независимыми переменными. Если одно из наблюдений является очевидным выбросом (как округ *Shelby* в этих данных), то линия регрессии стремиться "приблизится" к этому выбросу, с тем чтобы учесть его, насколько это возможно. В результате, при исключении соответствующего наблюдения, возникнет совершенно другая линия регрессии (и *B*-коэффициенты). Поэтому, если удаленный остаток сильно отличается от стандартизованного остатка, у вас есть основания полагать, что результаты регрессионного анализа существенно смещены соответствующим наблюдением. В данном примере удаленный остаток для округа *Shelby* является выбросом, который существенно влияет на анализ. Вы можете построить диаграмму рассеяния остатков относительно удаленных остатков с помощью опции *Остатки и удаленные остатки*. Ниже на диаграмме рассеяния явно заметен выброс.

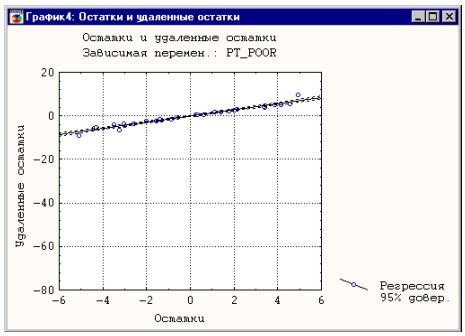


Copyright © StatSoft, 1995
StatSoft®

STATISTICA предоставляет интерактивное средство для удаления выбросов (*Кисть* - на панели инструментов для графики; подробнее см. в томе II). Позволяющее экспериментировать с удалением выбросов и позволяющее сразу же увидеть их влияние на линию регрессии. Когда это средство активизировано, курсор меняется на крестик и рядом с графиком высвечивается диалоговое окно *Кисть* (см. ниже). Вы можете (временно) интерактивно исключать отдельные точки данных из графика, отметив (1) опцию *Автообновление* и (2) поле *Выключить* из блока *Действие*; а затем щелкнув мышкой на точке, которую нужно удалить, совместив ее с крестиком курсора.



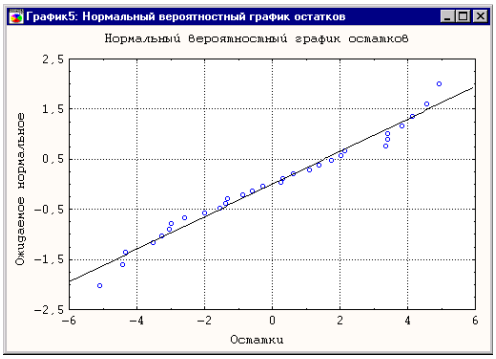
Щелчок на точке будет теперь автоматически удалять ее (временно) из графика, помечать ее (подсвечивать) в *графическом редакторе данных*, так что вы можете исследовать данные и пересчитывать подогнанную регрессию (см. график ниже).



Отметим, что удаленные точки можно "возвратить", щелкнув по кнопке *Отменить выбор* диалогового окна *Закрашивание*.

Нормальные вероятностные графики. Из окна *Анализ остатков* пользователь получает большому количеству дополнительных графиков, каждый из которых кратко обсуждается далее в этой главе начиная со страницы 1695. Большинство этих графиков более или менее просто интерпретируются. Тем не менее, здесь мы дадим интерпретацию нормального вероятностного графика, поскольку он наиболее часто используется при анализе справедливости предположений регрессии.

Как было замечено ранее, множественная линейная регрессия предполагает линейную связь между переменными в уравнении, и *нормальное* распределение остатков. Если эти предположения нарушаются, окончательные заключения могут оказаться неточными. Нормальный вероятностный график остатков наглядно показывает наличие или отсутствие больших отклонений от высказанных предположений. Нажмите кнопку *Нормальный* для построения этого графика.



Этот график строится следующим образом. Сначала остатки регрессии ранжируются. Для этих упорядоченных остатков вычисляются *z*-значения (т.е. стандартные значения нормального распределения), *исходя из предположения*, что данные имеют нормальное

StatSoft
Copyright © StatSoft, 1995

распределение. Эти z-значения откладываются по оси *Y* на графике (подробнее см. в разделе *Примечания*, на стр. 1699).

Если наблюдаемые остатки (отложенные по оси *X*) нормально распределены, то все значения будут располагаться на графике вблизи прямой линии; на данном графике все точки лежат очень близко к прямой линии. Если остатки не распределены нормально, то они будут отклоняться от линии. На этом графике также могут стать заметны выбросы.

Если имеющаяся модель плохо согласуется с данными, и данные на графике, похоже, образуют некоторую структуру (например, облако наблюдений принимает *S*-образную форму) около линии регрессии, то, возможно, будет полезным применение некоторого преобразования зависимой переменной (например, логарифмирование с целью "поджать" хвост распределения, и т.п.; см. также краткое обсуждение преобразований Бокса-Кокса и Бокса-Тидвелла в разделе *Примечания*, стр. 1702). Обсуждение подобных методов лежит за рамками данного руководства (в книге Neter, Wasserman и Kutner, 1985, стр. 134, авторы предлагают превосходное обсуждение преобразований, как средств борьбы с ненормальностью и нелинейностью). Однако слишком часто исследователи просто принимают свои данные, не пытаясь присмотреться к их структуре или проверить их на соответствие своим предположениям, что приводит к ошибочным заключениям. По этой причине одной из основных задач, стоявшей перед разработчиками пользовательского интерфейса модуля *Множественной регрессии* было максимально возможное упрощение (графического) анализа остатков.

Пример 2: Пошаговая регрессия

Файл данных

Следующий пример основан на файле данных *Job_prof.sta* (взятом из работы Neter, Wasserman and Kutner, 1989, стр. 473). Ниже приведена распечатка части этого файла в виде электронной таблицы.

Данные: JOB_PROF.STA 5п * 25н

Job proficiency data set from Neter, Wasserman, & Kutr

	1 TEST1	2 TEST2	3 TEST3	4 TEST4	5 JOB_PROF
1	86	110	100	87	88
2	62	97	99	100	80
3	110	107	103	103	96
4	101	117	93	95	76
5	100	101	95	88	80
6	78	85	95	84	73
7	120	77	80	74	58
8	105	122	116	102	116
9	112	119	106	105	104
10	120	89	105	97	99
11	87	81	90	88	64
12	133	120	113	108	126
13	140	121	96	89	94
14	84	113	98	78	71
15	106	102	109	109	111
16	109	129	102	108	109

Первые четыре переменные (*Test1*-*Test4*) представляют четыре различных теста на пригодность, направленных каждому из 25 претендентов, участвующих в конкурсе на замещение конторских должностей в компании. Все 25 претендентов были приняты на работу вне зависимости от набранных ими баллов по тестам. По истечении испытательного срока рабочие качества каждого из служащих были оценены, и показатель профессиональной пригодности был записан в качестве значений переменной *Job_prof*.

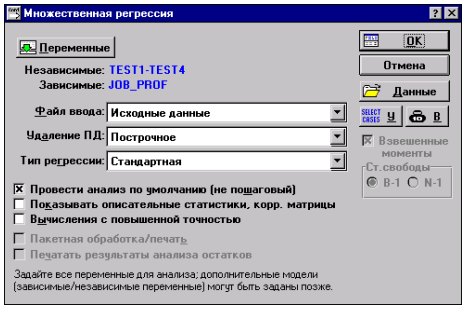
Цель исследования

С использованием пошаговой регрессии мы проанализируем переменные (или подмножество переменных), позволяющие наиболее точно предсказать профессиональную пригодность претендента. Таким образом, зависимой переменной будет *Job_prof*, а переменные *Test1*-*Test4* играют роль независимых переменных.

Copyright © StatSoft, 1995

Начало анализа

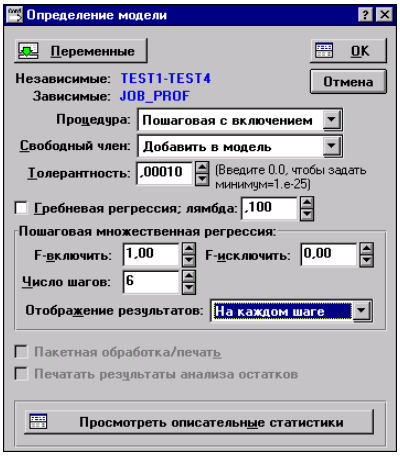
После запуска модуля *Множественной регрессии*, откройте файл данных *Job_prof.sta*. В стартовой панели *Множественной регрессии*, щелкните на кнопке *Переменные* и выберите переменную *Job_prof* в качестве зависимой переменной, а переменные *Test1-Test4* в качестве независимых.



После этого отмените выбранную по умолчанию опцию *Провести анализ по умолчанию (не пошаговый)* и нажмите *OK* для перехода в следующее диалоговое окно *Определение модели*.

Определение модели для пошаговой регрессии

Вы можете выбрать следующие процедуры для проведения анализа данных: *стандартная*, *пошаговая с включением* или *пошаговая с исключением*. Наиболее часто используемой процедурой является процедура с *пошаговым включением предикторов,* которая осуществляет выбор предикторов на каждом шаге, добавляя или удаляя их из модели, исходя из заданного пользователем критерия (более подробную информацию см. в работе Neter, Wasserman and Kutner, 1989, а также d разделе *Примечания*, стр. 1700). Поэтому мы используем пошаговую регрессию с включением предикторов для анализа данных нашего примера.

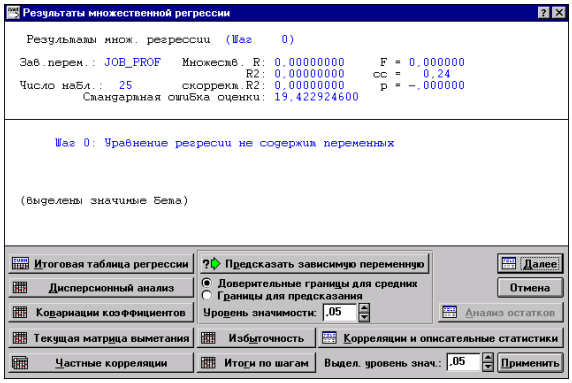


В диалоговом окне *Определение модели*, щелкните мышкой по комбинированному окну *Процедура* и выберите *Пошаговая с включением*. Свободный член будет включаться в модель, а, следовательно, установите опцию *Свободный член* равной *Добавить в модель*. Теперь, после того, как стали доступными опции панели *Пошаговая множественная регрессия*, вы можете изменить здесь значения F-включить – параметр *F- включить* и F-исключить - *F-исключить*. Однако, для проведения анализа в данном примере мы примем значения по умолчанию, равные *1.0* и *0* соответственно. Для просмотра результатов на каждом шаге анализа выберите опцию *На каждом шаге* в комбинированном окне *Отображение результатов*. После этого, оставив все остальные установки без изменений, нажмите *OK* для запуска пошаговой регрессии с включением предикторов.

Шаг 0

Сначала на экран будет выведено диалоговое окно для модели, в которую не включена ни одна независимая переменная.

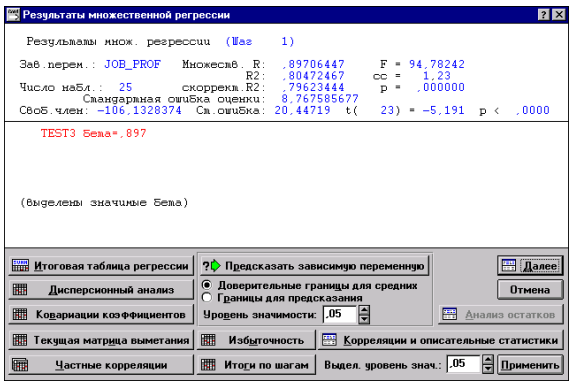
StatSoft
Copyright © StatSoft, 1995



На этом этапе вы можете просмотреть описательные статистики (нажмите *Корреляции и описательные статистики*) или корреляции переменных, не включенных в модель (нажмите *Частные корреляции*).

Шаг 1

Щелкните мышкой по кнопке *Далее*, чтобы приступить к следующему шагу анализа. На первом шаге производится раздельное оценивание каждой независимой переменной, и переменная, имеющая наибольшее значение *F-критерия*, которое к тому же больше или равно значения *F- включить*, добавляется в уравнение регрессии.

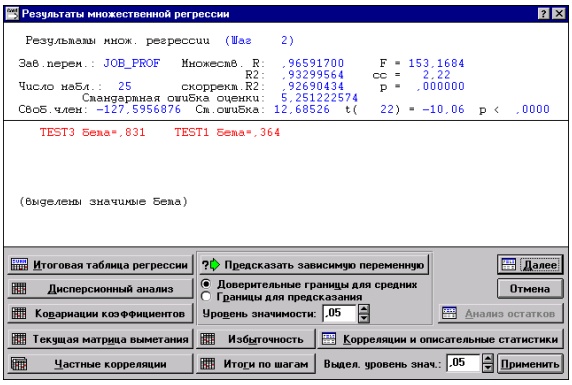


Здесь переменная *Test3* проходит значение *F-включить* (*F*>1.0) и добавляется в модель. Щелкните мышкой по кнопке *Итоги по шагам* для отображения таблицы результатов с итоговой информацией по выполненным шагам регрессионного анализа.

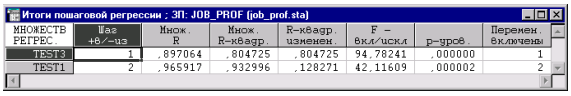
Щелкните на кнопке *Далее* диалогового окна *Результаты множественной регрессии* для выполнения следующего шага.

Шаг 2

Далее, на последующих шагах, при добавлении переменной в модель (исходя из значения *F-включить*), процедура пошаговой регрессии с включением предикторов будет исследовать переменные, включенные в модель, и, исходя из *значения F-исключить*, будет определять, следует ли удалить из модели какую-либо переменную.



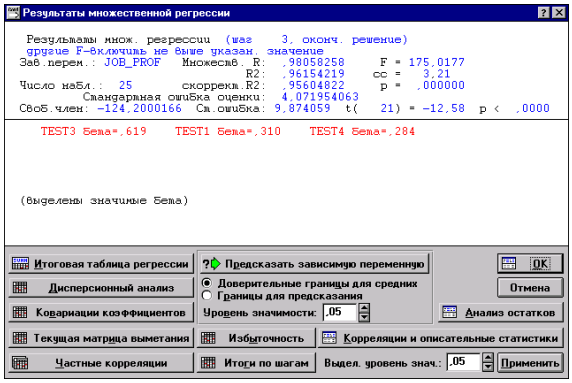
На этом шаге, переменная *Test1* включается в модель. Щелчок мышкой по кнопке *Итоги по шагам* откроет следующее диалоговое окно с итоговыми результатами.



И снова, нажмите кнопку *Далее* в диалоговом окне *Результаты множественной регрессии* для

Copyright © StatSoft, 1995
StatSoft®

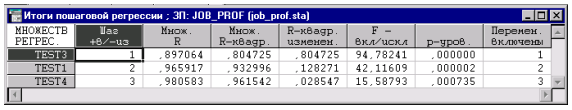
перехода к шагу 3 в пошаговом анализе с включением предикторов.



Шаг 3 (Окончательное решение)

Осталось две переменные, которые необходимо просчитать (*Test2* и *Test4*). На этом шаге наибольшее значение *F-критерия* соответствует переменной *Test4*, поэтому, она добавляется в модель. При расчете переменной *Test2*, значение *F-критерия* оказалось меньше, чем значение *F-включить*, равное *1.0*, следовательно, эта переменная не может быть включена в модель.

Таблица результатов *Итоги по шагам* теперь содержит результаты по всем переменным, которые были добавлены и оставлены в модели.



МНОЖЕСТВ РЕГРЕС	Шаг +В/-Из	Множ. R	Множ. R-квадр.	R-квадр. изменен.	F - вкл/искл	p-уров.	Перемен. включены
TEST3	1	,897064	,804725	,804725	94,78241	,000000	1
TEST1	2	,965917	,932996	,128271	42,11609	,000002	2
TEST4	3	,980583	,961542	,028547	15,58793	,000735	3

Тогда в соответствии с процедурой *пошаговой регрессии с включением предикторов* подмножество испытательных тестов (независимых переменных), которое наилучшим образом предсказывает оценки профессиональной пригодности (зависимую переменную), содержит тесты *Test3*, *Test1* и *Test4*. Поэтому регрессионное уравнение будет следующим:

$$y = B_0+B_1*X_3+B_2*X_1+B_3*X_4$$

Для получения регрессионных коэффициентов из таблицы результатов с итоговой информацией для регрессии нажмите *Итоговая таблица регрессии*.

	БЕТА	Ст. ош. БЕТА	B	Ст. ош. B	t(21)	p-уров.
Св.член			-124,200	9,874059	-12,5784	,000000
TEST3	,618670	,069224	1,357	,151832	8,9373	,000000
TEST1	,309670	,045646	,296	,043679	6,7841	,000001
TEST4	,284405	,072035	,517	,131054	3,9482	,000735

Окончательно, полученное регрессионное уравнение имеет следующий вид:

$$y = -124.200+1.357*X_3+0.296*X_1+0.517*X_4$$

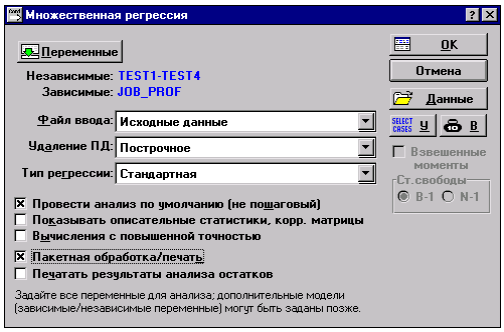
StatSoft

Copyright © StatSoft, 1995

ДИАЛОГОВЫЕ ОКНА, ОПЦИИ, СТАТИСТИКИ

Стартовая панель

После запуска модуля *Множественной регрессии* на экране появляется *стартовая панель* модуля множественной регрессии, в которой вы можете определить переменные для анализа, тип файла данных и установить другие опции определяющие ход проведения анализа (см. ниже).



Переменные

Кнопка вызывает стандартное диалоговое окно выбора переменных, в котором можно выбрать переменные для регрессионного анализа (корреляционная матрица будет вычислена для всех выбранных переменных).

Файл ввода

Выберите в качестве файла входных данных исходные данные или корреляционную матрицу. Корреляционную матрицу можно создать ранее или в самом модуле *Множественная регрессия*, или с помощью опции *Быстрые основные статистики* (см. главу 10), а также в других модулях *STATISTICA* (например, в модуле *Основные статистики и таблицы*, глава 9; *Факторный анализ*, том III; *Канонический анализ*, том III или *Надежность и позиционный анализ*, том III). Корреляционную матрицу можно также ввести в электронную таблицу с помощью клавиатуры (см. описание *Формата матричного файла* в системе *STATISTICA*, глава 7).

Удаление пропущенных данных

Это поле активно только при использовании файла исходных данных (см. выше). Пропущенные данные могут удаляться *построчно*, заменяться *средними значениями* или удаляться *попарно* при выборе соответствующей опции в поле *Удаление ПД*.

Построчное удаление пропущенных данных. При выборе этой опции при проведении анализа используются только те наблюдения, которые не имеют пропущенных значений в выбранных для анализа переменных.

Замена средним. При выборе этой опции пропущенные значения в каждой переменной заменяются средним, вычисленным по имеющимся наблюдениям соответствующей переменной (только при проведении анализа, а не в самом файле данных).

Попарное удаление пропущенных данных. Если выбрана эта опция, то при вычислении парных корреляций удаляются наблюдения, имеющие пропущенные значения в соответствующих переменных. При проведении последующего анализа все статистические критерии будут основываться на минимальном числе допустимых наблюдений, обнаруженном во всех выбранных переменных.

Тип регрессии

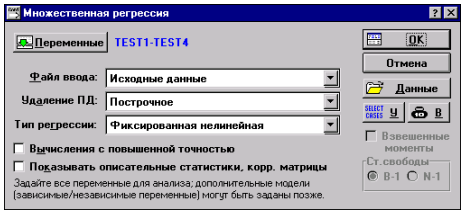
Пользователь может выбрать *стандартную* или *фиксированную нелинейную* регрессию.

Copyright © StatSoft, 1995

StatSoft®

Стандартная регрессия. По умолчанию выбрано проведение анализа с помощью *стандартной* множественной регрессии. При этом вычисляется стандартная корреляционная матрица всех выбранных переменных.

Фиксированная нелинейная регрессия. При выборе этого типа регрессии стартовая панель принимает следующий вид:



Этот тип регрессии позволяет использовать различные преобразования независимых переменных (см. *Фиксированная нелинейная регрессия*, стр. 1679). Таким образом, новые переменные будут созданы, как преобразования исходных, и включены в общую корреляционную матрицу. Обычно этот прием применяется для введения в уравнение регрессии линейных, квадратичных, кубических и др. членов. Заметим, что *STATISTICA* также включает модуль *Нелинейного оценивания* (том III) для оценивания моделей нелинейной регрессии любой сложности (включая *пробит* и *логит* регрессию и др.).

Провести анализ по умолчанию

После выбора этой опции и нажатия кнопки *OK* на стартовой панели, программа использует для анализа установки, принятые по умолчанию (т.е. *Стандартная* регрессионная модель, включающая свободный член) в диалоговом окне *Определение модели* и перейдет в диалоговое окно *Результаты регрессии*. Если эта опция отменена, то при щелчке мышью на кнопке *OK* стартовой панели откроется диалоговое окно *Определение модели*, в котором вы можете выбрать как тип регрессионного анализа (например, пошаговый, гребневый и др.), так и другие опции.

Показывать описательные статистики и корреляционную матрицу

Установите этот переключатель, чтобы открыть диалоговое окно *Просмотр описательных статистик* (стр. 1680). Хотя диалоговое окно *Просмотр описательных статистик* доступно на любом этапе анализа, при выборе этой опции из стартовой панели описательные статистики будут основываться на *попарном* исключении пропущенных данных, если выбрана опция *Попарное*. В диалоговом окне *Просмотр описательных статистик* вы можете просмотреть подробные описательные статистики (в том числе число пар наблюдений, по которым был вычислен коэффициент корреляции для каждой пары переменных). Чтобы продолжить анализ и открыть диалоговое окно *Определение модели*, нажмите *OK* в диалоговом окне *Просмотр описательных статистик*.

Вычисления с повышенной точностью

Установите этот переключатель для использования более точного алгоритма вычисления корреляционной матрицы, используемой при проведении множественной регрессии. Эту опцию следует установить, если анализируемые переменные имеют чрезвычайно малую *относительную* дисперсию (дисперсия, деленная на среднее), например, если все значения переменной находятся в диапазоне от 10000.000001 до 10000.000002. Отметим, что имеется разница между малой дисперсией и малой *относительной* дисперсией (например, *относительная* дисперсия значений из диапазона от .0000000000000001 до .0000000000000002 не является малой). Для

StatSoft
Copyright © StatSoft, 1995

крайне больших файлов данных с несколькими тысячами наблюдений вычисления при установке этого переключателя могут быть несколько медленнее. Выполняемые в процессе *Множественной регрессии* вычисления с двойной точностью оптимальны в смысле достижения максимальной точности, так что почти для всех наборов данных, полученных с помощью реальных измерений (т.е. не созданных искусственно), результаты этих вычислений идентичны результатам, полученным с использованием опции повышенной точности вычислений.

Пакетная обработка и печать

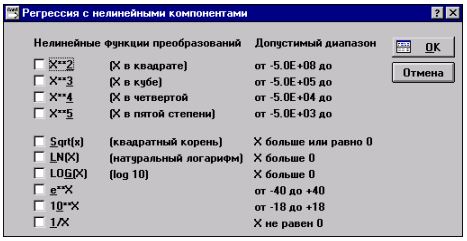
Эта опция доступна только в том случае, когда установлен переключатель *Выполнить анализ по умолчанию* (см. выше) и в диалоговом окне *Параметры страницы/вывода* выбран *принтер*, *файл на диске*, и/или *Окно текста/вывода* (щелкните на поле *Вывод* строки состояния, чтобы открыть это диалоговое окно). Выбор опции *Пакетная обработка/печать* определяет способ печати выходных результатов регрессии. По умолчанию выбрана *интерактивная обработка* (т.е. переключатель сброшен). В этом режиме вы можете (интерактивно) определять, какую из таблиц результатов следует печатать. В режиме *Пакетная обработка/печать* таблицы результатов автоматически печатаются, посылаются в файл принтера и/или в *Окно текста/вывода* (в зависимости от того, что выбрано в окне *Параметры страницы/вывода*). Если эта опция установлена, программа выполнит заданный регрессионный анализ (или набор анализов, если задано более одной зависимой переменной), “распечатает” результаты и возвратится в диалоговое окно с заголовком *Определение модели*.

Печатать результаты анализа остатков

Эта опция доступна только в том случае, если установлен переключатель *Провести анализ по умолчанию* (см. выше) и в диалоговом окне *Параметры страницы/вывода* выбран *принтер*, *файл на диске* и/или *окно текста/вывода* одновременно с выбранной опцией *Пакетная обработка/печать* (см. выше). Если выбрана эта опция, распечатанные результаты по умолчанию будут включать таблицу результатов с различными статистиками остатков по каждому наблюдению.

Фиксированная нелинейная регрессия

Диалоговое окно *Регрессия с нелинейными компонентами* доступно только после выбора в стартовой панели опции *Фиксированная нелинейная* регрессия. Это диалоговое окно открывается после нажатия кнопки *OK* в стартовой панели.



Данное окно позволяет задавать нелинейные преобразования выбранных переменных. Преобразования помечаются флажком рядом с соответствующей функцией. Доступные преобразования и допустимые диапазоны показаны в окне выше. Можно выбрать несколько преобразований.

Для каждой выбранной переменной создаются (в оперативной памяти компьютера, а не в исходном файле данных) новые временные переменные, содержащие результаты

Copyright © StatSoft, 1995

преобразований (новая переменная будет создана для каждого выбранного преобразования по каждой выбранной переменной). Имя новой переменной будет являться комбинацией номера соответствующей исходной переменной и выбранного преобразования, например, *V4**2*, *10**V7* и т.д.

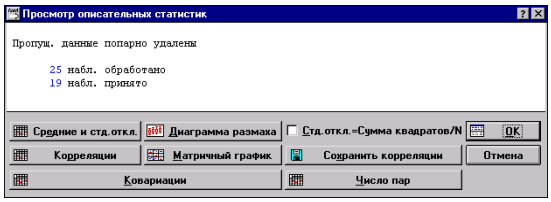
Код *пропущенного значения* по умолчанию (-9999) будет назначен значениям новых (преобразованных) переменных, если:

(a) Исходная (не преобразованная) переменная содержит пропущенное значение для того же наблюдения;

(b) Значение наблюдения не попадает в допустимый диапазон.

Просмотр описательных статистик

Промежуточное диалоговое окно позволяет просмотреть определенные описательные статистики для ранее выбранных переменных (и любых производных переменных), включая таблицу с числом пар наблюдений, по которым вычислен каждый коэффициент корреляции (если в стартовой панели было выбрано *Попарное удаление ПД*).



Средние и стандартные отклонения

Эта опция выводит таблицу результатов с выборочными средними и стандартными отклонениями. Выборочное среднее (или просто среднее) является несмещенной *оценкой* среднего популяции. По умолчанию, стандартное отклонение вычисляется как корень квадратный из суммы квадратов отклонений значений переменной от выборочного среднего, деленный на *n*-1. Полученная в результате статистика является оценкой *сигмы* популяции (эта статистика позволяет сделать выводы о разбросе значений переменной).

Вы можете использовать опцию *Стд.откл. = Сумма квадратов/n* (см. ниже) для вычисления стандартного отклонения как суммы квадратов, деленной на число допустимых наблюдений *n*. Получаемая статистика является выборочным значением стандартного отклонения *на данной выборке* (по этой статистике нельзя сделать никаких выводов о популяции, из которой была получена выборка).

Графиком, выбранным по умолчанию для этой таблицы, является гистограмма частот с наложенным графиком нормального распределения. (Эта опция недоступна для *Фиксированной нелинейной регрессии*, стр. 1679) График может быть очень полезен для быстрой проверки предположения о нормальности соответствующей переменной.

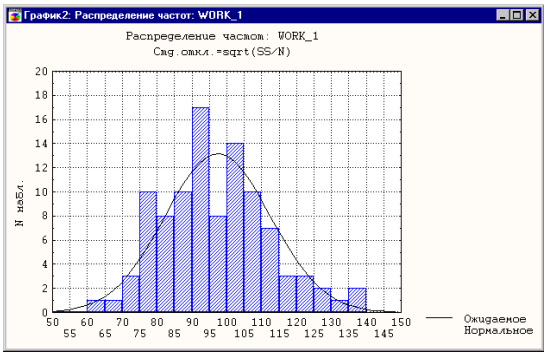
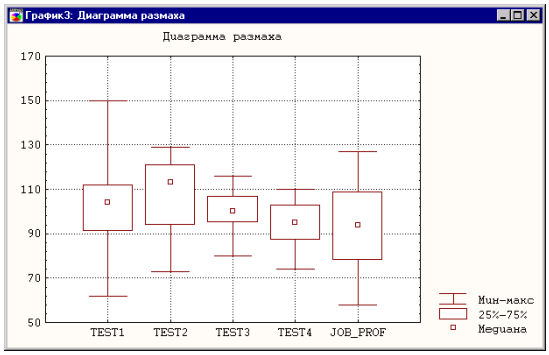


Диаграмма размаха

Эта опция позволяет строить диаграммы размаха (на котором показаны медианы и квартили или средние и стандартные отклонения и пр.) для выбранных переменных.

StatSoft®
Copyright © StatSoft, 1995

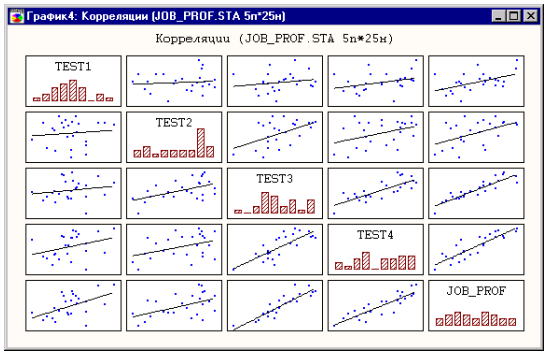


Корреляции

Эта опция выводит таблицу результатов с корреляциями. Графиком, выбранным по умолчанию для этой таблицы, является диаграмма рассеяния для каждой пары переменных. На графике показана как подогнанная методом наименьших квадратов прямая регрессии, так и 95% доверительные границы для математического ожидания зависимой переменной при данном значении независимой переменной. Данные доверительные границы представляют собой гиперболы.

Матричный график

Эта опция строит матрицу диаграмм рассеяния для выбранных переменных.



Ковариации

Эта опция строит таблицу результатов с ковариациями. Графиком, выбранным по умолчанию для этой таблицы результатов, является диаграмма рассеяния соответствующих двух переменных. На графике показана как подогнанная регрессионная прямая, так и 95% доверительные границы для математического ожидания зависимой переменной при данном значении независимой переменной. Данные доверительные границы представляют собой гиперболы.

Сохранить корреляции

При нажатии этой кнопки открывается диалоговое окно *Сохранить матрицу как*, в котором вы можете задать сохранение корреляционной матрицы в виде стандартного матричного файла системы *STATISTICA* (см. *Формат матричного файла*, глава 7).

Число пар

Опция доступна только при анализе *исходных данных* с *попарным удалением пропущенных значений*. Отображается число пар наблюдений, по которым построен каждый коэффициент корреляции.

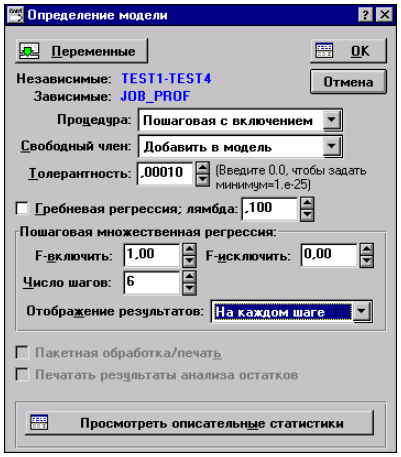
Стд.откл. = *Сумма квадратов/n*

Этот переключатель влияет на метод вычисления стандартных отклонений (*Средние и стандартные отклонения)* и дисперсий/ковариаций (опция *Ковариации*). Если этот переключатель установлен, полученные статистики будут *описывать* стандартные отклонения, дисперсии и ковариации в *изучаемой выборке* (см. *Средние и стандартные отклонения*, выше). Установка этого переключателя влияет только на соответствующие значения в таблицах результатов. Она не влияет на какие-либо последующие вычисления.

Copyright © StatSoft, 1995

Определение модели

Выберите независимые и зависимые переменные для множественного регрессионного анализа в диалоговом окне *Определение модели*.



При выборе нескольких зависимых переменных программа будет запускаться для каждой переменной из списка зависимых переменных.

Переменные

Выбор этой опции приведет к появлению стандартного диалогового окна для выбора двух переменных, в котором выбираются зависимые и независимые переменные. Если выбрано более одной зависимой переменной, то регрессионный анализ выполняется последовательно для каждой переменной из списка зависимых переменных.

Процедура

Эта опция позволяет выбрать вид регрессионного анализа.

Стандартная регрессия. При выборе этой опции все переменные будут включены в уравнение регрессии одним блоком (т.е. на одном шаге итерации).

Пошаговая регрессия с включением предикторов. При выборе этой опции независимые переменные будут по отдельности включаться или исключаться из модели на каждом шаге регрессии (в зависимости от вашего выбора для значений *F - включить* или *F - исключить*, см. ниже) до тех пор, пока не будет получена "наилучшая" регрессионная модель.

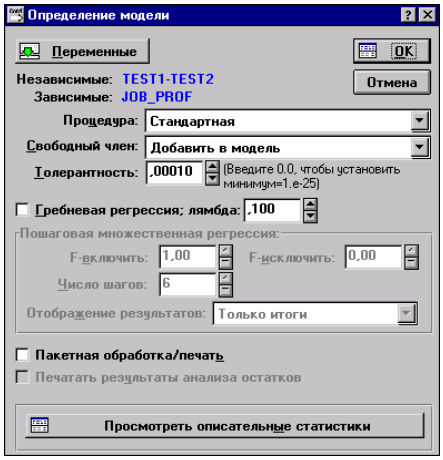
Пошаговая регрессия с исключением предикторов. При выборе этой опции независимые переменные будут исключаться из модели по одной переменной на шаге (в зависимости от вашего выбора для значений *F - включить* или *F - исключить*, см. ниже) до тех пор, пока не будет получена "наилучшая" регрессионная модель.

Пошаговая регрессия по блокам (иерархическая регрессия)

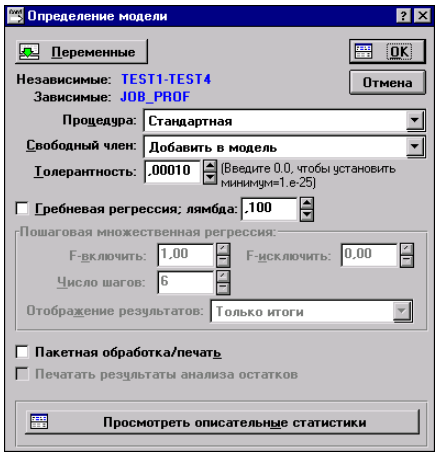
Модуль *Множественной регрессии* автоматически следит за последовательно проводимыми анализами. При выборе *стандартной* регрессии программа автоматически оценивает *R-квадрат,* а также вычисляет увеличение или уменьшение *R-квадрат* для последовательных шагов анализа *одной и той же* зависимой переменной.

Например, выберите в стартовой панели *Множественной регрессии* все представляющие интерес независимые переменные и отмените опцию *Провести анализ по умолчанию* (например, для файла данных *Job_prof.sta* выберите в качестве зависимой переменной *Job_prof*, а в качестве независимых переменных *Test1-Test4*). Нажмите *OK* для перехода в окно *Определение модели*. В этом окне нажмите кнопку *Переменные* и выберите подмножество интересующих вас независимых переменных (пусть это будут, например, переменные *Test1-Test2*).

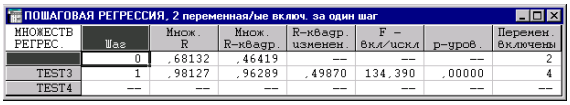
StatSoft®
Copyright © StatSoft, 1995



Теперь нажмите *OK* и в открывшемся диалоговом окне *Результаты множественной регрессии*, нажмите *Отменить* для возврата в диалоговое окно *Определение модели*. Вновь нажмите *Переменные* и выберите (наряду с ранее выбранными независимыми переменными) дополнительные независимые переменные (в итоге вы выберите *Test1-Test4*).



Нажмите *OK* для перехода в диалоговое окно *Результаты множественной регрессии*. Программа автоматически оценит *R-квадрат* и вычислит приращения *R-квадрата*, вызванное включением дополнительных независимых переменных (*Test3* и *Test4*). Нажмите *Итоги по шагам* для вызова таблицы результатов с этими значениями.


ПОШАГОВАЯ РЕГРЕССИЯ, 2 переменная/ые включ. за один шаг

МНОЖЕСТВ РЕГРЕС.	Шаг	Множ. R	Множ. R-квадр.	R-квадр. изменен.	F - вкл/искл	p-уров.	Перемен. включены
	0	,68132	,46419	--	--	--	2
TEST3	1	,98127	,96289	,49870	134,390	,00000	4
TEST4	--	--	--	--	--	--	--

Свободный член

Эта опция позволяет задать регрессионное уравнение, *включающее* свободный член (если выбрана опция *Добавить в модель*), или *не содержащее* свободный член (свободный член приравнивается к нулю и регрессия проходит через начало координат, если выбрана опция *Установить равными нулю*). Регрессия без свободного члена часто используется при анализе экономических данных в случаях, когда, по определению, линия регрессии, описывающая связь между некоторыми переменными, должна проходить через начало координат.

Например, если изучается зависимость дохода от сбора налогов и валового национального продукта (ВНП), то, очевидно, что при нулевом ВНП, доход от налогов также равен нулю. Однако, в большинстве приложений (в частности, в общественных и естественных науках) интересующие переменные измеряются в более или менее произвольных шкалах и нулевая отметка не имеет особого значения. Поэтому, по умолчанию, в поле *Свободный член* выбрано значение *Добавить в модель*.

Толерантность

Толерантность переменной определяется как 1 минус квадрат множественной корреляции этой переменной со всеми другими независимыми переменными уравнения регрессии. Поэтому, чем меньше толерантность переменной, тем в большей степени ее вклад в регрессию является избыточным (т.е. она является избыточной при заданных значениях других независимых переменных). Если толерантность каких-либо переменных в уравнении регрессии равна нулю

Copyright © StatSoft, 1995

(или очень близка к нулю), то оценки не могут быть вычислены (так как получаемая матрица плохо обусловлена и не может быть обращена).

Опция позволяет задать допустимое при проведении анализа значение толерантности. Минимальное допустимое значение этого параметра равно $1.00E^{-25}$ (т.е. число с 24 нулями после десятичной точки). Однако не рекомендуется устанавливать этот параметр на такое экстремально малое значение. Если толерантность переменной на входе в регрессионное уравнение меньше установленного по умолчанию значения толерантности (*.0001*), то это означает, что эта переменная на 99 процентов является излишней (идентичной) для переменных, уже включенных в уравнение. Принудительное включение чрезвычайно избыточных переменных в регрессионное уравнение не только сомнительно с точки зрения уместности получаемых результатов, но и приводит к очень ненадежным оценкам.

F-включить, F-исключить

Эти опции доступны только в случае выбора пошаговой регрессии с *включением* или *исключением* предикторов (см. выше). Значение *F-включить* заданное здесь, определяет, насколько значимым должен быть вклад переменной в регрессию, чтобы она была добавлена в уравнение. Значение *F-исключить*, определяет насколько "незначимым" должен быть вклад переменной в уравнение регрессии, чтобы она могла быть исключена из регрессионного уравнения. Значимость определяется для изменения F-критерия при изменении набора переменных (включении/исключении переменной).

Если в регрессии с *пошаговым включением предикторов* желательно принудительно включить все (или почти все) переменные в уравнение (по одной на каждом шаге), то значение *F-включить* должно быть установлено минимальным (*0.0001*), а значение *F-исключить* должно быть установлено на свой минимум (*0.0*; *F-исключить* должно всегда быть меньше *F-включить*).

Если в регрессии с *пошаговым исключением предикторов* желательно удалить все переменные из уравнения (по одной на каждом шаге), то значение *F- включить* должно быть установлено равным очень большому числу (например, 999), а значение *F-исключить* должно быть установлено на сходное по величине значение (например, 998; напомним, что значение *F-включить* всегда должно быть меньше, чем значение *F-исключить*).

Число шагов

Эта опция доступна только в случае выбора пошаговой регрессии с *включением* или *исключением* предикторов (см. выше). Эта опция позволяет пользователю выбирать максимальное число шагов, которое процедура оценивания выполнит до вывода окончательных результатов пошагового регрессионного анализа.

Отображение результатов

Эта опция влияет на отображение результатов пошагового (с *включением* или *исключением* предикторов) множественного регрессионного анализа. Вы можете выбрать для просмотра только окончательные (итоговые) результаты пошагового регрессионного анализа (выберите *Только итоги*) или результаты каждого шага анализа (выберите *На каждом шаге*).

Гребневая регрессия

Гребневая (или ридж) регрессия используется, когда независимые переменные очень сильно коррелируют друг с другом, и, поэтому, устойчивые оценки регрессионных

StatSoft®

Copyright © StatSoft, 1995

коэффициентов не могут быть получены с помощью обычного метода наименьших квадратов (см. книги Hoerl, 1962; Schmidt и Muller, 1978; Rozeboom, 1979). А именно, константа (*лямбда*) добавляется к диагонали корреляционной матрицы, которая затем перестандартизуется, так чтобы все диагональные элементы были равны 1.0 (при этом внедиагональные элементы делятся на некоторую константу).

Другими словами, гребневая регрессия искусственно занижает коэффициенты корреляции, так, что могут быть вычислены более устойчивые (хотя и смещенные) оценки (*бета-коэффициенты*). Заметим, что стандартные ошибки для *B* и *бета-коэффициентов* вычисляются с использованием формул стандартной регрессии. После выбора этой опции нужно ввести значение константы *лямбда*.

Пакетная обработка и печать

Эта опция доступна только в том случае, если в окне *Параметры страницы/вывода* была выбрана опция печати на *принтер*, *в файл на диске* и/или в *окно текста/вывода* (щелкните на поле *Вывод* строки состояния, чтобы открыть это диалоговое окно). Выбор опции *Пакетная обработка/печать* определяет способ распечатки результатов регрессии.

По умолчанию, выбрана *интерактивная обработка* (т.е. переключатель *Пакетная обработка/печать* сброшен). В этом режиме вы можете (интерактивно) определять, какую из таблиц результатов следует печатать. В режиме *Пакетной обработки/печати* таблицы результатов автоматически печатаются, посылаются в файл принтера и/или в *окно текста/вывода* (в зависимости от того, что было выбрано в окне *Параметры страницы/вывод*).

Если эта опция установлена, программа выполняет заданный регрессионный анализ (или последовательность анализов, если задано более одной зависимой переменной), “распечатывает” результаты и возвращает управление окну *Определение модели*. Эта опция особенно полезна, если желательно распечатать результаты пошаговой регрессии после каждого шага для нескольких зависимых переменных.

Печатать результаты анализа остатков

Эта опция доступна только в том случае, если в диалоговом окне *Параметры страницы/вывода* была выбрана печать *на принтер*, *в файл на диске* и/или в окно *текста/вывода*, а также выбрана опция *Пакетная обработка/печать* (см. выше). Если эта опция отмечена, распечатка по умолчанию будет включать таблицы с различными статистиками для остатков по каждому наблюдению.

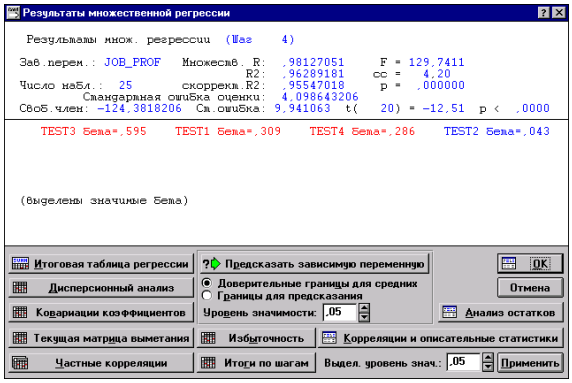
Просмотреть описательные статистики

Нажмите эту кнопку для возврата в диалоговое окно *Просмотр описательных статистик*, в котором вы можете выбрать просмотр средних и стандартных отклонений, корреляционной и ковариационной матрицы, а также сохранить корреляционную матрицу в стандартном матричном файле системы *STATISTICA*.

Результаты множественной регрессии

Диалоговое окно *Результаты множественной регрессии* показывает сводку результатов текущего регрессионного анализа и предоставляет возможность просмотра некоторых специфических результатов.

Copyright © StatSoft, 1995



Информационное поле

Следующие результаты выводятся в этой части диалогового окна:

Зависимая переменная. Отображается имя зависимой переменной.

Число наблюдений. Отображается число наблюдений, принятых к обработке (*n*) (зависит от выбранного способа обработки *пропущенных данных*).

Множественный *R*. Отображается коэффициент множественной корреляции (между зависимой и независимыми переменными).

***R*-квадрат.** Отображается коэффициент множественной детерминации, определяемый как:

```
R-квадрат = 1 - (SS остатков/Общая SS)
```

Равен квадрату коэффициента множественной корреляции, показывает, какую долю дисперсии зависимой переменной объясняет уравнение регрессии.

Скорректированный *R*-квадрат. *R-квадрат* корректируется делением суммы квадратов ошибок и общей суммы квадратов на соответствующее им число степеней свободы.

```
R-квадрат (скорректированный) =
      1-[(SS остатков/cc)/(Общая SS/cc)]
```

Стандартная ошибка оценки. Эта величина измеряет рассеяние наблюдаемых значений относительно линии регрессии.

Свободный член. Если выбрана регрессия, включающая свободный член (см. *Определение модели*), то будет приведена оценка свободного члена.

Стандартная ошибка. Отображается стандартная ошибка оценки свободного члена.

***t*(число ст.св.) и *p*-значение.** *t*-статистика и соответствующее ей *p*-значение используются для проверки гипотезы о равенстве нулю свободного члена в уравнении регрессии.

***F-критерий*, ст.св. и *p*-значение:** Статистика *F*-критерия и соответствующий уровень *p* используются в качестве общего *F*-критерия для проверки гипотезы о зависимости предикторов и отклика. Имеем:

```
F = MS регрессии/MS остатков
```

Поле параметров модели

В этой части окна отображаются и выделяются статистически значимые коэффициенты регрессии (*бета-коэффициенты*) для переменных, включенных в анализ. Критерий статистической значимости (выделяемый уровень значимости – *альфа*) может быть выбран из интервала [0.0001,0.5). По умолчанию он равен *.05* (см. ниже).

Итоговая таблица регрессии

Эта опция строит таблицы результатов со стандартизованными (*бета*) и нестандартизованными (*B*) регрессионными коэфициентами (весами), их стандартными ошибками и уровнями значимости.

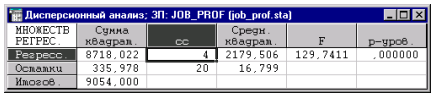
StatSoft

Copyright © StatSoft, 1995

[Стандартизованные коэффициенты оцениваются по стандартизованным данным, имеющим выборочное среднее 0 и дисперсию 1.] Итоговые статистики регрессионного анализа (в том числе *R*, *R-квадрат* и др.) отображаются в заголовке таблицы.

Дисперсионный анализ

Нажатие этой кнопки строит полную таблицу дисперсионного анализа (ANOVA) для данного регрессионного уравнения.



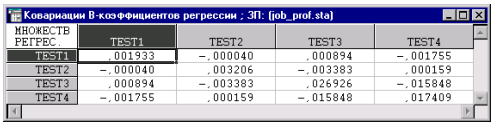
Дисперсионный анализ, скорректированный на среднее

Опция доступна только, когда регрессионная модель не включает свободный член. В этом случае можно вычислить значение множественного R-квадрата, основываясь или на вариации относительно нуля или относительно среднего. По умолчанию, значение R-квадрата, приводимое в итоговом окне, соответствует первому случаю, т.е. равно доле вариации зависимой переменной относительно нуля, объясняемой предикторами (см. формулу для R-квадрата выше). Если вы нажмете данную кнопку, *STATISTICA* построит таблицу дисперсионного анализа, скорректированного на среднее (при вычислении сумм квадратов будет вычитаться выборочное среднее). По поводу альтернативных способов вычисления R-квадрата смотрите работу Kvalseth (1985).

Ковариации коэффициентов

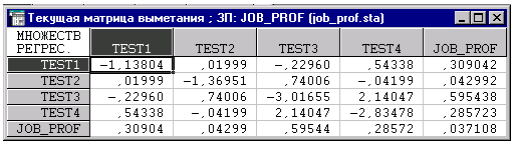
Опция строит 2 таблицы: (1) таблицу с корреляциями регрессионных коэффициентов

(2) таблицу с выборочными дисперсиями (на диагонали) и ковариациями регрессионных коэффициентов.



Текущая матрица выметания

Эта опция строит таблицу результатов с текущей матрицей выметания.



Обращение матрицы в процедуре *Множественной регрессии* производится методом выметания. Матрица выметания для всех независимых переменных, включенных на данном этапе в регрессионное уравнение, равна -1, умноженной на матрицу, обратную к корреляционной матрице этих переменных (умножение на –1 означает, что знак каждого элемента матрицы изменен на противоположный). Диагональные элементы для переменных, не включенных в уравнение, можно интерпретировать как значения (*1 - R-квадрат*), при рассмотрении соответствующей переменной в качестве зависимой и использовании всех текущих независимых переменных в уравнении.

Copyright © StatSoft, 1995

StatSoft®

Частные корреляции

Нажатие этой кнопки открывает таблицу результатов, содержащую для каждой переменной:

(1) Коэффициент *бета (в)* (стандартизованный коэффициент для соответствующей переменной, если бы она была включена в уравнение регрессии как независимая переменная);

(2) *Частную корреляцию* (между соответствующей переменными и зависимой переменной после учета влияния всех остальных независимых переменных в уравнении);

(3) *Получастную корреляцию* (корреляцию между *нескорректированной* зависимой переменной и соответствующей независимой переменной после учета влияния всех остальных независимых переменных в уравнении);

(4) *Толерантность* (определяется как 1 минус квадрат множественной корреляции между соответствующей переменной и всеми независимыми переменными в уравнении регрессии);

(5) *Минимальная толерантность* (наименьшее значение толерантности среди всех независимых переменных в уравнении, которое получилось бы при включении соответствующей переменной в качестве дополнительной независимой переменной);

(6) значение *t*-статистики;

(7) уровень значимости для *t*-статистики.

Эти статистики отображаются отдельно для переменных, *не включенных* в текущее уравнение регрессии, и для переменных, *включенных* в уравнение (если такие имеются).

Итоги по шагам

Эта опция доступна в диалоговом окне *Результаты множественной регрессии* только в том случае, если (1) пользователь выбрал регрессию с *пошаговым включением* или *исключением* предикторов, или если (2) в предыдущем *стандартном* регрессионном анализе исследовалась та же самая зависимая переменная, и прежний список независимых переменных являлся подмножеством текущего списка независимых переменных или наоборот (см. стр. 1682). Таким образом, в последнем случае можно вычислить приращения *R-квадрата*, вызванное исключением или включением нескольких переменных на одном шаге (иерархический анализ). Например, если в первом анализе переменные с номерами от 1 до 5 были выбраны в качестве независимых переменных, а в последующем анализе с *той же зависимой переменной* пользователь задал переменные с номерами от 1 до 3 в качестве независимых переменных, то выбор этой опции приведет к построению таблицы со значениями *R-квадрата*, приращениями *R-квадрата*, *F-критерием для исключения* и числом переменных, удаленных на этом единственном шаге (две переменные в нашем примере). Если была задана регрессия с *пошаговым включением* или *исключением* предикторов, таблица результатов будет содержать приращения *R-квадрата* на каждом шаге.

Предсказать зависимую переменную

Опция открывает окно, в котором можно ввести значения независимых переменных и вычислить (предсказать) значение зависимой переменной с помощью построенного уравнения регрессии.

Избыточность

Эта опция строит таблицу с различными показателями избыточности независимых

StatSoft®

Copyright © StatSoft, 1995

переменных (включенных или не включенных на данном этапе в уравнение регрессии). Для каждой переменной вычисляется:

(1) *толерантность* (определенную как *1 - R-квадрат* соответствующей переменной со всеми остальными переменными, включенными в уравнение),

(2) *R-квадрат* (между текущей переменной и всеми остальными переменными уравнения),

(3) *частную корреляцию* (между соответствующей переменной и зависимой переменной после учета влияния всех остальных независимых переменных в уравнении),

(4) *получастную корреляцию* (корреляцию между *нескорректированной* зависимой переменной и соответствующей переменной после учета влияния всех независимых переменных в уравнении).

Корреляции и описательные статистики

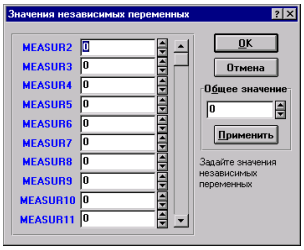
Кнопка открывает диалоговое окно *Просмотр описательных статистик* (стр. 1680). В этом диалоговом окне вы можете выбрать для просмотра таблицы средних и стандартных отклонений, корреляционную и ковариационную матрицы или сохранить корреляционную матрицу в формате матричного файла *STATISTICA*.

Выделяемый уровень значимости

По умолчанию выделяемый уровень значимости (*альфа*) равен *.05*. Если *p*-уровень меньше *альфа*, то этот предиктор выделяется цветом в таблице результатов как значимый. В соответствующем поле диалогового окна вы можете изменить значение выделяемого уровня ($.0001 < \alpha < .5$). Нажмите затем кнопку *Применить*, чтобы построить таблицу с выделением результатов на новом уровне.

Предсказание значений зависимой переменной

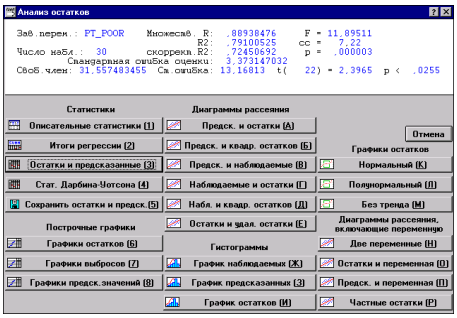
При нажатии кнопки *Предсказать зависимую переменную* окна *Результаты множественной регрессии* открывается диалоговое окно *Значения независимых переменных*.



В этом диалоговом окне можно задать конкретный набор значений, по которому будет вычислено предсказываемое значение зависимой переменной. Введите значения отдельно для каждой переменной текущего регрессионного уравнения или задайте одно общее значение. Затем нажмите кнопку *Применить*, чтобы присвоить его переменным. По введенным значениям независимых переменных *STATISTICA* вычислит предсказываемое значение зависимой переменной исходя из текущего регрессионного уравнения.

Анализ остатков

Это диалоговое окно доступно только при выборе *исходных данных* в качестве входных.



Copyright © StatSoft, 1995

StatSoft®

Корреляции и описательные статистики

Эта опция вызывает диалоговое окно *Просмотр описательных статистик* (стр. 1680), в котором можно просмотреть средние и стандартные отклонения, корреляционные и ковариационные матрицы, а также сохранить корреляционные матрицы в стандартном матричном файле системы *STATISTICA*.

Итоги регрессии

Нажатие этой кнопки строит таблицу результатов со стандартизованными (*Бета*) и нестандартизованными (*B*) коэффициентами уравнения регрессии, их стандартными ошибками и уровнем значимости. Итоговые статистики регрессионного анализа (например, *R*, *R-квадрат* и т.д.) будут выведены в заголовке этой таблицы.

Остатки и предсказанные

Эта опция открывает таблицу результатов с различными статистиками (основанными на остатках) для каждого наблюдения. Графиком, выбранным по умолчанию для этой таблицы, является пиктографик всех наблюдений (см. стр. 1671). Для построения этого графика щелкните на таблице правой кнопкой мыши и выберите из контекстного меню *Быстрые статистические графики* опцию *Пиктографик наблюдений*.

Наблюдаемые значения. Это наблюдаемые значения зависимой переменной.

Предсказанные значения. Это значения зависимой переменной, предсказанные уравнением регрессии.

Остатки. Это наблюдаемые значения минус предсказанные значения.

Стандартные предсказанные значения. Это стандартизованные предсказанные значения зависимой переменной.

Стандартизованные остатки. Это стандартизованные значения остатков (наблюдаемые значения минус предсказанные, деленные на корень квадратный из среднего квадрата остатков).

Стандартные ошибки предсказанных значений. Это стандартные ошибки нестандартизованных предсказанных значений.

Расстояния Махаланобиса. Независимые переменные в уравнении можно представлять точками в многомерном пространстве (каждое наблюдение изображается точкой). В этом пространстве можно построить точку, являющуюся средним всех независимых переменных. Эта “средняя точка” в многомерном пространстве называется *центроидом*, т.е. центром тяжести.

Расстояние Махаланобиса определяется как расстояние от наблюдаемой точки до центра тяжести в многомерном пространстве, определяемом коррелированными (неортогональными) независимыми переменными. (Если независимые переменные некоррелированы, расстояния Махаланобиса совпадают с обычными евклидовыми расстояниями.) Эта мера позволяет, в частности, определить является ли данное наблюдение выбросом по отношению к остальным значениям независимых переменных.

Удаленные остатки. Удаленные остатки, это стандартизованные значения остатков, соответствующих наблюдению, если бы оно было исключено из вычислений. Если удаленный остаток значительно отличается от соответствующего стандартизованного значения остатка, то, возможно, это наблюдение является выбросом, поскольку его исключение существенно изменяет уравнение регрессии.

Расстояния Кука. Это еще одна мера влияния соответствующего наблюдения на регрессионное уравнение. Эта величина показывает разницу

StatSoft®

Copyright © StatSoft, 1995

между вычисленными значениями *B* и значениями, которые получились бы при исключении соответствующего наблюдения. Все расстояния Кука должны быть примерно одинаковыми; если это не так, то имеются основания считать, что соответствующее наблюдение (или наблюдения) смещает оценки коэффициентов регрессии.

Средства борьбы с выбросами. Целью всех описанных статистик является обнаружение выбросов. Напомним, что, особенно при малых *N* (меньших 100), оценки множественной регрессии (*B*-коэффициенты) не являются устойчивыми. Другими словами, отдельные экстремальные наблюдения могут оказать значительное влияние на окончательные оценки. Следовательно, желательно всегда просматривать эти статистики (используя эти или описанные далее опции) и повторить полный регрессионный анализ вновь после удаления всех выбросов.

Другой альтернативой является проведение анализа данных, используя регрессию, основанную на *абсолютных отклонениях*, а не на *квадратах отклонений*, тем самым "ослабляется" влияние выбросов. Модуль *Нелинейного оценивания* (см. том III) позволяет провести такого рода анализ.

Статистика Дарбина-Уотсона

Статистика Дарбина-Уотсона полезна для оценивания наличия или отсутствия сериальной корреляции между остатками (т.е. наличия или отсутствия зависимости между остатками для *соседних* наблюдений; в частности, это может служить доказательством зависимости наблюдений в файле данных).

Отметим, что все критерии значимости в множественной регрессии предполагают, что данные являются случайной выборкой из независимых наблюдений. Если это не так, то оценки (*B*-коэффициенты) могут быть более неустойчивыми, чем это гарантируют их уровни значимости.

Интуитивно понятно, например, что заполнение опросника 100 раз одним и тем же субъектом предоставит меньше информации о популяции в целом, чем заполнение его же случайно отобранными 100 индивидуумами, каждый из которых ответил на вопросы по одному разу. В первом случае наблюдения не являются независимыми друг от друга (один и тот же респондент будет давать похожие ответы в повторяющихся опросниках), в то время как во втором случае наблюдения являются независимыми (для различных людей). [Статистика Дарбина-Уотсона позволяет построить критерий проверки гипотезы о том, что остатки некоррелированы, против альтернативы, что они связаны авторегрессионной зависимостью первого порядка. При этом возможны варианты: коэффициент авторегрессии больше нуля, меньше нуля или не равен нулю. Критерий определяется двумя точками, которые зависят как от числа наблюдений, по которым построена регрессия, так и от числа предикторов.]

При нажатии этой кнопки появляется таблица результатов, содержащая значение статистики *D* Дарбина-Уотсона и сериальную корреляцию (т.е. корреляцию между соседними остатками).

Сохранить остатки и предсказанные

Эта опция полезна, если необходимо сохранить предсказанные значения и/или остатки для последующего анализа в других модулях *STATISTICA*. Нажмите эту кнопку, чтобы открыть диалоговое окно *Сохранение предсказанных значений/остатков*, в котором вы можете выбрать имя файла для сохранения данных. Затем появится окно выбора переменных, позволяющее выбрать дополнительные переменные для сохранения

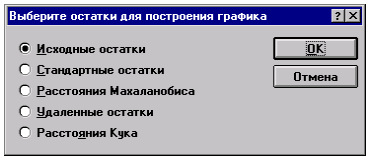
Copyright © StatSoft, 1995

вместе с предсказанными значениями и остатками.

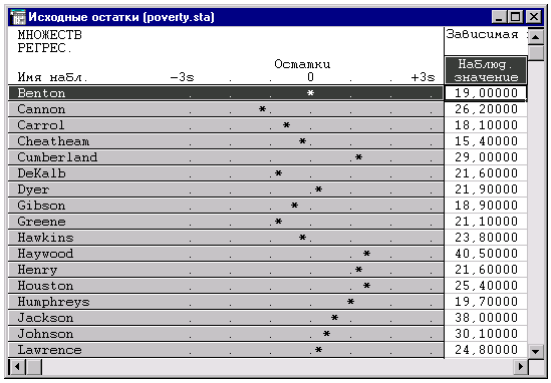
Отметим, что все значения, выбранные в этом списке, будут сохранены вместе с остатками.

Графики остатков

После нажатия этой кнопки появится диалоговое окно, в котором вы можете выбрать тип остатков, который будет построен на графике (см. ниже). Доступными типами остатков являются *исходные остатки*, *стандартизованные остатки*, *расстояния Махаланобиса*, *удаленные остатки* и *расстояния Кука*. Описание этих статистик дано выше (см. опцию *Остатки и предсказанные*).



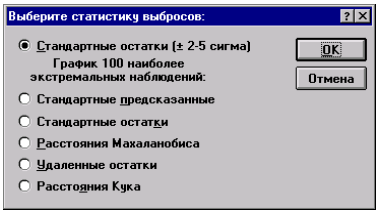
После выбора желаемого типа остатков откроется таблица результатов, в которой выбранные остатки изображены построчно в крайнем левом столбце.



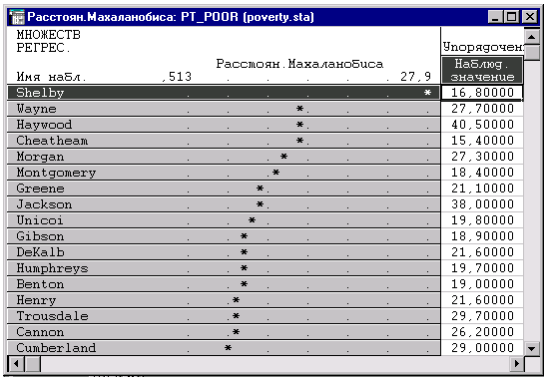
Одновременно в остальных столбцах таблицы вычисляются значения *всех* статистик остатков.

Графики выбросов

После нажатия этой кнопки открывается диалоговое окно, позволяющее выбрать для отображения: (1) только наблюдения, для которых абсолютное значение стандартизированного остатка больше *+2;* (2) *100* наиболее удаленных наблюдений, отсортированных в порядке убывания одной из следующих статистик: *стандартизованных предсказанных* значений, *стандартизованных остатков*, *расстояний Махаланобиса*, *удаленных остатков*, *расстояний Кука* (подробно описаны в разделе *Остатки и предсказанные*).



Таким образом, эта опция позволяет определить выбросы даже в очень большом файле данных, поскольку отображаются только наиболее сильно выделяющиеся наблюдения. В этой таблице выбранные статистики отображаются в крайнем левом столбце; в остальных столбцах таблицы показаны все другие статистики остатков.

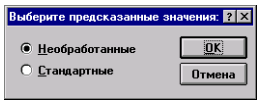


StatSoft®

Copyright © StatSoft, 1995

Графики предсказанных значений

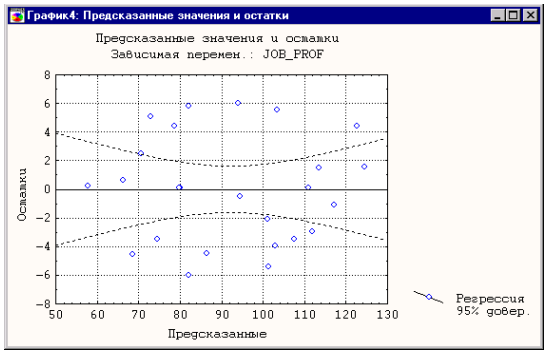
После нажатия этой кнопки вы можете выбрать для отображения *необработанные предсказанные* или *стандартизованные предсказанные* значения.



В таблице результатов выбранные значения отображаются в крайнем левом столбце; одновременно в остальных столбцах показаны значения других статистик остатков.

Предсказанные и остатки

Опция строит диаграмму рассеяния предсказанных значений (по оси *X*) и нестандартизованных остатков (по оси *Y*). На диаграмме также отображаются верхние и нижние доверительные границы.

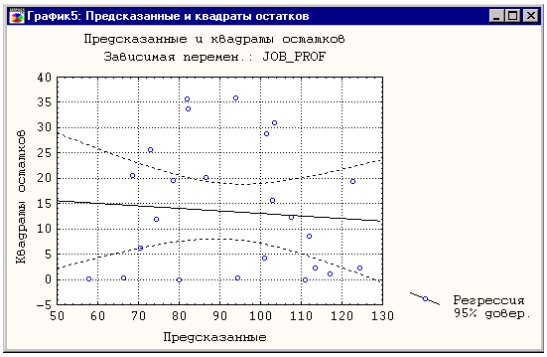


Этот график очень полезен для проверки предположения о *линейности* связи между независимыми переменными и зависимой переменной. Если зависимость линейна, то можно ожидать, что значения остатков образуют однородное "облако" около центральной линии. Если имеет место нелинейность, то может возникнуть своеобразное расположение точек, сильно отличающееся от однородного "облака".

Например, если действительная зависимость между переменными является нелинейной, то остатки могут располагаться в виде перевернутой буквы *U* около центральной линии. Такое расположение показывает, что предсказанные уравнением регрессии значения устойчиво завышаются на краях интервала и занижаются в центральной части. В этом случае может помочь полиномиальная регрессия, а именно, можно попробовать включить в уравнение регрессии квадраты и кубы исходных независимых переменных (X^2, X^3 и т.д.). Технически это легко сделать при проведении *фиксированной нелинейной* регрессии (см. стр. 1679).

Предсказанные и квадраты остатков

Эта опция строит диаграмму рассеяния предсказанных значений (ось *X*) и квадратов нестандартизованных остатков (ось *Y*).

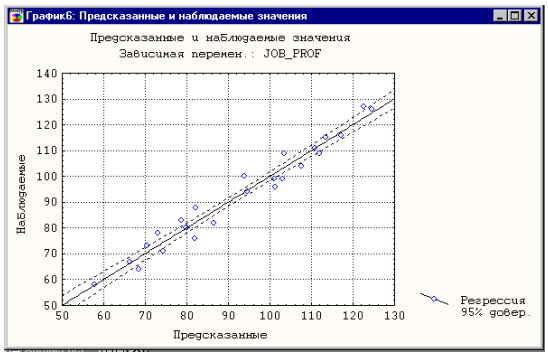


Copyright © StatSoft, 1995

StatSoft®

Предсказанные и наблюдаемые

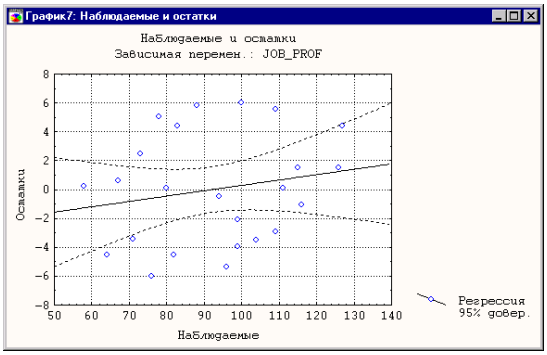
Эта опция позволяет построить диаграмму рассеяния предсказанных и наблюдаемых значений.

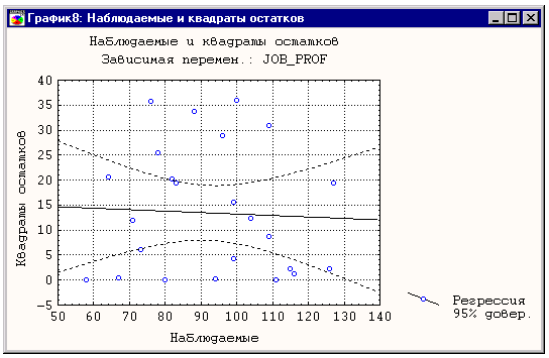


Этот график особенно полезен для нахождения потенциальных кластеров (групп) наблюдений, прогнозы для которых не точны. Например, при анализе данных о выполнении работ (с использованием некоторого критерия интенсивности работы в качестве предиктора), может оказаться, что критерий дает хорошие оценки для большинства служащих, однако для отдельной группы служащих прогнозы окажутся устойчиво заниженными (общая проблема тестов, которая с трудом может быть выражена вербально). Возможно, для этой группы служащих следует найти другой показатель и еще раз провести анализ.

Наблюдаемые и остатки, наблюдаемые и квадраты остатков

Эти графики также очень полезны для обнаружения выбросов и групп наблюдений, предсказанные значения для которых постоянно завышают или занижают точные значения (см. выше).





Остатки и удаленные остатки

Удаленные остатки являются стандартизованными остатками, которые получились бы при исключении соответствующего наблюдения из процедуры оценивания коэффициентов регрессионного уравнения. Таким образом, если имеются большие расхождения между удаленными остатками и обычными стандартизованными остатками, то можно заключить, что значения регрессионных коэффициентов не являются достаточно устойчивыми и в большой степени зависят от исключения отдельных наблюдений (см. также выше обсуждение опции *Остатки и предсказанные значения*).

StatSoft®

Copyright © StatSoft, 1995

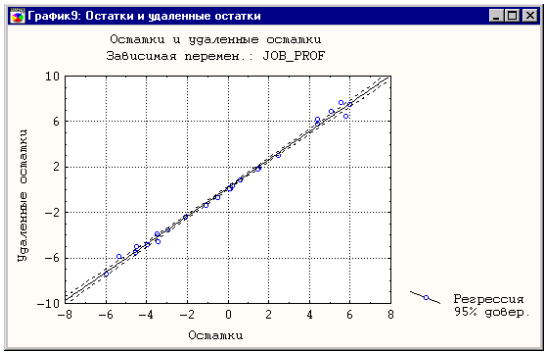


График наблюдаемых

Эта опция строит гистограмму (показывающую распределение частот) наблюдаемых значений зависимой переменной и наложит на нее график подогнанного нормального распределения.

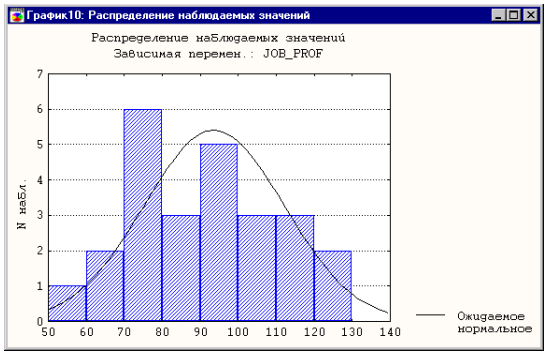
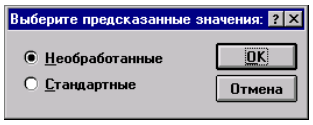


График предсказанных

После нажатия этой кнопки вы можете выбрать для отображения на гистограмме либо *необработанных* (нестандартизованных) *предсказанных* значений либо *стандартизованных предсказанных* значений в диалоговом окне *Выберите предсказанные значения*.



После нажатия кнопки *OK* этого диалогового окна будет построен следующий график.

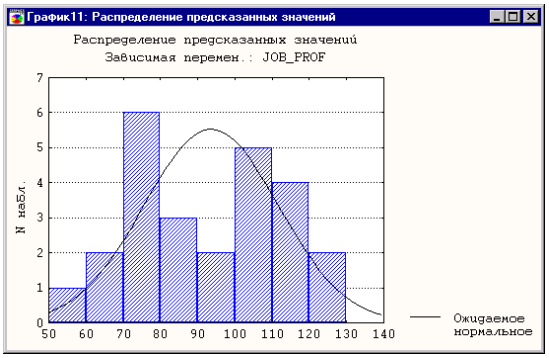
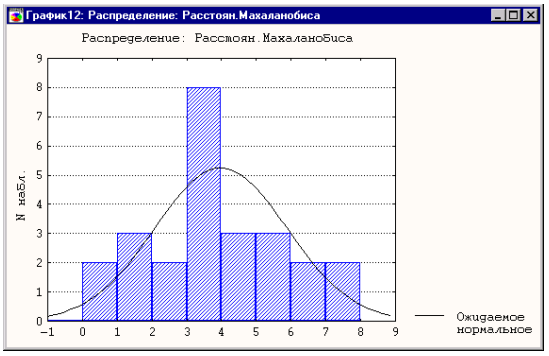


График остатков

После нажатия этой кнопки откроется окно, в котором вы можете выбрать тип остатков для построения гистограммы. Доступными типами остатков являются *исходные остатки*, *стандартизованные остатки*, *расстояния Махаланобиса*, *удаленные остатки* и *расстояния Кука*. (Краткое описание этих статистик дано выше в разделе *Остатки и предсказанные значения*, стр. 1690).

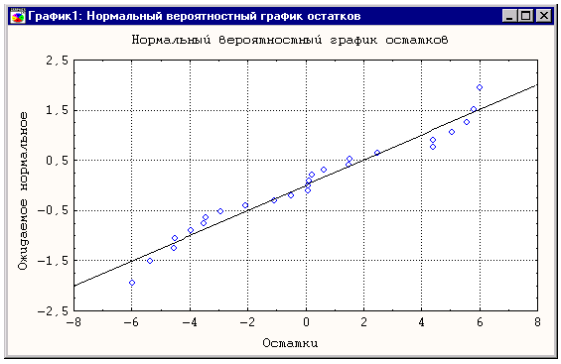


Нормальный график остатков

При проведении анализа с помощью множественной регрессии предполагается, что

Copyright © StatSoft, 1995

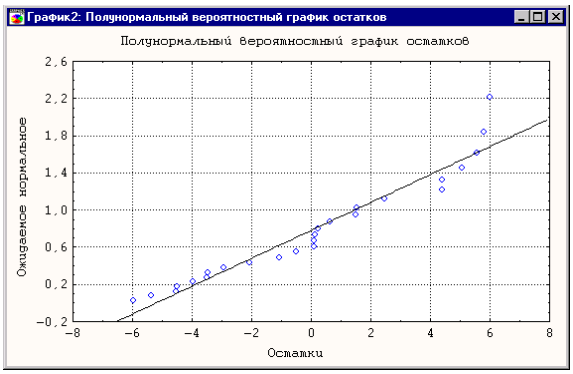
значения остатков (наблюдаемые значения минус предсказанные значения) нормально распределены, и функция регрессии (т.е. связь между зависимой и независимыми переменными) в действительности линейна. Если какое-либо из этих предположений существенно нарушается, то может получиться завышение или занижение оценок коэффициентов регрессии (*B*-коэффициентов), а также завышение или занижение значимости критериев. Если же "все обстоит хорошо" и модель адекватна данным, то можно ожидать, что значения остатков нормально распределены. *Нормальный вероятностный график* обеспечивает быстрый способ визуальной проверки того, в какой мере распределение остатков подчиняется нормальному закону.



Если остатки имеют распределение, отличное от нормального, то точки на графике будут отклоняться от прямой. На этом графике также становятся заметны выбросы. Если модель плохо согласуется с наблюдениями, и данные располагаются специальным образом около прямой линии (например, имеют *S*-образный вид), то, возможно, требуется нелинейное преобразование зависимой переменной (например, логарифмирование для "подтягивания" хвостов распределения).

Полунормальный график остатков

Полунормальный вероятностный график строится так же, как и стандартный нормальный вероятностный график за исключением того, что рассматривается только положительная часть нормального распределения. Следовательно, только положительные нормальные значения изображаются на оси *Y*.



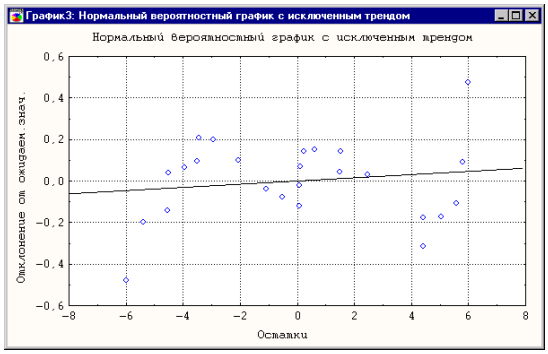
Этот график используется в основном в тех случаях, когда нас не интересует знак остатка, а когда роль играют только абсолютные значения остатков, безотносительно их знаков.

Нормальный график остатков без тренда

Нормальный вероятностный график без тренда строится так же, как и стандартный нормальный вероятностный график за исключением того, что до построения графика удаляется линейный тренд. Это часто "расширяет" график и облегчает исследование деталей отклонений от нормальности.

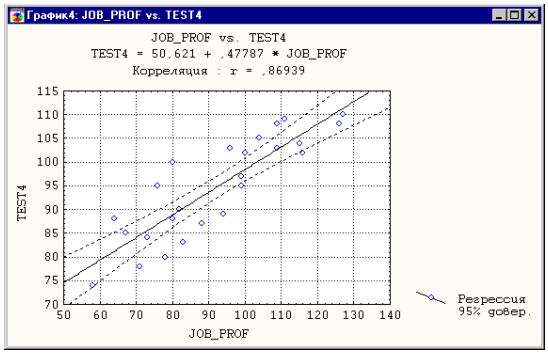
StatSoft

Copyright © StatSoft, 1995



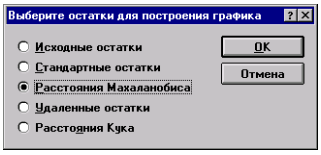
Две переменные

Эта опция позволяет построить диаграмму рассеяния любых двух переменных из файла данных, независимо от того, были ли эти переменные включены в анализ или нет.



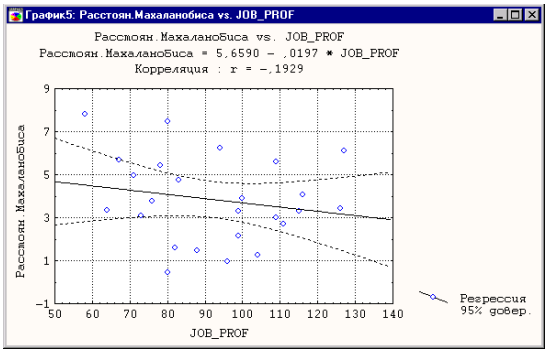
Остатки и переменная

Эта опция позволяет построить диаграммы рассеяния остатков различных типов (зависимой переменной) относительно любой переменной из файла данных, независимо от того, была она включена в регрессионный анализ или нет. После нажатия этой кнопки открывается диалоговое окно *Выберите остатки для построения графика*, в котором вы можете выбрать тип остатков для этой диаграммы рассеяния.



Доступными типами остатков являются *исходные остатки*, *стандартные остатки*, *расстояния Махаланобиса*, *удаленные остатки* и *расстояния Кука*. Краткое описание этих статистик дано в разделе *Остатки и предсказанные значения*.

После этого программа предложит выбрать вторую переменную для построения графика. Эта опция позволяет исследовать зависимость между остаточными значениями зависимой переменной и другими переменными.

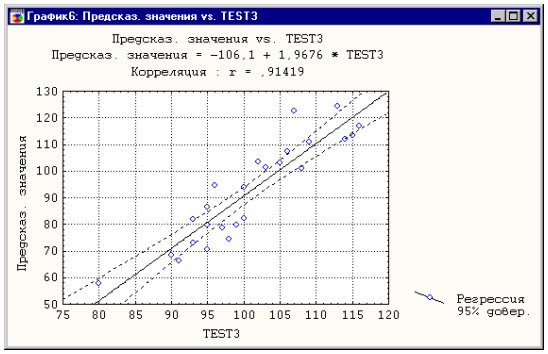


Предсказанные и переменная

Опция позволяет построить диаграмму рассеяния предсказанных значений выбранного типа относительно любой переменной из файла данных, независимо от того, была соответствующая переменная включена в регрессионный анализ или нет. После нажатия кнопки открывается диалоговое окно, в котором можно выбрать использование исходных или стандартизованных предсказанных значений. Далее программа предложит выбрать переменную для этого графика.

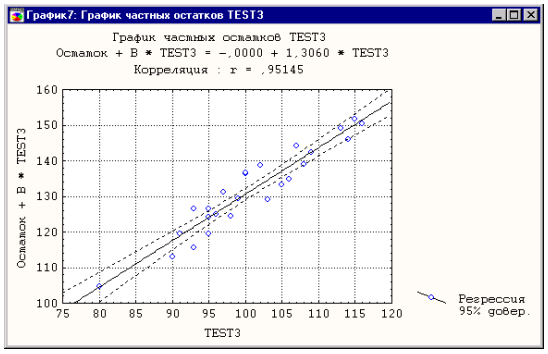
Copyright © StatSoft, 1995

StatSoft®



Частные остатки

Эта опция позволяет построить график *частных остатков* для любой переменной, включенной в настоящий момент в регрессионное уравнение. После нажатия этой кнопки программа предложит выбрать независимую переменную из включенных в регрессионное уравнение.



На графике частных остатков, остаток плюс вклад соответствующей независимой переменной в регрессию (т.е. *остаток* + b_i*x_{ii}) отображается напротив значения выбранной независимой переменной. Подробнее этот график обсуждается в книге Larsen and McCleary (1972).

StatSoft®

Copyright © StatSoft, 1995

ПРИМЕЧАНИЯ

Исходные данные на входе, удаление пропущенных данных

Если файл исходных данных используется в качестве входных данных, то модуль *Множественной регрессии* сначала вычисляет корреляционную матрицу. Пропущенные значения могут удаляться *построчно*, *попарно*, или же *заменяться на средние значения* (см. *Первоначальные понятия*, глава 8).

Если запрошено *попарное* удаление пропущенных данных, то последующий регрессионный анализ и тесты на значимость будут вычисляться на основе числа степеней свободы (сс), соответствующего наименьшему числу допустимых наблюдений по выбранным переменным (число допустимых наблюдений для каждой пары выбранных переменных можно просмотреть). Следовательно, результаты регрессионного анализа с использованием попарного удаления пропущенных данных должны интерпретироваться с осторожностью, поскольку они являются только приближением, основанным на наименьшем числе допустимых наблюдений.

При выборе *построчного* удаления пропущенных данных вычисление корреляционной матрицы будет более быстрым и тестовые статистики (например, *R-квадрат*, *F*), вычисленные в последующем анализе множественной регрессии будут более точными.

Матрица на входе

Если корреляционная матрица выбрана в качестве исходных данных модуля *Множественной регрессии*, то из нее извлекается подматрица по выбранным переменным. Очевидно, предсказанные значения и остатки не могут быть вычислены, если в качестве исходных данных используется корреляционная матрица.

Исходный файл с корреляционной матрицей должен быть записан в формате матричного файла *STATISTICA*. Такие файлы могут быть созданы в каждом модуле *STATISTICA* с помощью опции *Быстрые основные статистики* (см. главу 10); некоторые модули *STATISTICA* имеют также дополнительные опции, предназначенные для создания таких файлов (например, *Основные статистик и таблицы*, *Множественная регрессия*, *Факторный анализ*, *Канонический анализ*, *Теория надежности*), или же пользователь может создать файлы такого типа напрямую, непосредственным вводом данных в электронную таблицу с помощью клавиатуры (см. *Формат матричного файла*, глава 7).

Опция пакетной обработки

В дополнение к языку макрокоманд (*SCL*), доступному во всех модулях *STATISTICA*, модуль *Множественной регрессии* предлагает также и опцию пакетной обработки. Пользователь может задать до 99 зависимых переменных в диалоговом окне *Определение модели* (щелчок по кнопке *Переменные* откроет диалоговое окно выбора переменных). Если задан *список* зависимых переменных, программа пройдет по списку, выполняя анализ множественной регрессии (в соответствии с текущими спецификациями) для каждой зависимой переменной. Результаты заданных таким образом анализов могут печататься один за другим (без участия пользователя), посредством задания принтера в качестве устройства вывода в диалоговом окне *Параметры страницы/вывода* и установки опции *Пакетная обработка и печать* в

Copyright © StatSoft, 1995

стартовой панели модуля *Множественная регрессия*.

Использование попарного удаления пропущенных данных

Следует с большой осторожностью подходить к интерпретации результатов анализа множественной регрессии при включенной опции попарного удаления пропущенных данных. Теоретически, наблюдение не должно приниматься в расчет, если оно имеет пропущенные данные хотя бы для одной из переменных, включенных в анализ множественной регрессии. Модуль *Множественной регрессии* будет выполнять регрессионный анализ, считая корреляционную матрицу (вычисленную с использованием попарного удаления пропущенных данных) построенной только на заполненных наблюдениях (т.е. без каких-либо пропущенных данных; модуль *Множественной регрессии* будет использовать минимальное число наблюдений для каждой переменной регрессионного уравнения для определения числа степеней свободы в тестах на значимость и при оценке *Суммы квадратов*). Рекомендуется всегда сравнивать результаты, полученные с использованием попарного удаления пропущенных данных, с результатами, полученными с использованием построчного удаления пропущенных данных или замены пропущенных данных средними, для обнаружения больших расхождений.

Задание значений *F-включить /исключить* для пошаговой регрессии

В пошаговой регрессии переменная включается в регрессионное уравнение, если значение *F*-статистики для нее больше заданного минимального значения (т.е. *F-включить*, задаваемого в диалоговом окне *Определение модели*), и исключается из регрессионного уравнения, если значение *F*-статистики для нее меньше заданного минимального значения (т.е. *F-исключить* в диалоговом окне *Определение модели*). Следовательно, значение *F-включить* всегда должно задаваться большим значения *F-исключить*; в противном случае переменная будет включаться и исключаться в бесконечном цикле. Модуль *Множественной регрессии* выдаст сообщение об ошибке, если пользователь нарушит это правило.

Если *пошаговая регрессия с исключением предикторов* запрошена в диалоговом окне *Определение модели*, то программа сначала включит все переменные в регрессионное уравнение (шаг 0), а затем найдет переменную с наименьшим значением *F-исключить*. Переменные далее будут последовательно исключаться на шагах регрессии до тех пор, пока ни одна из переменных регрессионного уравнения не будет иметь значения *F-исключить* меньше заданного порога. Поэтому, если желательно удалить *все* переменные из регрессионного уравнения, по одной на каждом шаге, то следует задать очень большое значение *F-включить* (например, *999*) и очень большое значение *F-исключить* (например, *998*; напомним, что значение *F-исключить* должно быть всегда меньшим значения *F-включить*).

Если задана *пошаговая регрессия с включением предикторов*, то на каждом шаге программа будет проходить по списку независимых переменных еще не включенных в регрессионное уравнение и включать переменную с наибольшим значением *F-включить*. Пошаговый процесс закончится тогда, когда для каждой независимой переменной (не включенной в регрессионное уравнение) значение *F-включить* не будет превышать заданное пороговое значение *F-включить*. Поэтому, если желательно включить все

StatSoft®

Copyright © StatSoft, 1995

переменные в регрессионное уравнение, по одной на каждом шаге, следует задать очень маленькое значение *F-включить* (например, *0.0001*), а значение *F-исключить* установить равным нулю (введите *0.0*).

Включение переменных по блокам (иерархический анализ)

Опции *пошагового включения* и *пошагового исключения* приводят к поочередному включению или исключению переменных из уравнения регрессии. Часто желательно оценить общий эффект от включения нескольких переменных в регрессионное уравнение на одном шаге или на последовательности шагов. Необходимые для этого статистики автоматически вычисляются модулем *Множественной регрессии* при надлежащем задании списка независимых переменных в диалоговом окне *Определение модели*.

При выборе *стандартной* множественной регрессии программа всегда сравнивает переменные, выбранные для данного анализа, с переменными, выбранными в предыдущем анализе. Вначале программа проверяет: для анализа использовалась одна и та же зависимая переменная. Если нет, то приращение *R-квадрата* по сравнению с предыдущим анализом не может быть осмысленным. Если да, то приращение имеет смысл (это приращение квадрата множественной корреляции одной и той же зависимой переменной с группой независимых переменных) и его можно вычислить. Далее, программа проверяет: (1) является ли список независимых переменных подмножеством заданного ранее списка независимых переменных (т.е. часть переменных была исключена) или (2) является ранее заданный список независимых переменных подмножеством списка, заданного к настоящему моменту (т.е. переменные включались). Если одно из этих двух условий выполняется, то модуль *Множественной регрессии* вычисляет пошаговое увеличение или уменьшение *R-квадрата* и соответствующие статистики (доступные с помощью кнопки *Итоги по шагам* диалогового окна *Результаты множественной регрессии*; см. пример на стр. 1682).

Плохо обусловленная матрица

Говорят, что корреляционная матрица плохо обусловлена, если имеется линейная зависимость между переменными. Например, если одна независимая переменная является суммой двух других независимых переменных, то множественная регрессия не может быть выполнена. На практике плохая обусловленность матрицы возникает при использовании нескольких сильно взаимно коррелированных независимых переменных.

Имеется два способа, которые можно применять для решения этой проблемы. Во-первых, используйте пошаговую регрессию с включением для добавления предикторов по одному на каждом шаге. Это позволяет следить за собственным вкладом каждой независимой переменной в предсказание по мере их включения в регрессионное уравнение. Опция *Избыточность* в диалоговом окне *Результаты множественной регрессии* особенно полезна в этом отношении. Для каждой независимой переменной она позволяет вычислять значение *R-квадрата* со всеми остальными переменными, входящими на данном этапе в регрессионное уравнение, *1-R-квадрат* (*Толерантность*), а также *частные* и (часть) *получастные* корреляции. Эти статистики позволят идентифицировать переменные, которые являются в значительно степени избыточными в совокупности с другими предикторами.

Copyright © StatSoft, 1995

Гребневая регрессия

Второй возможностью является использование *гребневой регрессии* (выберите опцию *Гребневая регрессия* в диалоговом окне *Определение модели*, стр. 1684). В гребневой регрессии (см. Hoerl, 1962) от пользователя требуется задать значение для *лямбда*. Это значение затем добавляется к диагонали корреляционной матрицы; в результате значения корреляций искусственно занижаются. Хотя получающиеся оценки не являются несмещенными, они могут быть в действительности более точными и устойчивыми (см. обсуждение этого в Schmidt and Muller, 1978; см. также Rozeboom, 1979, по поводу критических оценок этого метода).

Некоторые полезные программы на языке *STATISTICA BASIC*

Система *STATISTICA* включает несколько примеров программ в *STATISTICA BASIC* (в подкаталоге *STBASIC*), являющихся полезными для решения определенных регрессионных задач.

Исключительно большие задачи множественной регрессии

Программа *Regressn.stb* выполнит регрессионный анализ с очень большим количеством переменных-предикторов (используемые в вычислениях массивы динамически размещаются в программе, так что размер анализируемой проблемы ограничен только объемом памяти, доступной вашей системе). Программа позволяет оценивать регрессионные модели со свободным членом и без него, а также вычислять статистики остатков и рисовать графики.

Оценивание методом взвешенных наименьших квадратов

Программа *Wls.stb* выполняет анализ по методу взвешенных наименьших квадратов и вычисляет различные, альтернативные R-квадрат, статистики (см. Kvelseth, 1985). Вместо итеративного не использующего производные метода оценивания, описанного в главе *Нелинейное оценивание* (том III), программа *Wls.stb* использует выражения матричной алгебры (например, см. Neter, Wassermann, и Kutner, 1985) для оценивания параметров метода взвешенных наименьших квадратов. Таким образом, для больших файлов данных, программа *Wls.stb* является более эффективной и часто гораздо более быстрой.

Оценивание двухступенчатым методом наименьших квадратов

Программа *2stls.stb* представляет пример регрессии с 2х-ступенчатым методом наименьших квадратов. Эта программа может легко быть расширена или приспособлена для выполнения, например, анализа с итеративным повторным взвешиванием по методу наименьших квадратов.

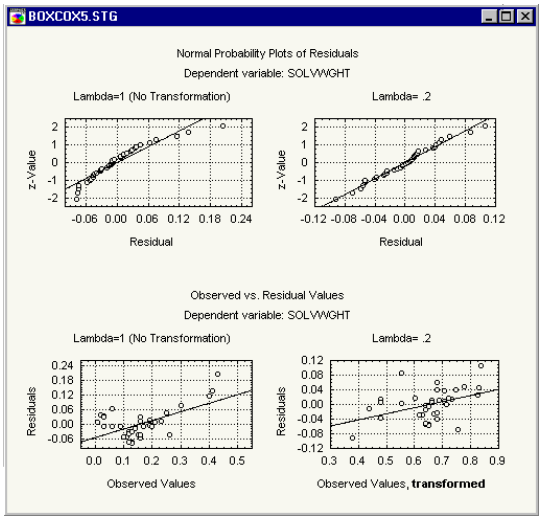
Преобразования Бокса-Кокса и Бокса-Тидвелла

Часто, когда остатки, полученные при анализе обычным методом наименьших квадратов, не следуют нормальному закону распределения, возможно, применить определенные преобразования зависимой или независимых переменных для решения этой проблемы. Семейство простых степенных преобразований для положительных значений отклика *y* может быть описано следующим образом:

$$z = y^{\lambda}, \quad \lambda \neq 0$$
$$= \text{натур.логарифм } (y), \quad \lambda = 0$$

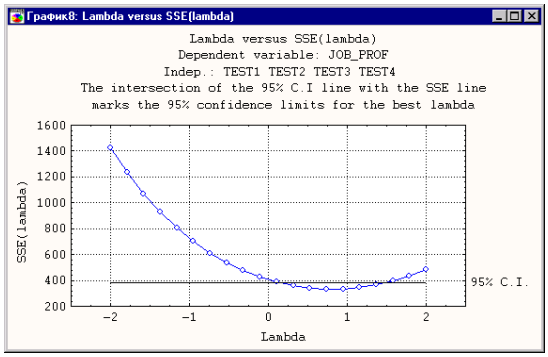
StatSoft®

Copyright © StatSoft, 1995



Заметим, что при $\lambda = 1$ (*lambda =1*) преобразование не выполняется; при $\lambda = -1$ результатом преобразования является величина, обратная *y*, при $\lambda = 1/2$ результатом преобразования будет корень квадратный из y. Таким образом, некоторые различные преобразования зависимой переменной могут быть выражены в этой общей форме.

Программа *Boxcox.stb* вычисляет оценку максимума правдоподобия для *лямбда* в преобразовании *Бокса-Кокса* для зависимой переменной (Box и Cox, 1964; см. также Mason, Gunst, и Hess, 1989) и позволяет строить графики для сумм квадратов ошибок, как функции от различных значений *лямбда*.



Тот же класс степенных преобразований может быть применен к предикторным (независимым) переменным в задаче линейной регрессии. Программа *Boxtid.stb* вычисляет параметры *лямбда* для этих, так называемых, преобразований *Бокса-Тидвелла* независимых переменных (см. Mason, Gunst, и Hess, 1989).

Copyright © StatSoft, 1995

StatSoft®

ТЕХНИЧЕСКАЯ ИНФОРМАЦИЯ

Общий обзор

Программа модуля *Множественная регрессия* состоит из двух основных частей. Первая часть вычисляет корреляционную матрицу (или извлекает корреляционную матрицу, если в качестве исходных данных выбрана матрица) в соответствии со спецификациями, заданными пользователем (способом обработки пропущенных данных, условиями выбора случаев и т.д.). Вторая часть фактически выполняет регрессионный анализ.

Вычисления при проведении множественной регрессии, обращение матрицы

Обращение матрицы осуществляется с помощью метода выметания (см. Dempster, 1969, стр. 62). Регрессионные коэффициенты, остаточные *суммы квадратов*, толерантности и частные корреляции также вычисляются в процессе операции выметания (см. также Jennrich, 1977).

Проверка статистической значимости

Используются стандартные формулы для вычисления *F*-статистики, связанной с множественным *R,* и для *t*-статистик, соответствующих регрессионным коэффициентам (см. например Cooley и Lohnes, 1971; Darlington, 1990; Lindeman, Merenda, и Gold, 1980; Morrison, 1967; Neter, Wasserman, и Kutner, 1985; Pedhazur, 1973; Stevens, 1986; Younger, 1985).

Остатки

Стандартная ошибка остатка вычисляется, как квадратный корень из:

$$[1-1/n-(X_{raw}-X_{mean})*C^{-1}*(X_{raw}-X_{mean})']*RMS$$

где

X_{raw} вектор исходных данных для независимых переменных

X_{mean} вектор средних для независимых переменных

C^{-1} матрица, обратная к матрице попарных произведений отклонений для независимых переменных

n число допустимых наблюдений

RMS средний квадрат остатков

Члены *1/n* и X_{mean} опускаются при отсутствии свободного члена (регрессия принудительно проходит через начало координат).

Стандартизованные остатки получаются делением каждого остатка на его стандартную ошибку.

Расстояние Махаланобиса является расстоянием от наблюдения до центроида всех наблюдений в пространстве, заданном независимыми переменными. Оно вычисляется следующим образом:

$$(n-1)*(X_{raw}-X_{mean})*C^{-1}*(X_{raw}-X_{mean})'$$

где

X_{raw} вектор исходных данных для независимых переменных

X_{mean} вектор средних для независимых переменных

C^{-1} матрица, обратная к матрице попарных произведений отклонений для независимых переменных

n число допустимых наблюдений

Члены *1/n* и X_{mean} опускаются при отсутствии свободного члена (регрессия принудительно проходит через начало координат). См. в разделе *Примеры* пример того, как расстояния

StatSoft

Copyright © StatSoft, 1995

Махаланобиса могут помочь в обнаружении выбросов.

Удаленные остатки являются остатками, которые получились бы, не будь данное наблюдение включено в оценивание регрессионного уравнения. Они вычисляются посредством деления обычных остатков на:

$$1-(1/n)-(X_{raw}-X_{mean})*C^{-1}*(X_{raw}-X_{mean})'$$

где

X_{raw} вектор исходных данных для независимых переменных

X_{mean} вектор средних для независимых переменных

C^{-1} матрица, обратная к матрице попарных произведений отклонений для независимых переменных

n число допустимых наблюдений

Члены *1/n* и X_{mean} опускаются при отсутствии свободного члена (регрессия принудительно проходит через начало координат). См. в разделе *Примеры* пример того, как удаленные остатки могут помочь в обнаружении выбросов.

Расстояния Кука (Cook, 1977) полезны для оценки изменений во всех остатках, которые возникли бы при выбрасывании соответствующего наблюдения из регрессионного анализа. Они определяются следующим образом:

$$\{DR^2* [1/n + MD/(n-1)]\}/[(No.Vars + 1)*RSM]$$

где

DR *удаленный остаток*

MD расстояние Махаланобиса

RSM средний квадрат остатков

Если свободный член отсутствует, *n-1* заменяется на *n*, член *1/n* опускается и член *+1* (добавление 1 к числу независимых переменных) также выбрасывается.

Вычисления для нормальных вероятностных графиков

Следующие формулы используются для преобразования рангов (вычисленных по остаткам) к ожидаемым значениям вероятности нормального распределения, т.е. соответствующим нормальным *z*-значениям.

Нормальный вероятностный график. Нормальное вероятностное значение z_j для *j*-ого значения (ранга) переменной с *N* наблюдениями вычисляется следующим образом:

$$z_j = \Phi^{-1} [(3*j-1)/(3*N+1)]$$

где Φ^{-1} является функцией, обратной к функции нормального распределения (и преобразовывающей нормальную вероятность *p* в нормальное значение *z*).

Полунормальный вероятностный график. Здесь, полунормальное вероятностное значение z_j для *j*-ого значения (ранга) переменной с *N* наблюдениями вычисляется следующим образом:

$$z_j = \Phi^{-1} [(3*N+3*j-1)/(6*N+1)]$$

где Φ^{-1} также является функцией, обратной к функции нормального распределения.

Нормальный вероятностный график без тренда . На этом графике каждое значение (x_j) стандартизуется посредством вычитания *среднего* и деления на соответствующее стандартное отклонение (*s*). Нормальное вероятностное значение без тренда z_j для *j*-ого значения (ранга) переменной с *N* наблюдениями вычисляется следующим образом:

$$z_j = \Phi^{-1} [(3*j-1)/(3*N+1)] - (x_j-mean)/s$$

где Φ^{-1} также является функцией, обратной к функции нормального распределения.

Copyright © StatSoft, 1995

ПРЕДМЕТНЫЙ УКАЗАТЕЛЬ

А

Анализ остатков, 1661, 1669

В

Взвешенный метод наименьших квадратов, 1702
Выбросы
  графики, 1692
  интерактивное средство удаления выбросов, 1672
  построчный график, 1670
  расстояния Кука, 1690, 1705
  расстояния Махаланобиса, 1690
  средства борьбы с выбросами, 1691
  удаленные остатки, 1690

Г

График для двух переменных, 1697
Гребневая регрессия, 1684, 1702

Д

Дарбина-Уотсона статистика, 1691
Двухступенчатый метод наименьших квадратов, 1702
Дисперсионный анализ, скорректированный на среднее, 1687

З

Замена пропущенных данных средним, 1677, 1699
Значение F-включить, 1684
Значение F-исключить, 1674, 1684

И

Иерархическая регрессия, 1682, 1701
Избыточность, 1688
Итоги регрессии по шагам, 1688

К

Кисть, 1672
Ковариации коэффициентов, 1687
Ковариации переменных, 1681
Корреляции переменных, 1681
Коэффициент R-квадрат, 1659, 1686
Коэффициент детерминации, 1659
Коэффициент множественной корреляции R, 1659, 1686

М

Матрица выметания, 1687
Множественная регрессия
  анализ остатков, 1661, 1669
  вычислительные аспекты, 1658
  задание множественной регрессии, 1668
  коэффициент R-квадрат, 1659
  коэффициент множественной корреляции R, 1659
  мультиколлинеарность, 1661
  общее назначение, 1657
  остаточная дисперсия, 1659
  оценивание методом наименьших квадратов, 1658
  пакетная обработка, 1699
  плохая обусловленность матрицы, 1661
  предсказанные значения и остатки, 1659
  расстояния Махаланобиса, 1670, 1704
  регрессионные коэффициенты, 1668
  стандартная ошибка оценки, 1686
  уравнение регрессии, 1658
  частные корреляции, 1668, 1688
Мультиколлинеарность, 1661

Н

Наблюдаемые значения, 1690
Нормальные вероятностные графики, 1672, 1705

О

Остатки, 1659, 1690
  графики, 1692
  нормальный график, 1695, 1705
  нормальный график без тренда, 1696, 1705
  полунормальный график, 1696, 1705
  построчный график, 1669
  удаленные, 1671
  частные, 1698
Остаточная дисперсия, 1659
Оценивание методом наименьших квадратов, 1658

П

Параметр F-включить, 1700
Параметр F-исключить, 1700
Плохая обусловленность матрицы, 1661
Плохо обусловленная матрица, 1701
Повышенная точность вычислений, 1678
Получастные корреляция, 1669
Попарное удаление пропущенных данных, 1677, 1699, 1700
Построчное удаление пропущенных данных, 1677, 1699
Построчный график выбросов, 1670
Построчный график остатков, 1669
Пошаговая регрессия, 1673
Пошаговая регрессия с включением предикторов, 1673, 1682, 1700
Пошаговая регрессия с исключением предикторов, 1682, 1700
Предположения регрессии
  линейность, 1660
  нормальность, 1660
Предсказанные значения, 1690
Предсказанные значения и остатки, 1659
Преобразования Бокса-Кокса и Бокса-Тидвелла, 1702
Программы на языке *STATISTICA BASIC* и их использование для множественной регрессии, 1702
Пропущенные данные
  замена средним, 1677, 1699
  попарное удаление, 1677, 1699, 1700
  построчное удаление, 1677, 1699

Р

Расстояния Кука, 1690, 1705

StatSoft®

Copyright © StatSoft, 1995

Расстояния Махаланобиса, 1670, 1690, 1704
Регрессионные коэффициенты, 1668
Регрессионный анализ
взвешенный метод наименьших квадратов, 1702
гребневая регрессия, 1684, 1702
двухступенчатый метод наименьших квадратов, 1702
дисперсионный анализ, скорректированный на среднее, 1687
значение F-включить, 1684, 1700
значение F-исключить, 1674, 1700
иерархическая регрессия, 1682, 1701
избыточность, 1688
коэффициент R-квадрат, 1686
коэффициент множественной корреляции R, 1686
пошаговая регрессия, 1673
пошаговая регрессия по блокам, 1682, 1701
пошаговая регрессия с включением предикторов, 1673, 1682, 1700
пошаговая регрессия с исключением предикторов, 1682, 1700
преобразования Бокса-Кокса и Бокса-Тидвелла, 1702
программы на языке *STATISTICA BASIC*, 1702
свободный член, 1683
скорректированный R-квадрат, 1686
стандартная регрессия, 1665, 1682
толерантность, 1683
фиксированная нелинейная регрессия, 1679

С

Свободный член, 1683
Сериальная корреляция остатков, 1691
Скорректированный R-квадрат, 1686
Стандартизованные остатки, 1690
Стандартная ошибка оценки, 1686
Стандартная регрессия, 1665, 1682
Стандартные остатки, 1704
Стандартные ошибки предсказанных значений, 1690
Стандартные предсказанные значения, 1690

Т

Толерантность, 1683

У

Удаленные остатки, 1671, 1690, 1705

Ф

Фиксированная нелинейная регрессия, 1678, 1679

Ч

Частные корреляции, 1658, 1668, 1688
Частные остатки, 1698
Число пар, 1681

Copyright © StatSoft, 1995

StatSoft®

Copyright © StatSoft, 1995

Глава 13:

Дисперсионный анализ (*ANOVA/MANOVA*)

Оглавление

Вводный обзор ... 1715
Основные идеи ... 1715
Сложные планы ... 1719
Ковариационный анализ ... 1721
Многомерные планы:многомерный дисперсионный и ковариационный анализ ... 1722
Анализ контрастов и апостериорные критерии ... 1723
Сумма квадратов типа I, II, III, и IV ... 1724
Принимаемый по умолчанию вычислительный подход в модуле *Дисперсионный анализ* ... 1727
Анализ планов с пропущенными ячейками (неполные планы) ... 1728
Предположения и эффекты нарушения предположений ... 1730
Обзор программы ... 1735
Примеры ... 1738
Диалоговые окна, опции, статистики ... 1757
Стартовая панель ... 1757
Задание планов с повторными измерениями ... 1760
Задание гнездовых (вложенных) планов ... 1761
Задание случайных факторов (эффектов) ... 1762
Фиксированные и переменные ковариаты ... 1762
Диалоговое окно *Результаты* ... 1763
Опции вывода ... 1765
Примечания, часть I: Примеры анализа планов общего типа ... 1784
Примечания, часть II: Примеры дисперсионного анализа несбалансированных и неполных планов ... 1806
Технические замечания ... 1820
Предметный указатель ... 1822

Подробное оглавление

ВВОДНЫЙ ОБЗОР ... 1715
Основные идеи ... 1715
Разбиение суммы квадратов. ... 1715
Многофакторный дисперсионный анализ ... 1717
Эффекты взаимодействия ... 1718
Сложные планы ... 1719
Межгрупповые и внутригрупповые планы (планы с повторными измерениями) ... 1720
Неполные (гнездовые) планы ... 1720

Copyright © StatSoft, 1995

StatSoft®

Ковариационный анализ ... 1721
Основная идея ... 1721
Фиксированные ковариаты ... 1721
Переменные ковариаты ... 1722
Многомерные планы:многомерный дисперсионный и ковариационный анализ ... 1722
Межгрупповые планы ... 1722
Планы с повторными измерениями ... 1723
Суммирование значений переменных и многомерный дисперсионный анализ ... 1723
Анализ контрастов и апостериорные критерии ... 1723
Почему сравниваются отдельные множества средних? ... 1723
Анализ контрастов ... 1724
Апостериорные сравнения ... 1724
Сумма квадратов типа I, II, III и IV ... 1724
Многомерная регрессия и дисперсионный анализ ... 1724
Несбалансированные и сбалансированные планы ... 1725
Сумма квадратов типа I, II, III и IV ... 1726
Интерпретация гипотезы о сумме квадратов типа I, II, и III ... 1726
Принимаемый по умолчанию вычислительный подход в модуле *Дисперсионный анализ* ... 1727
Спланированные сравнения ... 1728
Анализ планов с пропущенными ячейками (неполные планы) ... 1728
Регрессионный подход ... 1728
Средние ячеек, подход дисперсионного анализа, гипотезы типа IV ... 1729
Пропущенные ячейки и проверка специфического эффекта ... 1729
Пропущенные ячейки и объединенные эффекты/члены ошибки ... 1730
Предположения и эффекты нарушения предположений ... 1730
Отклонение от предположения о нормальности распределений ... 1730
Однородность дисперсии ... 1730
Однородность дисперсии и ковариаций ... 1731
Сферичность и сложная симметрия: причины использования многомерного подхода к повторным измерениям в дисперсионном анализе ... 1732
Многомерный дисперсионный анализ и структурное моделирование уравнений. ... 1734

ОБЗОР ПРОГРАММЫ ... 1735
Особенности модуля *Дисперсионный анализ* ... 1735
Ковариаты ... 1736
Анализ контрастов ... 1736
Графики ... 1736
Альтернативные процедуры ... 1737

ПРИМЕРЫ ... 1738
Пример 1: План 2x3 одномерного дисперсионного анализа ... 1738
Файл данных ... 1738
Выбор анализа ... 1738
Анализ результатов ... 1739
Апостериорные сравнения средних ... 1740

StatSoft®
Copyright © StatSoft, 1995

Проверка предположений ... 1741
Резюме ... 1743
Пример 2: План 2 x 3 x 3 (межгрупповой с повторными измерениями) ... 1744
Обзор ... 1744
Проблема исследования ... 1744
Выбор плана ... 1745
Анализ результатов ... 1746
Многомерный подход ... 1748
Резюме ... 1751
Пример 3: Многомерный ковариационный анализ ... 1751
Обзор ... 1751
Файл данных ... 1751
Задание плана ... 1752
Анализ результатов ... 1752
Исследование гипотезы параллельности. ... 1755

ДИАЛОГОВЫЕ ОКНА, ОПЦИИ, СТАТИСТИКИ ... 1757
Стартовая панель ... 1757
Переменные (зависимые и независимые) ... 1757
Ковариаты ... 1758
Коды для межгрупповых факторов. ... 1758
Планы с повторными измерениями (внутригрупповая сумма квадратов) ... 1758
Гнездовые (вложенные) планы ... 1759
Случайные факторы. ... 1759
Изолированная контрольная группа ... 1759
Регрессионный подход ... 1759
Задание планов с повторными измерениями ... 1760
Задание более одного фактора повторных измерений ... 1760
Многомерные планы с повторными измерениями ... 1760
Задание кода для изолированной контрольной группы ... 1761
Задание гнездовых (вложенных) планов ... 1761
Гнездовые (вложенные) планы ... 1762
Задание случайных факторов (эффектов) ... 1762
Планы с несколькими случайными эффектами ... 1762
Фиксированные и переменные ковариаты ... 1762
Переменные ковариаты ... 1762
Диалоговое окно *Результаты* ... 1763
Таблица всех эффектов ... 1763
Средние/графики ... 1763
Специфический эффект/средние/графики ... 1763
Спланированные сравнения ... 1763
Опции ... 1764
Объединенные эффекты и члены ошибки ... 1764
Апостериорные сравнения ... 1764

Copyright © StatSoft, 1995

Описательные статистики и графики ... 1764
Регрессия внутри ячеек ... 1764
Опции вывода ... 1765
Распечатка плана ... 1765
Опции вывода ... 1765
Типы гипотез ... 1765
Критерии для внутригрупповых планов (планов с повторными измерениями) ... 1766
Ковариационный анализ ... 1766
Многомерный дисперсионный анализ ... 1767
Проверка специфического эффекта ... 1767
Результаты регрессии (ковариаты) ... 1767
Сводка результатов для ковариат ... 1768
B-веса и *бета - веса* ... 1768
Выбор эффектов для графиков средних (взаимодействий) или таблиц средних ... 1768
Отобразить ... 1769
Назначение факторов осям графика ... 1769
Апостериорные сравнения ... 1769
Средние ... 1770
Критерий НЗР или спланированные сравнения ... 1770
Критерий Шеффе ... 1770
Критерий Ньюмена - Кеулса и критические области ... 1770
Критерий множественных сравнений Дункана и критические области ... 1770
Критерий ДЗР Тьюки (достоверно значимой разности) ... 1771
Критерий ДЗР Тьюки для неравных объемов выборок (критерии Спйотвола и Столайна) ... 1771
Описательные статистики и критерии проверки предположений дисперсионного анализа ... 1771
Описательные статистики ... 1771
Наглядные графики ... 1772
Исследование предположений дисперсионного анализа ... 1773
Выбор группы для таблицы или графика ... 1776
Критерии однородности дисперсии ... 1776
Многомерные критерии ... 1777
Сжатие и расширение графика ствол и листья ... 1777
Объединенные эффекты и члены ошибки ... 1778
Эффект ... 1778
Ошибка ... 1778
Задание объединенных эффектов (объединение эффектов) ... 1779
Порядок эффектов в последовательных (тип I) суммах квадратов ... 1779
Переупорядочивание эффектов для сумм квадратов типа I ... 1780
Уровень значимости, выделяемый в таблице всех эффектов ... 1780
Таблицы и графики средних значений ... 1780
Опции для спланированных сравнений ... 1781
Ввод контрастов отдельно для каждого фактора ... 1781
Задание множества контрастов ... 1781
Вставка значения ... 1782

StatSoft®

Copyright © StatSoft, 1995

Предопределенные контрасты ... 1782
Типы предопределенных контрастов ... 1782
Полиномиальные коэффициенты ... 1782
Предопределенные контрасты ... 1783
Задание контрастных векторов ... 1783

ПРИМЕЧАНИЯ, ЧАСТЬ I: ПРИМЕРЫ АНАЛИЗА ПЛАНОВ ОБЩЕГО ТИПА ... 1784

Задание одномерных и многомерных межгрупповых планов ... 1784
Однофакторный межгрупповой дисперсионный анализ и анализ контрастов ... 1784
Многофакторный межгрупповой дисперсионный анализ и анализ контрастов ... 1786
Задание кодов для изолированной контрольной группы ... 1788
Задание внутригрупповых (с повторными измерениями) одномерных и многомерных планов ... 1790
Однофакторный внутригрупповой (с повторными измерениями) дисперсионный анализ ... 1790
Однофакторный внутригрупповой дисперсионный анализ с многомерными повторными измерениями (*MANOVA*) ... 1792
Многофакторные внутригрупповые (с повторными измерениями) планы дисперсионного анализа ... 1792
Многомерные многофакторные внутригрупповые (с повторными измерениями) планы ... 1795
Меж- и внутригрупповые одномерные и многомерные планы ... 1795
Многомерный дисперсионный анализ зависимых измерений ... 1797
Гнездовые планы ... 1797
Простые межгрупповые гнездовые планы ... 1797
Простые внутригрупповые (с повторными измерениями) планы ... 1798
Латинские квадраты, греко-латинские квадраты, обобщенные латинские квадраты ... 1799
Другие специальные планы ... 1800
Планы с одним наблюдением в ячейке ... 1800
Случайные блочные планы ... 1800
Случайные эффекты ... 1801
Сравнение экспериментов ... 1802
Задание плана ... 1802
Исследование главных эффектов и взаимодействий ... 1802
Фиксированные и переменные ковариаты в межгрупповых и внутригрупповых планах ... 1802
Фиксированные ковариаты ... 1802
Переменные ковариаты ... 1803
Анализ полиномиальных контрастов ... 1804

ПРИМЕЧАНИЯ, ЧАСТЬ II: ПРИМЕРЫ ДИСПЕРСИОННОГО АНАЛИЗА НЕСБАЛАНСИРОВАННЫХ И НЕПОЛНЫХ ПЛАНОВ ... 1806

Пример1: Несбалансированные планы и сумма квадратов типа I и II ... 1806
Пример 2: Простой гнездовой (вложенный) план ... 1807
Пример 3: Трехфакторный гнездовой (вложенный) план. ... 1808
Пример 4: Анализ плана с повторными измерениями как гнездового (вложенного) плана со случайными эффектами ... 1809
Пример 5: Планы с расщепленными делянками ... 1811
Пример 6: Планы с полосными делянками ... 1812

Copyright © StatSoft, 1995

Пример 7: Планы с расщепленными делянками с неравным числом подделянок 1813
Пример 8: Планы типа квадрата Юдена 1814
Планы с пропущенными ячейками 1814
Пример 9: План 3 x 3 с пропущенными ячейками 1814
Пример 10: Взвешенные средние Анализ гнездового плана с неравным числом наблюдений в ячейках 1815
Пример 11: Гнездовой план 4 x 11 с неравным количеством уровней 1817
Пример 12: Сложные планы с большим количеством пропущенных ячеек (Исследование гипотез типа IV) 1817

ТЕХНИЧЕСКИЕ ЗАМЕЧАНИЯ 1820
Вычислительные процедуры 1820
Использование библиотеки матричных функций *STATISTICA BASIC* в многомерном дисперсионном анализе 1820
Многомерный дисперсионный анализ и моделирование структурными уравнениями 1820

ПРЕДМЕТНЫЙ УКАЗАТЕЛЬ 1822

StatSoft®

Copyright © StatSoft, 1995

Глава 13

Дисперсионный анализ (ANOVA/MANOVA)

ВВОДНЫЙ ОБЗОР

В этом разделе мы рассмотрим основные методы, предположения и терминологию дисперсионного анализа. Дополнительная информация о дисперсионном анализе некоторых специальных планов содержится далее в соответствующих разделах этой главы (так, межгрупповые планы даны на стр. 1784, планы с повторными измерениями - на стр. 1790). Подробные примеры задания планов и проведения анализа рассматриваются в разделе *Примеры* (стр. 1738).

Отметим, что в англоязычной литературе дисперсионный анализ обычно называется анализом вариации. Поэтому, для краткости, ниже мы иногда будем использовать термин *ANOVA* (*Analysis of variation*) для обычного дисперсионного анализа и термин *MANOVA* для многомерного дисперсионного анализа. В этом разделе мы последовательно рассмотрим основные идеи дисперсионного анализа (*ANOVA*), ковариационного анализа (*ANCOVA*), многомерного дисперсионного анализа (*MANOVA*) и многомерного ковариационного анализа (*MANCOVA*). После краткого обсуждения достоинств анализа контрастов и апостериорных критериев рассмотрим предположения, на которых основаны методы дисперсионного анализа. Ближе к концу этого раздела поясняются преимущества многомерного подхода для анализа повторных измерений по сравнению с традиционным одномерным подходом.

Основные идеи

Цель дисперсионного анализа. Основной целью дисперсионного анализа является исследование значимости различия между средними. Глава *Элементарные понятия статистики* (глава 8) содержит краткое введение в исследование статистической значимости. Если вы просто сравниваете средние в двух выборках, дисперсионный анализ даст тот же результат, что и обычный *t*-критерий для независимых выборок (если сравниваются две независимые группы объектов или наблюдений) или *t*-критерий для зависимых выборок (если сравниваются две переменные на одном и том же множестве объектов или наблюдений). Если вы не достаточно знакомы с этими критериями, рекомендуем обратиться к вводному обзору главы *Основные статистики и таблицы* (глава 9).

Откуда произошло название *Дисперсионный анализ*? Может показаться странным, что процедура сравнения средних называется дисперсионным анализом. В действительности, это связано с тем, что при исследовании статистической значимости различия между средними, мы на самом деле анализируем дисперсии.

Разбиение суммы квадратов.

Для выборки объема n выборочная дисперсия вычисляется как сумма квадратов отклонений от выборочного среднего, деленная на n-1 (объем выборки минус единица). Таким образом, при фиксированном объеме выборки n дисперсия есть функция суммы квадратов (отклонений), обозначаемая, для краткости, *SS* (от английского Sum of Squares – Сумма Квадратов). В основе

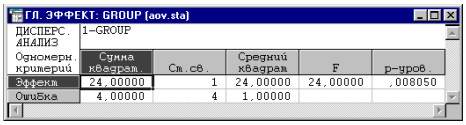
Copyright © StatSoft, 1995

StatSoft®

дисперсионного анализа лежит разделение (или разбиение) дисперсии на части. Рассмотрим следующий набор данных:

	Группа 1	Группа 2
Наблюдение 1	2	6
Наблюдение 2	3	7
Наблюдение 3	1	5
Среднее	2	6
Сумма квадратов (*SS*)	2	2
Общее среднее	4	
Общая сумма квадратов	28	

Средние двух групп существенно различны (2 и 6 соответственно). Сумма квадратов отклонений *внутри* каждой группы равна 2. Складывая их, получаем 4. Если теперь повторить эти вычисления *без учета* групповой принадлежности, то есть, если вычислить *SS* исходя из общего среднего этих двух выборок, то получим 28. Иными словами, дисперсия (сумма квадратов), основанная на внутригрупповой изменчивости, приводит к гораздо меньшим значениям, чем при вычислении на основе общей изменчивости (относительно общего среднего). Причина этого, очевидно, заключается в существенной разнице между средними значениями, и это различие между средними и объясняет существующее различии между суммами квадратов. В самом деле, если использовать для анализа приведенных данных модуль *Дисперсионный анализ*, будут получены следующие результаты:



Как видно из таблицы, общая сумма квадратов *SS* = 28 разбита на сумму квадратов, обусловленную *внутригрупповой* изменчивостью (*2+2=4*; см. вторую строку таблицы) и сумму квадратов, обусловленную различием средних значений. (28-(2+2)=24; см первую строку таблицы).

***SS* ошибок и *SS* эффекта.** Внутригрупповая изменчивость (*SS*) обычно называется дисперсией *ошибки.* Это означает, что обычно при проведении эксперимента она не может быть предсказана или объяснена. С другой стороны, *SS эффекта* (или межгрупповую изменчивость) можно объяснить различием между средними значениями в изучаемых группах. Иными словами, принадлежность к некоторой группе *объясняет* межгрупповую изменчивость, т.к. нам известно, что эти группы обладают разными средними значениями.

Проверка значимости. Основные идеи проверки статистической значимости обсуждаются в главе *Элементарные понятия статистики* (глава 8). В этой же главе объясняются причины, по которым многие критерии используют отношение объясненной и необъясненной дисперсии. Примером такого использования является сам дисперсионный анализ. Проверка значимости в дисперсионном анализе основана на сравнении дисперсии, обусловленной межгрупповым разбросом (называемой *средним квадратом эффекта* или $MS_{эффект}$) и дисперсии, обусловленной внутригрупповым разбросом (называемой *средним квадратом ошибки* или $MS_{ошибка}$). Если верна нулевая гипотеза (равенство средних в двух популяциях), то можно ожидать сравнительно небольшое различие в выборочных средних из-за случайной изменчивости. Поэтому при нулевой гипотезе внутригрупповая дисперсия будет практически совпадать с общей дисперсией, подсчитанной без учета группой принадлежности. Полученные внутригрупповые дисперсии можно сравнить с помощью *F*-критерия, проверяющего, действительно ли отношение дисперсий значимо больше 1. В рассмотренном выше примере *F*-критерий показывает, что различие между средними статистически значимо.

StatSoft®

Copyright © StatSoft, 1995

Основная логика дисперсионного анализа. Подводя итоги, можно сказать, что целью дисперсионного анализа является проверка статистической значимости разницы между средними (для групп или переменных). Эта проверка проводится с помощью анализа дисперсии, т.е. с помощью разбиения общей дисперсии (вариации) на части, одна из которых обусловлена случайной ошибкой (то есть внутригрупповой изменчивостью), а вторая связана с различием средних значений. Последняя компонента дисперсии затем используется для анализа статистической значимости различия между средними значениями. Если это различие значимо, нулевая гипотеза отвергается и принимается альтернативная гипотеза о существовании различия между средними.

Зависимые и независимые переменные. Переменные, значения которых определяется с помощью измерений в ходе эксперимента (например, балл, набранный при тестировании), называются *зависимыми* переменными. Переменные, которыми можно управлять при проведении эксперимента (например, методы обучения или другие критерии, позволяющие разделить наблюдения на группы) называются *факторами* или *независимыми* переменными. Более подробно эти понятия описаны в главе *Элементарные понятия статистики* (глава 8).

Многофакторный дисперсионный анализ

В рассмотренном выше простом примере вы могли бы сразу вычислить t-критерий для независимых выборок, используя соответствующую опцию модуля *Основные статистики и таблицы.* Полученные результаты, естественно, совпадут с результатами дисперсионного анализа. Однако дисперсионный анализ содержит гибкие и мощные технические средства, которые могут быть использованы для гораздо более сложных исследований.

Множество факторов. Мир по своей природе сложен и многомерен. Ситуации, когда некоторое явление полностью описывается одной переменной, чрезвычайно редки. Например, если мы пытаемся научиться выращивать большие помидоры, следует рассматривать факторы, связанные с генетической структурой растений, типом почвы, освещенностью, температурой и т.д. Таким образом, при проведении типичного эксперимента приходится иметь дело с большим количеством факторов. Основная причина, по которой использование дисперсионного анализа предпочтительнее повторного сравнения двух выборок при разных уровнях факторов с помощью *t*-критерия, заключается в том, что дисперсионный анализ более *эффективен* и, для малых выборок, более информативен.

Управление факторами. Предположим, что в рассмотренном выше примере анализа двух выборок мы добавим еще один фактор, например, *Пол - Gender*. Пусть каждая группа состоит из 3 мужчин и 3 женщин. План этого эксперимента можно представить в виде таблицы 2 на 2:

	Эксперимент. Группа 1	Эксперимент. Группа 2
Мужчины	2	6
	3	7
	1	5
Среднее	2	6
Женщины	4	8
	5	9
	3	7
Среднее	4	8

До проведения вычислений, можно заметить, что в этом примере общая дисперсия имеет, по крайней мере, три источника:

(1) случайная ошибка (внутригрупповая дисперсия),

Copyright © StatSoft, 1995

StatSoft®

(2) изменчивость, связанная с принадлежностью к экспериментальной группе, и

(3) изменчивость, обусловленная полом объектов наблюдения.

(Отметим, что существует еще один возможный источник изменчивости – *взаимодействие факторов*, который мы обсудим позднее). Что произойдет, если мы не будем включать *пол – gender* как фактор при проведении анализа и вычислим обычный *t*-критерий? Если мы будем вычислять суммы квадратов, игнорируя *пол – gender* (т.е., объединяя объекты разного пола в одну группу при вычислении внутригрупповой дисперсии, получив при этом сумму квадратов для каждой группы равную *SS* =10, и общую сумму квадратов *SS* = 10+10 = 20), то получим большее значение внутригрупповой дисперсии, чем при более точном анализе с дополнительным разбиением на подгруппы по *полу - gender* (при этом внутригрупповые средние будут равны 2, а общая внутригрупповая сумма квадратов равна *SS* = 2+2+2+2 = 8). Это различие связано с тем, что среднее значение для *мужчин - males* меньше, чем среднее значение для *женщин – female*, и это различие в средних значениях увеличивает суммарную внутригрупповую изменчивость, если фактор пола не учитывается. Управление дисперсией ошибки увеличивает чувствительность (мощность) критерия.

На этом примере видно еще одно преимущество дисперсионного анализа по сравнению с обычным *t*-критерием для двух выборок. Дисперсионный анализ позволяет изучать каждый фактор, управляя значениями остальных факторов. Это, в действительности, и является основной причиной его большей статистической мощности (для получения значимых результатов требуются меньшие объемы выборок). По этой причине дисперсионный анализ даже на небольших выборках дает статистически более значимые результаты, чем простой *t*-критерий.

Эффекты взаимодействия

Существует еще одно преимущество применения дисперсионного анализа по сравнению с обычным *t*-критерием: дисперсионный анализ позволяет обнаружить *взаимодействие* между факторами и, следовательно, позволяет изучать более сложные модели. Для иллюстрации рассмотрим еще один пример.

Главные эффекты, попарные (двухфакторные) взаимодействия.

Предположим, что имеется две группы студентов, причем психологически студенты первой группы настроены на выполнение поставленных задач и более целеустремленны, чем студенты второй группы, состоящей из более ленивых студентов. Разобьем каждую группу случайным образом пополам и предложим одной половине в каждой группе сложное задание, а другой - легкое. После этого измерим, насколько напряженно студенты работают над этими заданиями. Средние значения для этого (вымышленного) исследования показаны в таблице:

Задание	Целеустремленные	Ленивые
Трудное	10	5
Легкое	5	10

Какой вывод можно сделать из этих результатов? Можно ли заключить, что: (1) над сложным заданием студенты трудятся более напряженно; (2) целеустремленные студенты работают упорнее, чем ленивые? Ни одно из этих утверждений не отражает сущность систематического характера средних, приведенных в таблице. Анализируя результаты, правильнее было бы сказать, что над сложными заданиями работают упорнее только целеустремленные студенты, в то время как над легкими заданиями только ленивые работают упорнее. Другими словами характер студентов и сложность задания *взаимодействуя* между собой влияют на затрачиваемое усилие. Это пример

StatSoft®

Copyright © StatSoft, 1995

парного взаимодействия между характером студентов и сложностью задания. Отметим, что утверждения 1 и 2 описывают *главные эффекты*.

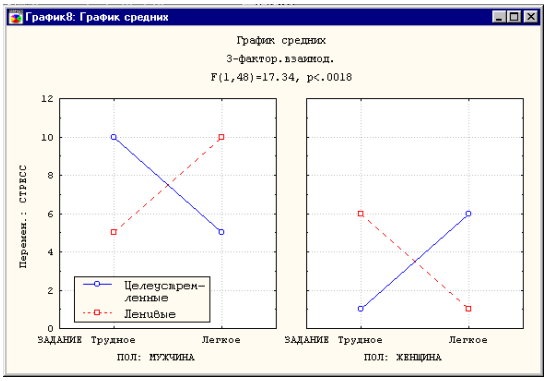
Взаимодействия высших порядков. В то время как объяснить попарные взаимодействия еще сравнительно легко, взаимодействия высших порядков объяснить значительно сложнее. Представим себе, что в рассматриваемый выше пример, введен еще один фактор *пол -Gender* и мы получили следующую таблицу средних значений:

	Женщины	
Задание	Целеустремленные	Ленивые
Трудное	10	5
Легкое	5	10

	Мужчины	
Задание	Целеустремленные	Ленивые
Трудное	1	6
Легкое	6	1

Какие теперь выводы можно сделать из полученных результатов? Графики средних позволяют легко интерпретировать сложные эффекты. Модуль дисперсионного анализа позволяет строить эти графики практически одним щелчком мышки.

Изображение на графиках внизу представляет собой изучаемое трехфакторное взаимодействие.



Глядя на графики, можно сказать, что у женщин существует взаимодействие между характером и сложностью теста: целеустремленные женщины работают над трудным заданием более напряженно, чем над легким. У мужчин это же взаимодействие носит обратный характер. Видно, что описание взаимодействия между факторами становится более запутанным.

Общий способ описания взаимодействий. В общем случае взаимодействие между факторами описывается в виде изменения одного эффекта под воздействием другого. В рассмотренном выше примере двухфакторное взаимодействие можно описать как изменение главного эффекта фактора, характеризующего сложность задачи, под воздействием фактора, описывающего характер студента. Для взаимодействия трех факторов из предыдущего параграфа можно сказать, что взаимодействие двух факторов (сложности задачи и характера студента) изменяется под воздействием *пола – Gender.* Если изучается взаимодействие четырех факторов, можно сказать, что взаимодействие трех факторов, изменяется под воздействием четвертого фактора, т.е. существуют различные типы взаимодействий на разных уровнях четвертого фактора. Оказалось, что во многих областях взаимодействие пяти или даже большего количества факторов не является чем-то необычным.

Сложные планы

Далее в этой главе будет разобрано много примеров сложных планов (межгрупповые планы описаны на стр. 1784, планы с повторными измерениями на стр. 1790) обсуждаются позднее. В следующем разделе будет дан обзор основных “кирпичиков”, из которых строятся эти сложные планы.

Copyright © StatSoft, 1995

StatSoft®

Межгрупповые и внутригрупповые планы (планы с повторными измерениями)

При сравнении двух различных групп обычно используется *t*-критерий для независимых выборок (из модуля *Основные статистики и таблицы*). Когда сравниваются две переменные на одном и том же множестве объектов (наблюдений), используется *t*-критерий для зависимых выборок. Для дисперсионного анализа также важно зависимы или нет выборки. Если имеются повторные измерения одних и тех же переменных (при разных условиях или в разное время) *для одних и тех же объектов*, то говорят о наличии *фактора повторных измерений* (называемого также *внутригрупповым фактором,* поскольку для оценки его значимости вычисляется внутригрупповая сумма квадратов). Если сравниваются разные группы объектов (например, мужчины и женщины, три штамма бактерий и т.п.), то разница между группами описывается *межгрупповым фактором.* Способы вычисления критериев значимости для двух описанных типов факторов различны, но общая их логика и интерпретации совпадает.

Меж- и внутригрупповые планы. Во многих случаях эксперимент требует включение в план и межгруппового фактора, и фактора повторных измерений. Например, измеряются математические навыки студентов женского и мужского пола (где *пол –Gender* -межгрупповой фактор) в начале и в конце семестра. Два измерения навыков каждого студента образуют внутригрупповой фактор (фактор повторных измерений). Интерпретация главных эффектов и взаимодействий для межгрупповых факторов и факторов повторных измерений совпадает, и оба типа факторов могут, очевидно, взаимодействовать между собой (например, женщины приобретают навыки в течение семестра, а мужчины их теряют).

Неполные (гнездовые) планы

Во многих случаях можно пренебречь эффектом взаимодействия. Это происходит или когда известно, что в популяции эффект взаимодействия отсутствует, или когда осуществление полного *факторного* плана невозможно. Например, изучается влияние четырех добавок к топливу на расход горючего. Выбираются четыре автомобиля и четыре водителя. Полный *факторный* эксперимент требует, чтобы каждая комбинация: добавка, водитель, автомобиль - появились хотя бы один раз. Для этого нужно не менее 4 x 4 x 4 = 64 групп испытаний, что требует слишком больших временных затрат. Кроме того, вряд ли существует взаимодействие между водителем и добавкой к топливу. Принимая это во внимание, можно использовать план *Латинские квадраты,* в котором содержится лишь 16 групп испытаний (четыре добавки обозначаются буквами A, B, C и D):

	Автомобиль 1	2	3	4
Водитель 1	A	B	C	D
Водитель 2	B	C	D	A
Водитель 3	C	D	A	B
Водитель 4	D	A	B	C

Латинские квадраты описаны в большинстве книг по планированию экспериментов (например, Hays, 1988; Lindman, 1974; Milliken and Johnson, 1984; Winer, 1962), и здесь они не будут детально обсуждаться. Отметим, что латинские квадраты это *неполные* планы, в которых участвуют не все комбинации уровней факторов. Например, водитель 1 управляет автомобилем 1 только с добавкой A, водитель 3 управляет автомобилем 1 только с добавкой C. Уровни фактора *добавок (*A, B, C и D) вложены в ячейки таблицы *автомобиль* x *водитель* – как яйца в гнезда. Это мнемоническое правило полезно для понимания природы *гнездовых или вложенных* планов. Модуль *Дисперсионный*

StatSoft®

Copyright © StatSoft, 1995

анализ предоставляет простые способы анализ планов такого типа (см. стр. 1797).

Ковариационный анализ

Основная идея

В разделе *Основные идеи* (стр. 1715) кратко обсуждалась идея управления факторами и то, каким образом включение аддитивных факторов позволяет уменьшать сумму квадратов ошибок и увеличивать статистическую мощность плана. Все это может быть распространено и на переменные с непрерывным множеством значений. Когда такие непрерывные переменные включаются в план в качестве факторов, они называются *ковариатами*.

Фиксированные ковариаты

Предположим, что сравниваются математические навыки двух групп студентов, которые обучались по двум различным учебникам. Предположим также, что имеются данные о коэффициенте интеллекта (IQ) для каждого студента. Можно предположить, что коэффициент интеллекта связан с математическими навыками, и использовать эту информацию. Для каждой из двух групп студентов можно вычислить коэффициент корреляции между IQ и математическими навыками. Используя этот коэффициент корреляции, можно выделить долю дисперсии в группах, объясняемую влиянием IQ и необъясняемую долю дисперсии (см. также *Элементарные понятия статистики* (глава 8) и *Основные статистики и таблицы* (глава 9)). Оставшаяся доля дисперсии используется при проведении анализа как дисперсия ошибки. Если имеется корреляция между IQ и математическими навыками, то можно существенно уменьшить дисперсии ошибки *SS/(n-1)*.

Влияние ковариат на *F* критерий. *F* критерий оценивает статистическую значимость различия средних значений в группах, при этом вычисляется отношение межгрупповой дисперсии (MS_{effect}) к дисперсии ошибок (MS_{error}). Если MS_{error} уменьшается, например, при учете фактора IQ, значение *F* увеличивается.

Множество ковариат. Рассуждения, использованные выше для одной ковариаты (IQ), легко распространяются на несколько ковариат. Например, кроме IQ, можно включить измерение мотивации, пространственного мышления и т.д. Вместо обычного коэффициента корреляции при этом используется множественный коэффициент корреляции (см. главу 12 *Множественная регрессия*).

Когда значение *F*-критерия уменьшается. Иногда введение ковариат в план эксперимента уменьшает значение *F*-критерия. Обычно это указывает на то, что ковариаты коррелированы не только с зависимой переменной (например, математическими навыками), но и с факторами (например, с разными учебниками). Предположим, что IQ измеряется в конце семестра, после почти годового обучения двух групп студентов по двум разным учебникам. Хотя студенты разбивались на группы случайным образом, может оказаться, что различие учебников настолько велико, что и IQ и математические навыки в разных группах будут сильно различаться. В этом случае, ковариаты не только уменьшают дисперсию ошибок, но и межгрупповую дисперсию. Другими словами, после контроля за разностью IQ в разных группах, разность в математических навыках уже будет несущественной. Можно сказать иначе. После “исключения” влияния IQ, неумышленно исключается и влияние учебника на развитие математических навыков.

Скорректированные средние. Когда ковариата влияет на межгрупповой фактор,

Copyright © StatSoft, 1995

StatSoft®

следует вычислять *скорректированные средние*, т.е. такие средние, которые получаются после удаления всех оценок ковариат.

Взаимодействие между ковариатами и факторами. Также как исследуется взаимодействие между факторами, можно исследовать взаимодействие между ковариатами и между группами факторов. Предположим, что один из учебников особенно подходит для умных студентов. Второй учебник для умных студентов скушен, а для менее умных студентов этот же учебник труден. В результате имеется положительная корреляция между IQ и результатом обучения в первой группе (более умные студенты, лучше результат) и нулевая или небольшая отрицательная корреляция во второй группе (чем умнее студент, тем менее вероятно приобретение математических навыков из второго учебника). В некоторых исследованиях эта ситуация обсуждается как пример нарушения предположений ковариационного анализа (см. *Предположения и эффекты нарушения предположений*, стр. 1730). Однако так как в модуле Дисперсионный анализ используются самые общие способы ковариационного анализа, можно, в частности, оценить статистическую значимость взаимодействия между факторами и ковариатами.

Переменные ковариаты

В то время как фиксированные ковариаты обсуждаются в учебниках достаточно часто, переменные ковариаты упоминаются намного реже. Обычно, при проведении экспериментов с повторными измерениями, нас интересуют различия в измерениях одних и тех же величин в разные моменты времени. А именно, нас интересует значимость этих различий. Если одновременно с измерениями зависимых переменных проводится измерение ковариат, можно вычислить корреляцию между ковариатой и зависимой переменной.

Например, можно изучать интерес к математике и математические навыки в начале и в конце семестра. Интересно было бы проверить, коррелированы ли между собой изменения в интересе к математике с изменением математических навыков.

Модуль *Дисперсионный анализ* в *STATISTICA* автоматически оценивает статистическую значимость изменения ковариат в тех планах, где это возможно.

Многомерные планы: многомерный дисперсионный и ковариационный анализ

Межгрупповые планы

Все рассматриваемые ранее примеры включали только одну зависимую переменную. Когда одновременно имеется несколько зависимых переменных, возрастает лишь сложность вычислений, а содержание и основные принципы не меняются.

Например, проводится исследование двух различных учебников. При этом изучаются успехи студентов в изучении физики и математики. В этом случае имеются две зависимые переменные и нужно выяснить, как влияют на них одновременно два разных учебника. Для этого можно воспользоваться многомерным дисперсионным анализом (MANOVA). Вместо одномерного F критерия, используется многомерный F критерий (λ-критерий Уилкса), основанный на сравнении ковариационной матрицы ошибок и межгрупповой ковариационной матрицы.

Если зависимые переменные коррелированы между собой, то эта корреляция должна учитываться при вычислении критерия значимости. Очевидно, если одно и то же измерение повторяется дважды, то ничего нового

StatSoft®

Copyright © StatSoft, 1995

получить при этом нельзя. Если к имеющемуся измерению добавляется коррелированное с ним измерение, то получается некоторая новая информация, но при этом новая переменная содержит избыточную информацию, которая отражается в ковариации между переменными.

Интерпретация результатов. Если общий многомерный критерий значим, можно заключить, что соответствующий эффект (например, тип учебника) значим. Однако встают следующие вопросы. Влияет ли тип учебника на улучшение только математических навыков, только физических навыков, или одновременно на улучшение тех и других навыков. В действительности, после получения значимого многомерного критерия, для отдельного главного эффекта или взаимодействия исследуется одномерный *F* критерий. Другими словами, отдельно исследуются зависимые переменные, которые вносят вклад в значимость многомерного критерия.

Планы с повторными измерениями

Если измеряются математические и физические навыки студентов в начале семестра и в конце, то это и есть повторные измерения. Изучение критерия значимости в таких планах это логическое развитие одномерного случая (см. стр. 1720). Заметим, что методы многомерного дисперсионного анализа обычно также используются для исследования значимости одномерных факторов повторных измерений, имеющих более чем два уровня. Соответствующие применения будут рассмотрены позднее в этой части.

Суммирование значений переменных и многомерный дисперсионный анализ

Даже опытные пользователи одномерного и многомерного дисперсионного анализа часто приходят в затруднение, получая разные результаты при применении многомерного дисперсионного анализа, например, для трех переменных, и при применении одномерного дисперсионного анализа к сумме этих трех переменных, как к одной переменной.

Идея *суммирования* переменных состоит в том, что каждая переменная содержит в себе некоторую истинную переменную, которая и исследуется, а также случайную ошибку измерения. Поэтому при усреднении значений переменных, ошибка измерения будет ближе к 0 для всех измерений и усредненное значений будет более надежным. На самом деле, в этом случае применение дисперсионного анализа к сумме переменных разумно и является мощным методом. Однако если зависимые переменные по своей природе многомерны, суммирование значений переменных неуместно.

Например, пусть зависимые переменные состоят из четырех показателей *успеха в обществе*. Каждый показатель характеризует совершенно независимую сторону человеческой деятельности (например, профессиональный успех, преуспеваемость в бизнесе, семейное благополучие и т.д.). Сложение этих переменных подобно сложению яблока и апельсина. Сумма этих переменных не будет подходящим одномерным показателем. Поэтому с такими данными нужно обходится как с многомерными показателями в *многомерном дисперсионном анализе*.

Анализ контрастов и апостериорные критерии

Почему сравниваются отдельные множества средних?

Обычно гипотезы относительно экспериментальных данных формулируются не просто в терминах главных эффектов или

Copyright © StatSoft, 1995

взаимодействий. Примером может служить такая гипотеза: некоторый учебник повышает математические навыки только у студентов мужского пола, в то время как другой учебник примерно одинаково эффективен для обоих полов, но все же менее эффективен для мужчин. Можно предсказать, что эффективность учебника взаимодействует с полом студента. Однако этот прогноз касается также *природы* взаимодействия. Ожидается значительное различие между полами для обучающихся по одной книге и практически не зависимые от пола результаты для обучающихся по другой книге. Такой тип гипотез обычно исследуется с помощью анализа контрастов.

Анализ контрастов

Если говорить коротко, то анализ контрастов позволяет оценивать статистическую значимость некоторых линейных комбинаций эффектов сложного плана. Анализ контрастов главный и обязательный элемент любого сложного плана дисперсионного анализа. Модуль *Дисперсионный анализ* имеет достаточно разнообразные возможности анализа контрастов, которые позволяют выделять и анализировать любые типы сравнений средних (см. стр. 1785 для описания контрастов).

Апостериорные сравнения

Иногда в результате обработки эксперимента обнаруживается неожиданный эффект. Хотя в большинстве случаев творческий исследователь сможет объяснить любой результат, это не дает возможностей для дальнейшего анализа и получения оценок для прогноза. Эта проблема является одной из тех, для которых используются *апостериорные критерии*, то есть критерии, не использующие *априорные* гипотезы. Для иллюстрации рассмотрим следующий эксперимент. Предположим, что на 100 карточках записаны числа от 1 до 10. Опустив все эти карточки в шапку, мы случайным образом выбираем 20 раз по 5 карточек, и вычисляем для каждой выборки среднее значение (среднее чисел, записанных на карточки). Можно ли ожидать, что найдутся две выборки, у которых средние значения значимо отличаются? Это очень правдоподобно! Выбирая две выборки с максимальным и минимальным средним, можно получить разность средних, сильно отличающуюся от разности средних, например, первых двух выборок. Эту разность можно исследовать, например, с помощью анализа контрастов. Если не вдаваться в детали, то существует несколько, так называемых *апостериорных* критериев, которые основаны в точности на первом сценарии (взятие экстремальных средних из 20 выборок), т. е. эти критерии основаны на выборе наиболее отличающихся средних для сравнения всех средних значений в плане. Эти критерии применяются для того, чтобы чисто случайно не получить искусственный эффект, например, обнаружить значимое различие между средними, когда его нет. Модуль *Дисперсионный анализ* предлагает широкий выбор таких критериев. Когда в эксперименте, связанном с несколькими группами, встречаются неожиданные результаты, то используются *апостериорные* процедуры для исследования статистической значимости полученных результатов.

Сумма квадратов типа I, II, III и IV

Многомерная регрессия и дисперсионный анализ

Существует тесная взаимосвязь между методом многомерной регрессии и дисперсионным анализом (анализом вариаций). И в том и в другом методе исследуется линейная модель. Для более подробного обсуждения линейной

StatSoft®

Copyright © StatSoft, 1995

модели можно обратиться к разделу *Технические замечания* (стр. 1820). Если говорить коротко, то практически все планы эксперимента можно исследовать с помощью многомерной регрессии. Рассмотрим следующий простой межгрупповой 2 x 2 план.

DV	A	B	AxB
3	1	1	1
4	1	1	1
4	1	-1	-1
5	1	-1	-1
6	-1	1	-1
6	-1	1	-1
3	-1	-1	1
2	-1	-1	1

Столбцы A и B содержат коды, характеризующие уровни факторов A и B, столбец AxB содержит произведение двух столбцов A и B. Мы можем анализировать эти данные с помощью многомерной регрессии. Переменная *DV* определяется как зависимая переменная, переменные от *A* до *AxB* как независимые переменные. Исследование значимости для коэффициентов регрессии будет совпадать с вычислениями в дисперсионном анализе значимости главных эффектов факторов *A* и *B* и эффекта взаимодействия *AxB*.

Несбалансированные и сбалансированные планы

При вычислении корреляционной матрицы для всех переменных, например, для данных, изображенных выше, можно заметить, что главные эффекты факторов *A* и *B* и эффект взаимодействия *AxB* некоррелированы. Это свойство эффектов называют также ортогональностью. Говорят, что эффекты *A* и *B* - *ортогональны* или *независимы* друг от друга. Если все эффекты в плане ортогональны друг другу, как в приведенном выше примере, то говорят, что план *сбалансирован*.

Сбалансированные планы обладают “хорошим свойством”. Вычисления при анализе таких планов очень просты. Все вычисления сводятся к вычислению корреляции между эффектами и зависимыми переменными. Так как эффекты ортогональны, частные корреляции (как в полной *многомерной* регрессии) не вычисляются. Однако в реальной жизни планы не всегда сбалансированы.

Рассмотрим реальные данные с неравным числом наблюдений в ячейках.

	Фактор B	
Фактор A	B1	B2
A1	3	4, 5
A2	6, 6, 7	2

Если закодировать эти данные как выше и вычислить корреляционную матрицу для всех переменных, то окажется, что факторы плана коррелированы друг с другом. Факторы в плане теперь не ортогональны и такие планы называются *несбалансированными.* Заметим, что в рассматриваемом примере, корреляция между факторами полностью связана с различием частот 1 и -1 в столбцах матрицы данных. Другими словами, планы экспериментов с неравными объемами ячеек (точнее, непропорциональными объемами) будут несбалансированными, это означает, что главные эффекты и взаимодействия будут смешиваться. В этом случае для вычисления статистической значимости эффектов нужно полностью вычислять многомерную регрессию. Здесь имеется несколько стратегий.

Copyright © StatSoft, 1995

Сумма квадратов типа I, II, III и IV

Сумма квадратов типа I и III. Для изучения значимости каждого фактора в многомерной модели можно вычислять частную корреляцию каждого фактора, при условии, что все другие факторы уже учтены в модели. Можно также вводить факторы в модель пошаговым способом, фиксируя все факторы, уже введенные в модель и игнорируя все остальные факторы. Вообще, в этом и состоит различие между *типом III* и *типом I* суммы квадратов (эта терминология была введена в SAS, см. например, SAS, 1982; подробное обсуждение можно также найти в Searle, 1987, стр. 461; Woodward, Bonett, and Brecht, 1990, стр. 216; или Milliken and Johnson, 1984, стр. 138).

Сумма квадратов типа II. Следующая "промежуточная" стратегия формирования модели состоит: в контроле всех главных эффектов при исследовании значимости отдельного главного эффекта; в контроле всех главных эффектов и всех попарных взаимодействий, когда исследуется значимость отдельного попарного взаимодействия; в контроле всех главных эффектов всех попарных взаимодействий и всех взаимодействий трех факторов; при исследовании отдельного взаимодействия трех факторов и т. д. Суммы квадратов для эффектов, вычисляемые таким способом, называются *типом II* суммы квадратов. Итак, *тип II* суммы квадратов контролирует все эффекты того же порядка и ниже, игнорируя все эффекты более высокого порядка.

Сумма квадратов типа IV. Наконец, для некоторых специальных планов с пропущенными ячейками (неполными планами) можно вычислять, так называемые, *типа IV* суммы квадратов. Этот метод будет обсуждаться позднее в связи с неполными планами (планами с пропущенными ячейками).

Интерпретация гипотезы о сумме квадратов типа I, II, и III

Сумму квадратов *типа III* легче всего интерпретировать. Напомним, что суммы квадратов *типа III* исследуют эффекты после контроля всех других эффектов. Например, после нахождения статистически значимого *типа III* эффекта для фактора *A* в модуле *Дисперсионный анализ*, можно сказать, что существует единственный значимый эффект фактора *A*, после введения всех других эффектов (факторов) и соответственно интерпретировать этот эффект. Вероятно в 99% всех приложений дисперсионного анализа именно этот тип критерия интересует исследователя. Этот тип суммы квадратов обычно вычисляется в модуле *Дисперсионный анализ* по умолчанию, независимо от того выбрана опция *Регрессионный подход* или нет (стандартные подходы принятые в модуле *Дисперсионный анализ* обсуждаются ниже).

Значимые эффекты, полученные с помощью сумм квадратов *типа* или *типа II* суммы квадратов интерпретировать не так легко. Лучше всего их интерпретировать в контексте пошаговой многомерной регрессии (см., например, обсуждение в главе 12). Если при использовании суммы квадратов *типа I* главный эффект фактора B оказался значим (после включения в модель фактора A, но перед добавлением взаимодействия между A и B), можно заключить, что существует значимый главный эффект фактора B, при условии, что нет взаимодействия между факторами A и B. (Если при использовании критерия *типа III*, фактор B также оказался значимым, то можно заключить, что существует значимый главный эффект фактора B, после введения в модель всех других факторов и их взаимодействий).

StatSoft®

Copyright © StatSoft, 1995

В терминах маргинальных средних гипотезы *типа I* и *типа II* обычно не имеют простой интерпретации. В этих случаях говорят, что нельзя интерпретировать значимость эффектов, рассматривая только маргинальные средние. Скорее представленные *p* значений средних имеют отношение к сложной гипотезе, которая комбинирует средние и объем выборки. Например, *тип II* гипотезы для фактора A в простом примере плана 2 x 2, рассматриваемом ранее (стр. 1725) будут (см. Woodward, Bonett, and Brecht, 1990, стр. 219):

$$\Sigma(n_{ij}-n_{ji}^{2}/n_{.j})*u_{ij} = \Sigma\Sigma n_{ij}*n_{i'j}/n_{.j}*u_{i'j}$$

где

n_{ij} число наблюдений в ячейке

u_{ij} среднее значение в ячейке

$n_{.j}$ маргинальное среднее.

Если не вдаваться в детали (более подробно см. Milliken and Johnson, 1984, глава 10), то ясно, что это не простые гипотезы и в большинстве случаев ни одна из них не представляет особенного интереса у исследователя. Однако существуют случаи, когда гипотезы *типа I* могут быть интересны.

Принимаемый по умолчанию вычислительный подход в модуле *Дисперсионный анализ*

По умолчанию, если не отмечена опция *Регрессионный подход* (см. стартовую панель модуля *Дисперсионный анализ*, стр. 1757), модуль *Дисперсионный анализ* использует *модель средних по ячейкам*. Для этой модели характерно, что суммы квадратов для разных эффектов вычисляются для линейных комбинаций средних значений по ячейкам. В полном факторном эксперименте это приводит к суммам квадратов, которые совпадают с суммами квадратов, обсуждаемыми ранее как *тип III*. Однако в опции *Спланированные сравнения* (в окне *Результаты дисперсионного анализа*, см. стр. 1763), пользователь может проверять гипотезу относительно любой линейной комбинации взвешенных или невзвешенных средних по ячейкам. Таким образом, пользователь может проверять не только гипотезы *типа III*, но гипотезы любого типа (включая *тип IV*). Как будет показано в разделе *Примечания, часть II* (стр. 1806), этот общий подход особенно полезен, когда исследуются планы с пропущенными ячейками (так называемые неполные планы).

Для полных факторных планов этот подход полезно также использовать в тех случаях, когда хотят анализировать взвешенные маргинальные средние. Например, предположим, что в рассматриваемом ранее простом 2 x 2 плане (стр. 1725), нужно сравнить взвешенные (по уровням фактора *B*) маргинальные средние для фактора A. Это бывает полезным, когда распределение наблюдений по ячейкам не готовилось экспериментатором, а строилось случайно, и эта случайность отражается в распределении числа наблюдений по уровням фактора B в совокупности.

Например, имеется фактор - возраст вдов. Возможная выборка респондентов разбита на две группы: моложе 40 лет и старше 40 (фактор B). Второй фактор (фактор A) в плане - получали или нет социальную поддержку вдовы в некотором агентстве (при этом одни вдовы были выбраны случайно, другие служили в качестве контроля). В этом случае распределение вдов по возрастам в выборке отражает действительное распределение вдов по возрастам в совокупности. Оценке эффективности группы социальной поддержки вдов по *всем возрастам* будет соответствовать взвешенное среднее для двух возрастных групп (с весами соответствующими числу наблюдений в группе).

Copyright © StatSoft, 1995

Спланированные сравнения

Общая процедура для введения коэффициентов контрастов описана в разделе *Примечания, часть I* на стр. 1784. Заметим, что сумма введенных коэффициентов контрастов не обязательно равна 0 (нулю). Вместо этого программа будет автоматически вносить поправки, чтобы соответствующие гипотезы не смешивались с общим средним.

Для иллюстрации этого вернемся опять к простому 2 x 2 плану, рассмотренному ранее. Напомним, что числа наблюдений в ячейках этого несбалансированного плана -1, 2, 3, и 1. Предположим, что мы хотим сравнить взвешенные маргинальные средние для фактора A (взвешенные с частотой уровней фактора B). Можно ввести коэффициенты контраста:

```
1  2 -3 -1
```

Заметим, что эти коэффициенты не дают в сумме 0. Программа будет устанавливать коэффициенты так, что в сумме они будут давать 0, и при этом будут сохраняться их относительные значения, т. е.:

```
1/3  2/3 -3/4 -1/4
```

Эти контрасты будут сравнивать взвешенные средние для фактора A.

Гипотезы о главном среднем. Гипотеза, о том, что не взвешенное главное среднее равно 0 может исследоваться с помощью коэффициентов:

```
1  1  1  1
```

Гипотеза о том, что взвешенное главное среднее равно 0 проверяется с помощью:

```
1  2  3  1
```

Ни в одном случае программа не производит корректировки коэффициентов контрастов.

Анализ планов с пропущенными ячейками (неполные планы)

Факторные планы, содержащие пустые ячейки (обработка комбинаций ячеек, в которых нет наблюдений) называются неполными. В таких планах некоторые факторы обычно не ортогональны и некоторые взаимодействия не могут быть вычислены. Вообще не существует лучшего метода анализа таких планов.

Регрессионный подход

В некоторых старых программах, которые основаны на анализе планов дисперсионного анализа с помощью многомерной регрессии, факторы в неполных планах по умолчанию задаются обычным образом (как будто план полный). Затем производится многомерный регрессионный анализ для этих фиктивно закодированных факторов. К несчастью, этот метод приводит к результатам, которые очень трудно, или даже невозможно, интерпретировать, так как неясно, как каждый эффект участвует в линейной комбинации средних значений. Рассмотрим следующий простой пример.

Фактор A	Фактор B: B1	Фактор B: B2
A1	3	4, 5
A2	6, 6, 7	Пропущено

Если будет выполняться многомерная регрессия вида *Зависимая переменная = Константа + Фактор A + Фактор B*, то гипотеза о значимости факторов A и B в терминах линейных комбинаций средних выглядит так:

```
Фактор A: Ячейка A1,B1 = Ячейка A2,B1
Фактор B: Ячейка A1,B1 = Ячейка A1,B2
```


StatSoft®

Copyright © StatSoft, 1995

Этот случай прост. В более сложных планах невозможно фактически определить, что точно будет исследоваться.

Средние ячеек, подход дисперсионного анализа, гипотезы типа IV

Подход, который рекомендуется в литературе и который кажется предпочтительнее - исследование осмысленных (с точки зрения исследовательских задач) *априорных* гипотез о средних, наблюдаемых в ячейках плана. Подробное обсуждение этого подхода можно найти в Dodge (1985), Heiberger (1989), Milliken and Johnson (1984), Searle (1987), или Woodward, Bonett, and Brecht (1990). Суммы квадратов, ассоциированные с гипотезами о линейной комбинации средних в неполных планах, исследующие оценки части эффектов, называются также суммами квадратов *IV*.

Автоматическая генерация гипотез типа IV. Когда многофакторные планы имеют сложный характер пропущенных ячеек, желательно определить ортогональные (независимые) гипотезы, исследование которых эквивалентно исследованию главных эффектов или взаимодействий. Были развиты алгоритмические (вычислительные) стратегии (основанные на псевдообратной матрице плана) для генерирования подходящих весов для таких сравнений. К сожалению, окончательные гипотезы определяются не единственным образом. Конечно, они зависят от порядка, в котором эффекты были определены и редко допускают простую интерпретацию. Поэтому рекомендуется внимательно изучить характер пропущенных ячеек, затем формулировать гипотезы *типа IV,* которые наиболее содержательно соответствуют целям исследования. Затем исследовать эти гипотезы, используя опцию *Спланированные сравнения* в окне *Результаты* (стр. 1763). Самый легкий путь задать сравнения в этом случае - требовать введения вектора контрастов для всех факторов *вместе* в окне *Спланированные сравнения.* После вызова диалогового окна *Спланированные сравнения* будут показаны все группы текущего плана и помечены те, которые пропущены.

Пропущенные ячейки и проверка специфического эффекта

Существует несколько типов планов, в которых расположение пропущенных ячеек не случайно, но тщательно спланировано, что позволяет проводить простой анализ главных эффектов не затрагивая другие эффекты. Например, когда необходимое число ячеек в плане недоступно, часто используются планы *Латинские квадраты* (см. стр. 1720 и 1799) для оценивания главных эффектов нескольких факторов с большим числом уровней. Например, 4 x 4 x 4 x 4 факторный план требует 256 ячеек. В то же время можно использовать *Греко-латинский квадрат* для оценки главных эффектов, имея только 16 ячеек в плане (глава *Планирование эксперимента*, том IV, содержит детальное описание таких планов). Неполные планы, в которых главные эффекты (и некоторые взаимодействия) могут быть оценены с помощью простых линейных комбинаций средних, называются *сбалансированными неполными планами*.

В сбалансированных планах стандартный (по умолчанию) метод генерирования контрастов (весов) для главных эффектов и взаимодействий будет затем производить анализ таблицы дисперсий, в которой суммы квадратов для соответствующих эффектов не смешиваются друг с другом. Опция *Специфический эффекты* окна *Результаты* (стр. 1763) будет генерировать пропущенные контрасты, записывая ноль в пропущенные ячейки плана. Сразу после того, как будет запрошена опция *Специфический эффекты* для пользователя, изучающего

Copyright © StatSoft, 1995

некоторую гипотезу, появляется таблица результатов с фактическими весами. Заметим, что в сбалансированном плане, суммы квадратов соответствующих эффектов вычисляются только, если эти эффекты ортогональны (независимы) всем другим главным эффектам и взаимодействиям. В противном случае нужно воспользоваться опцией *Спланированные сравнения* для изучения содержательных сравнений между средними.

Пропущенные ячейки и объединенные эффекты/члены ошибки

Если опция *Регрессионное подход* в стартовой панели модуля *Дисперсионный анализ* не выбрана (стр. 1757), то при вычислении суммы квадратов для эффектов будет использоваться модель средних по ячейкам (установка по умолчанию). Если план не сбалансирован, то при объединении неортогональных эффектов (см. выше обсуждение опции *Пропущенные ячейки и специфический эффект*) можно получить сумму квадратов, состоящую из неортогональных (или перекрывающихся) компонент. Полученные при этом результаты, обычно не интерпретируемы. Поэтому нужно быть очень осторожным при выборе и реализации сложных неполных экспериментальных планов.

Существует много книг с детальным обсуждением планов разного типа. (Dodge, 1985; Heiberger, 1989; Lindman, 1974; Milliken and Johnson, 1984; Searle, 1987; Woodward and Bonett, 1990), но такого рода информация лежит вне границ этого учебника. Тем не менее, позднее в этом разделе будет продемонстрирован анализ различного типа планов (в 12 представленных примерах из разных учебников будет обсуждено применение планов дисперсионного анализа, см. раздел *Примечания, часть II*, стр. 1806).

Предположения и эффекты нарушения предположений

Отклонение от предположения о нормальности распределений

Предположим, что зависимая переменная измерена в числовой шкале (см. раздел *Элементарные понятия* глава 8). Предположим также, что зависимая переменная имеет нормальное распределение внутри каждой группы. *Дисперсионный анализ* содержит широкий набор графиков и статистик для обоснования этого предположения.

Эффекты нарушения. Вообще F критерий очень устойчив к отклонению от нормальности (подробные результаты см. в работе Lindman, 1974). Если эксцесс (см. *Основные статистики и таблицы*, глава 9) больше 0, то значение статистики F может стать очень маленьким. Нулевая гипотеза при этом принимается, хотя она может быть и не верна. Ситуация меняется на противоположную, когда эксцесс меньше 0. Асимметрия распределения обычно незначительно влияет на F статистику. Если число наблюдений в ячейке достаточно большое, то отклонение от нормальности не имеет особого значения в силу *центральной предельной теоремы*, в соответствии с которой, распределение среднего значения близко к нормальному, независимо от начального распределения. Подробное обсуждение устойчивости F статистики можно найти в Box and Anderson (1955), или Lindman (1974).

Однородность дисперсии

Предположения. Предполагается, что дисперсии разных групп плана одинаковы. Это предположение называется предположением об

StatSoft®

Copyright © StatSoft, 1995

однородности дисперсии. Вспомним, что в начале этого раздела, описывая вычисление суммы квадратов ошибок, мы производили суммирование внутри каждой группы. Если дисперсии в двух группах отличаются друг от друга, то сложение их не очень естественно и не дает оценки общей внутригрупповой дисперсии (так как в этом случае общей дисперсии вообще не существует). Модуль *Дисперсионный анализ - ANOVA/MANOVA* содержит большой набор статистических критериев обнаружения отклонения от предположений однородности дисперсии.

Эффекты нарушения. Линдман (Lindman 1974, стр. 33) показывает, что *F* критерий вполне устойчив относительно нарушения предположений однородности дисперсии (*неоднородность* дисперсии, см. также Box, 1954a, 1954b; Hsu, 1938).

Специальный случай: коррелированность средних и дисперсий. Бывают случаи, когда *F* статистика может *вводить в заблуждение.* Это бывает, когда в ячейках плана средние значения коррелированы с дисперсией. Модуль *Дисперсионный анализ* позволяет строить диаграммы рассеяния дисперсии или стандартного отклонения относительно средних для обнаружения такой корреляции. Причина, по которой такая корреляция опасна, состоит в следующем. Представим себе, что имеется 8 ячеек в плане, 7 из которых имеют почти одинаковое среднее, а в одной ячейке среднее намного больше остальных. Тогда *F* критерий может обнаружить статистически значимый эффект. Но предположим, что в ячейке с большим средним значением и дисперсия значительно больше остальных, т.е. среднее значение и дисперсия в ячейках зависимы (чем больше среднее, тем больше дисперсия). В этом случае большое среднее значение ненадежно, так как оно может быть вызвано большой дисперсией данных. Однако *F* статистика, основанная на *объединенной* дисперсии внутри ячеек, будет фиксировать большое среднее, хотя критерии, основанные на дисперсии в каждой ячейке, не все различия в средних будут считать значимыми.

Такой характер данных (большое среднее и большая дисперсия) - часто встречается, когда имеются резко выделяющиеся наблюдения. Одно или два резко выделяющихся наблюдений сильно смещают среднее значение и очень увеличивают дисперсию.

Однородность дисперсии и ковариаций

Предположения. В многомерных планах, с многомерными зависимыми измерениями, также применяются предположение об однородности дисперсии, описанные ранее. Однако так как существуют многомерные зависимые переменные, то требуется так же чтобы их взаимные корреляции (ковариации) были однородны по всем ячейкам плана. Модуль *Дисперсионный анализ* предлагает разные способы проверки этих предположений.

Эффекты нарушения. Многомерный аналог *F*- критерия - λ-критерий Уилкса. Не так много известно об устойчивости (робастности) λ-критерия Уилкса относительно нарушения указанных выше предположений. Тем не менее, так как интерпретация результатов модуля *Дисперсионный анализ* основывается обычно на значимости одномерных эффектов (после установления значимости общего критерия), обсуждение робастности касается, в основном, одномерного дисперсионного анализа. Поэтому должна быть внимательно исследована значимость одномерных эффектов.

Специальный случай: ковариационный анализ. Особенно серьезные нарушения

Copyright © StatSoft, 1995

однородности дисперсии/ковариаций могут происходить, когда в план включаются ковариаты. В частности, если корреляция между ковариатами и зависимыми измерениями различна в разных ячейках плана, может последовать неверное истолкование результатов. Следует помнить, что в ковариационном анализе, в сущности, проводится регрессионный анализ внутри каждой ячейки для того, чтобы выделить ту часть дисперсии, которая соответствует ковариате. Предположение об однородности дисперсии/ковариации предполагает, что этот регрессионный анализ проводится при следующем ограничении: все регрессионные уравнения (наклоны) для всех ячеек одинаковы. Если это не предполагается, то могут появиться большие ошибки. Модуль *Дисперсионный анализ* имеет несколько специальных критериев для проверки этого предположения. Можно посоветовать использовать эти критерии, для того, чтобы убедиться, что регрессионные уравнения для различных ячеек примерно одинаковы.

Сферичность и сложная симметрия: причины использования многомерного подхода к повторным измерениям в дисперсионном анализе

В планах, содержащих факторы повторных измерений с более чем двумя уровнями, применение одномерного дисперсионного анализа требует дополнительных предположений: предположения о сложной симметрии и предположения о сферичности. Эти предположения редко выполняются (см. ниже). Поэтому в последние годы многомерный дисперсионный анализ завоевал популярность в таких планах (оба подхода совмещены в модуле *Дисперсионный анализ*).

Предположение о сложной симметрии

Предположение о сложной симметрии состоит в том, что дисперсии (общие внутригрупповые) и ковариации (по группам) для различных повторных измерений однородны (одинаковы). Это достаточное условие для того, чтобы одномерный F критерий для повторных измерений был обоснованным (т.е. выданные F-значения в среднем соответствовали F-распределению). Однако в данном случае это условие не является необходимым.

Предположение о сферичности.

Предположение о сферичности является необходимым и достаточным условием того, чтобы F-критерий был обоснованным. Оно состоит в том, что внутри групп все наблюдения независимы и одинаково распределены. Природа этих предположений, а также влияние их нарушений обычно не очень хорошо описаны в книгах по дисперсионному анализу - эта будет описано в следующих параграфах. Там же будет показано, что результаты одномерного подхода могут отличаться от результатов многомерного подхода, и будет объяснено, что это означает.

Необходимость независимости гипотез.

Общий способ анализа данных в дисперсионном анализе – это *подгонка модели*. Если относительно модели, соответствующей данным, имеются некоторые *априорные* гипотезы, то дисперсия разбивается для проверки этих гипотез (критерии главных эффектов, взаимодействий). С точки зрения вычислений, этот подход генерирует некоторое множество контрастов (множество сравнений средних в плане). Однако если контрасты не независимы друг от друга, разбиение дисперсий становится бессодержательным. Например, если два контраста *A* и *B* тождественны и выделяется соответствующая им часть из дисперсии, то одна и та же часть выделяется дважды. Например, глупо и бессмысленно выделять две гипотезы: “среднее в ячейке 1 выше среднего в ячейке 2” и “среднее в ячейке 1 выше среднего в ячейке 2”.

StatSoft®

Copyright © StatSoft, 1995

Итак, гипотезы должны быть независимы или ортогональны.

Независимые гипотезы при повторных измерениях. Общий алгоритм, реализованный в модуле *Дисперсионный анализ*, будет пытаться для каждого эффекта генерировать независимые (ортогональные) контрасты (см. раздел *Технические замечания* стр.1820). Для фактора повторных измерений эти контрасты задают множество гипотез относительно *разностей* между уровнями рассматриваемого фактора. Однако если эти разности коррелированы внутри групп, то результирующие контрасты не являются больше независимыми. Например, в обучении, где обучающиеся измеряются три раза за один семестр, может случиться, что изменения между 1 и 2 измерением отрицательно коррелируют с изменением между 2 и 3 измерениями субъектов. Те, кто большую часть материала освоил между 1 и 2 измерениями, осваивают меньшую часть в течение того времени, которое прошло между 2 и 3 измерением. В действительности, для большинства случаев, где дисперсионный анализ используются при повторных измерениях, можно предположить, что изменения по уровням коррелированы по субъектам. Однако когда это случается, предположение о сложной симметрии и предположения о сферичности не выполняются и независимые контрасты не могут быть вычислены.

Влияние нарушений и способы их исправления. Когда предположения о сложной симметрии или о сферичности не выполняются, дисперсионный анализ может выдать ошибочные результаты. До того, как были достаточно разработаны многомерные процедуры, было предложено несколько предположений для компенсации нарушений этих предположений. (см., например, работы Greenhouse & Geisser, 1959 и Huynh & Feldt, 1970). Эти методы до сих пор широко используются (поэтому они представлены в модуле *Дисперсионный анализ*).

Подход многомерного дисперсионного анализа к повторным измерениям. В целом проблемы сложной симметрии и сферичности относятся к тому факту, что множества контрастов, включенных в исследование эффектов факторов повторных измерений (с числом уровней большим, чем 2) не независимы друг от друга. Однако им не обязательно быть независимыми, если используется *многомерный* критерий для одновременной проверки статистического значимости двух или более контрастов фактора повторных измерений. Это является причиной того, что методы многомерного дисперсионного анализа стали чаще использоваться для проверки значимости факторов одномерных повторных измерений с более чем 2 уровнями. Этот подход широко распространен, так как он, в общем случае, не требует предположения о сложной симметрии и предположения о сферичности.

Случаи, в которых подход многомерного дисперсионного анализа не может быть использован. Существуют примеры (планы), когда подход многомерного дисперсионного анализа не может быть применен. Обычно это случаи, когда имеется небольшое количество субъектов в плане и много уровней в факторе повторных измерений. Тогда для проведения многомерного анализа может быть слишком мало наблюдений. Например, если имеется 12 субъектов, $p = 4$ фактора повторных измерений, и каждый фактор имеет $k = 3$ уровней. Тогда взаимодействие 4-х факторов будет "расходовать" $(k-1)^p = 2^4 = 16$ степеней свободы. Однако имеется лишь 12 субъектов, следовательно, в этом примере многомерный тест не может быть проведен. Модуль *Дисперсионный анализ* самостоятельно обнаружит эти наблюдения и вычислит только одномерные критерии.

Copyright © StatSoft, 1995

Различия в одномерных и многомерных результатах. Если исследование включает большое количество повторных измерений, могут возникнуть случаи, когда одномерный подход дисперсионного анализа к повторным измерениям дает результаты, сильно отличающиеся от тех, которые были получены при многомерном подходе. Это означает, что разности между уровнями соответствующих повторных измерений коррелированы по субъектам. Иногда этот факт представляет некоторый самостоятельный интерес.

Многомерный дисперсионный анализ и структурное моделирование уравнений.

В последние годы моделирование структурных уравнений стало популярным, как альтернатива многомерному анализу дисперсии (см. например, Bagozzi and Yi, 1989; Bagozzi, Yi, and Singh, 1991; Cole, Maxwell, Arvey, and Salas, 1993). Этот подход позволяет проверять гипотезы не только о средних в разных группах, но так же и о корреляционных матрицах зависимых переменных. Например, можно ослабить предположения об однородности дисперсии и ковариаций и явно включить в модель для каждой группы дисперсии и ковариации ошибки. Модуль *STATISTICA Моделирование структурными уравнениями (SEPATH)* (см. том III) позволяет проводить такой анализ.

StatSoft®

Copyright © StatSoft, 1995

ОБЗОР ПРОГРАММЫ

Модуль *Дисперсионный анализ* - это универсальная процедура для проведения одномерного и многомерного дисперсионного анализа. Программа будет обрабатывать планы с фиксированными и случайными эффектами, проводить межгрупповой анализ и анализ планов с повторными измерениями, с одним или несколькими зависимыми переменными. Также программа будет вычислять эффекты фиксированных (постоянных) или изменяющихся ковариат. С помощью этой программы можно проводить анализ контрастов любой сложности, определенных пользователем, *апостериорный* анализ и анализ дискриминантной функции.

По умолчанию, модуль *Дисперсионный анализ* будет использовать подход *Модель средних*, но пользователь может выбрать для вычислений *тип I* (последовательный, по умолчанию, либо с порядком, заданным пользователем), *тип II*, или *тип III* суммы квадратов. Тип *IV* гипотез может быть использован для неполных планов.

Для анализа можно использовать обширные графические возможности (такие как графики взаимодействий, внутригрупповые распределения переменных, внутригрупповые корреляции и т.п.)

Особенности модуля *Дисперсионный анализ*

Стандартные результаты. Модуль *Дисперсионный анализ* может анализировать многофакторные межгрупповые и внутригрупповые планы с одной (*ANOVA*) или несколькими (*MANOVA*) зависимыми переменными. Программа будет автоматически вычислять все эффекты (одновременно) и отображать результаты в интерактивной таблице результатов. По этой таблице, пользователь может исследовать маргинальные средние или графики всех эффектов и взаимодействий.

Фиксированные и случайные эффекты. Программа будет автоматически обрабатывать фиксированные и случайные эффекты (по требованию пользователя).

Неполные планы, заданные члены ошибки. Модуль *Дисперсионный анализ* будет полностью анализировать наиболее стандартные неполные планы (гнездовые (вложенные) планы, латинские квадраты, греко-латинские квадраты, планы с одним измерением на ячейку, случайные блочные планы). Также пользователь может задавать члены объединенной ошибки.

Проверка предположений Для проверки предположений модуля *Дисперсионный анализ* можно использовать большое разнообразие графиков: графики распределений, стволовые и листовые графики, графики средних со стандартными отклонениями или дисперсиями, нормальные и полунормальные вероятностные графики, полунормальные графики корреляций внутри ячеек и т.д. Также имеется большое количество статистических процедур для проверки предположений: критерий Кохрена, Хартли , критерии Бартлетта, критерий Левена, непараметрический критерий Бокса, критерий Сена и Пури, критерий Колмогорова-Смирнова, критерий сферичности Моучли и т.д.

Специализированный вывод. Программа будет вычислять поправки Гринхауса - Гейзера и Хюнха - Фельдта для факторов повторных измерений с более чем двумя уровнями. Так же автоматически будут получены одномерные и многомерные результаты для факторов повторных измерений. Пользователь может исследовать матрицы (*SS*) сумм квадратов

Copyright © StatSoft, 1995

(гипотеза, ошибка) и когда нужно, программа будет проводить полный канонический анализ: вычислять собственные вектора, характеристические значения, процент дисперсии, соответствующий каждому собственному вектору, стандартизованные и нестандартизованные дискриминантные функции. Заметим, что модуль *Дискриминантный анализ*, том III, предлагает детальный поэтапный анализ дискриминантной функции.

Ковариаты

Фиксированные и переменные ковариаты. Пользователь может ввести любое количество ковариат в план любого типа. В планах с повторными измерениями модуль *Дисперсионный анализ* обрабатывает одновременно фиксированные (постоянные) ковариаты, которые одинаковы на разных уровнях факторов повторных измерений и переменные ковариаты, которые измеряются на каждом уровне факторов повторных измерений.

Статистические критерии. Может быть отдельно проверена гипотеза о параллельности (взаимодействие ковариат с межгрупповыми факторами). Все результаты одномерной и многомерной регрессии будут посчитаны и полученные значения могут быть выведены на экран или представлены графически.

Анализ контрастов

***Априорные* контрасты.** Контрасты любой сложности могут быть точно определены для всех аспектов плана: для межгрупповых факторов, внутригрупповых факторов, факторов повторных измерений, для многомерных зависимых переменных. Удобное окно позволяет редактировать и вводить коэффициенты контрастов. Эти окна позволяют простым нажатием мыши вводить наборы полиномиальных контрастов, отклонения контрастов, разности контрастов, контрасты Хелмета, простые контрасты и контрасты для факторов повторных измерений. Множества контрастных коэффициентов так же могут быть введены для всех межгрупповых или внутригрупповых факторов одновременно. Это полезно для анализа несбалансированных планов с пустыми ячейками (неполные планы). В этом случае пустые ячейки будут автоматически выделены редактором контрастных коэффициентов.

***Апостериорные* критерии.** Доступные апостериорные критерии включают *НЗР - LSD* (наименьшей значимой разности) критерий, критерий Ньюмена - Кеулса (и критические области), критерий Шеффе, критерий *ДЗР - HSD* (достоверно значимой разности) Тьюки и критерии *ДЗР* Спйотвола и Столайна (обобщение критерия Тьюки для неравных объемов выборок). *Апостериорные* критерии могут быть применены для маргинальных средних главных эффектов или взаимодействий. Они могут применяться также для эффектов, включающих межгрупповые факторы, факторы повторения измерений или и те, и другие вместе.

Графики

Также как и во всех модулях системы *STATISTICA*, модуль *Дисперсионный анализ* содержит большое разнообразие специальных графиков и диаграмм. Для проверки предположений о распределениях имеются графики нормальной вероятности, гистограммы, стволовые и листовые диаграммы. Доступен набор категоризованных графиков (для большого количества уровней или ячеек текущего плана). Он включает категоризованные графики нормальной вероятности, гистограммы и диаграммы размаха, диаграммы рассеяния и другие. Для того чтобы проверить независимость средних по ячейкам от дисперсий, можно

StatSoft®

Copyright © StatSoft, 1995

строить диаграммы рассеяния средних со стандартными отклонениями, дисперсией. Для проверки однородности ковариации по ячейкам, может быть построен график нормальной вероятности z-преобразований внутригрупповых корреляций. Наконец, для помощи в интерпретации результатов могут быть быстро получены графики средних. Пользователь имеет полный контроль над упорядочиванием факторов и может включать только выбранные зависимые переменные и/или ковариаты.

Альтернативные процедуры

Для некоторых специальных случаев могут использоваться следующие альтернативные методы. В области промышленных исследований и исследований контроля качества были разработаны специализированные методы планирования (планирование поверхности отклика, латинские квадраты, отсеивающие эксперименты Плакетта - Бермана, планы 2^{k-p}, 3^{k-p} и Бокса - Бенкена, ортогональные блоки Тагучи, *смешанные* планы и т.д.). Такие планы могут быть генерируются и анализируются в модуле *Планирование эксперимента* (см. том IV). Если цель исследования – разделить переменные по группам, то можно использовать *Дискриминантный анализ* (см. том III, *Дискриминантный анализ*). Модуль *Непараметрическая статистика и распределения* (см. главу 11) содержит различные непараметрические методы для сравнения групп или переменных (при повторных измерениях). Если множество данных содержит цензуризованные наблюдения (время продолжительности жизни, момент отказа) можно использовать модуль *Анализ выживаемости* (см. том III) для сравнения групп. Для проверки сложных гипотез, имеющих отношение к различным линейным или нелинейным внутригрупповым регрессионным моделям, можно использовать модуль *Нелинейное оценивание.* Для проверки гипотез о структуре средних в различных группах можно использовать модуль *Моделирование структурными уравнениями (SEPATH)* (см. том III). Наконец, для того, чтобы написать свои собственные процедуры можно использовать язык программирования *STATISTICA BASIC,* который включает в себя полную библиотеку алгебры матричных функций (включая функции вычисления обратной и псевдообратной матрицы, ортогонализацию, разложение по сингулярным собственным значениям и т.д.). Пример программы на этом языке - *Manova.stb* (в поддиректории *STBASIC)*, показывает, как с помощью нескольких строк, используя доступные матричные функции, может быть осуществлен многомерный дисперсионный анализ (MANOVA).

Copyright © StatSoft, 1995

StatSoft®

ПРИМЕРЫ

Пример 1: План 2x3 одномерного дисперсионного анализа

Файл данных

Рассмотрим пример полного факторного плана 2x3. В этом примере используются данные из Lindman (1974).

Предположим, проводится эксперимент по проблеме “природа и обучение”. Наблюдается поведение различных крыс в Т-лабиринте. Т-лабиринт—простейший лабиринт и задача крысы это научиться бежать прямо к пище, расположенной все время в одном конце лабиринта. Были использованы три типа крыс, классифицированные по способностям проходить Т-лабиринт: *умные - bright*, *средние - mixed*, и *глупые - dull*. В каждом из этих типов выбирается 4 животных, которые содержатся в свободной (стимулирующей) среде и 4 животных, которые держатся в клетке. Исследуемая величина—количество ошибок, сделанных каждой крысой при прохождении Т-лабиринта.
Данные по этому исследованию содержатся в файле данных *Rats.sta*, часть этого файла показана выше.

Данные: RATS.STA 3п * ...

Performance of rats in a

	1 ENVIRNMT	2 STRAIN	3 ERRORS
1	FREE	BRIGHT	26,000
2	FREE	BRIGHT	14,000
3	FREE	BRIGHT	41,000
4	FREE	BRIGHT	16,000
5	FREE	MIXED	41,000
6	FREE	MIXED	82,000
7	FREE	MIXED	26,000
8	FREE	MIXED	86,000
9	FREE	DULL	36,000
10	FREE	DULL	87,000
11	FREE	DULL	39,000
12	FREE	DULL	99,000
13	RESTRCTD	BRIGHT	51,000
14	RESTRCTD	BRIGHT	35,000

После запуска модуля *Дисперсионный анализ* следует открыть файл данных *Rats.sta*.

Выбор анализа.

Когда открывается модуль *Дисперсионный анализ,* появляется стартовая панель *Дисперсионный анализ*, в которой можно ввести характеристики плана.

Зависимые и независимые переменные

Модуль *Дисперсионный анализ* классифицирует переменные как независимые, зависимые и ковариаты (см. главу *Элементарные понятия статистики* главу 8). Ковариаты будут использованы в следующем примере. Независимые переменные - переменные, находящиеся под контролем экспериментатора. К этим переменным относятся *группирующие* переменные, *кодирующие* переменные или *межгрупповые* факторы. Эти переменные содержат коды, которые будут использованы для однозначного определения, к какой группе эксперимента принадлежит данный случай.

В файле данных *Rats.sta* для независимой переменной *Среда - Envirnmt* были использованы коды *1-свобода - free* и *2-клетка – restricted* для того, чтобы указать, где содержалась данная крыса на свободе или же в клетке. Для второй независимой переменной *Strain* использовались коды: *1-умный - bright*, *2-средний - mixed*, и *3-глуппый - dull*. Зависимая переменная в эксперименте–эта та переменная, которая зависит от одной или нескольких независимых переменных. В этом исследовании - это переменная *Ошибки - Errors*, которая содержит количество попыток, при которых крыса не находит пищу.

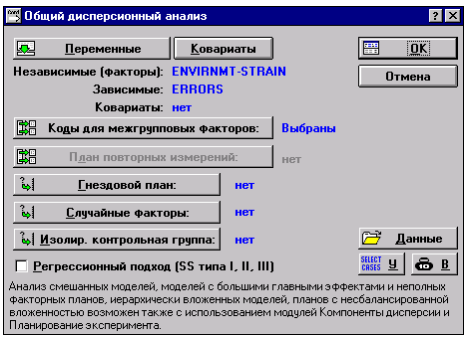
Выбор плана. Рассматриваемый план – это межгрупповой 2 (*Среда - Environment*) x 3 (*Tun - Strain*) план эксперимента. Переменные *Среда - Envirnmt* и *Tun - Strain* —независимые

StatSoft®

Copyright © StatSoft, 1995

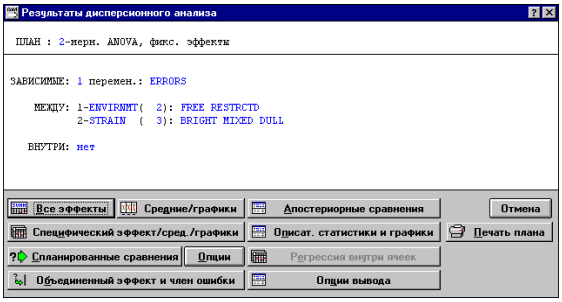
переменные, переменная *Ошибки - Errors* — зависимая переменная. Задать эти переменные можно с помощью кнопки *Переменные,* находящейся в стартовой панели. Затем с помощью кнопки *Коды* задаются коды, использующиеся для задания групп. Для введения всех кодов для каждой переменной, щелкните на кнопку *Все* для этой переменной. Для введения всех кодов для всех переменных нужно воспользоваться кнопкой *Выбрать все.*

Стартовая панель модуля *Дисперсионный анализ* выглядит так:

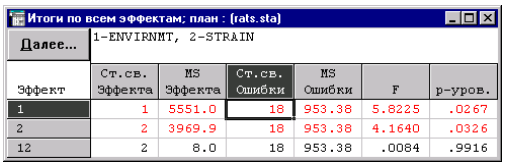


Анализ результатов

Для начала анализа нужно нажать на кнопку *OK* в стартовой панели. Откроется диалоговое окно *Результаты дисперсионного анализа.*



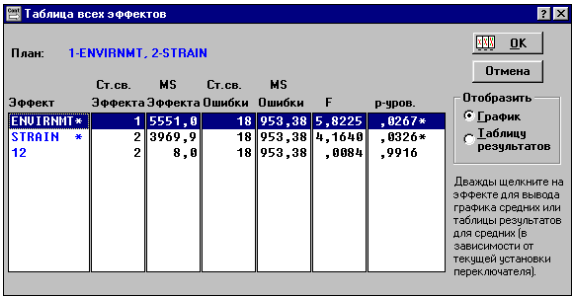
Этот диалоговое окно включает в себя описание плана и предлагает несколько опций для вывода. Нажмите кнопку *Все эффекты* для вывода таблицы результатов, содержащей все эффекты.



Эффект	Ст.св. Эффекта	MS Эффекта	Ст.св. Ошибки	MS Ошибки	F	p-уров.
1	1	5551.0	18	953.38	5.8225	.0267
2	2	3969.9	18	953.38	4.1640	.0326
12	2	8.0	18	953.38	.0084	.9916

Итоговая таблица дисперсионного анализа. Эта таблица объединяет главные результаты анализа. Можно заметить, что значимые эффекты (*p<.05*) в этой таблице выделены красным цветом. Пользователь может менять критерий значимости, нажав кнопку *Сервис* и введя затем нужный уровень α. В примере при уровне значимости α=0.05 главные эффекты факторов (*Среда -Envirnmt* и *Тип - Strain*) статистически значимы (*p<.05*), тогда как их взаимодействие незначимо (*p>.05*).

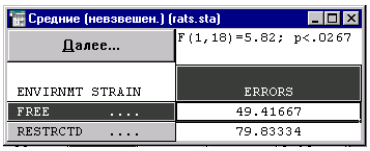
Просмотр маргинальных средних Теперь просмотрим маргинальные средние для главного эффекта фактора *Среда – Envirnmt* (заметим, что маргинальные средние считаются как невзвешенные средние). Нажмите на кнопку *Средние/графики* в диалоговом окне *Результаты дисперсионного анализа.* Откроется окно *Таблица всех эффектов.*



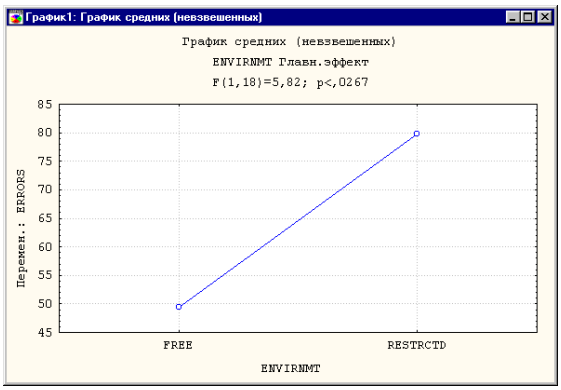
Эффект	Ст.св. Эффекта	MS Эффекта	Ст.св. Ошибки	MS Ошибки	F	p-уров.
ENVIRNMT *	1	5551,0	18	953,38	5,8225	,0267*
STRAIN *	2	3969,9	18	953,38	4,1640	,0326*
12	2	8,0	18	953,38	,0084	,9916

В этом окне выделите главный эффект *Среда - Envirnmt* и опцию *Таблицу результатов* в качестве параметра *Отобразить* (см. выше). После этого щелкните *OK* для создания таблицы результатов маргинальных средних для выделенного эффекта.

Copyright © StatSoft, 1995
StatSoft®



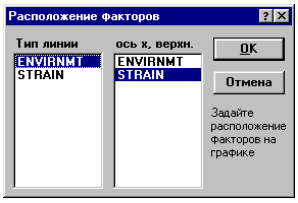
По умолчанию, для всех таблиц результатов с маргинальными средними, графиком является график средних. Для создания графика средних фактора *Среда* нужно вернуться в окно *Таблица всех эффектов* (нажать на кнопку *Далее* в таблице результатов) и выбрать значение *График* в качестве опции *Отобразить*, а затем снова нажмите *OK*.



Видно, что крысы, содержавшиеся в клетках, сделали большее количество ошибок, чем крысы содержавшиеся на свободе. Теперь можно посмотреть на график взаимодействия факторов *Среды - Environmt* и *Тип - Strain*.

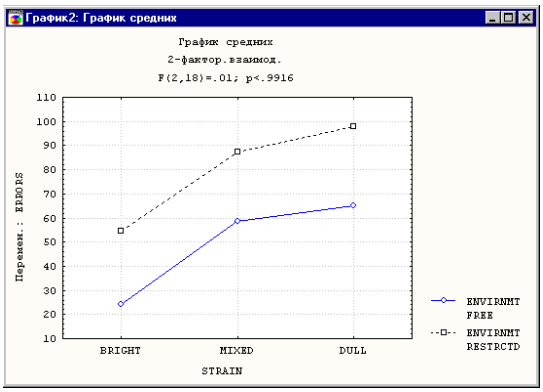
Просмотр графика взаимодействия

Вернемся еще раз к диалогу *Таблица всех эффектов* и на этот раз выделим эффект взаимодействия. Если теперь щелкнуть *OK*, появится следующее окно:



Заметим, что можно контролировать порядок, в котором выбираются взаимодействующие факторы. Для этого примера (выбрав факторы, как показано выше) получим график средних.

График, расположенный ниже, хорошо иллюстрирует результаты исследования, т.е. характер двух главных эффектов. Крысы, содержавшиеся в клетках (пунктирная линия), сделали большее количество ошибок, чем крысы, содержавшиеся на свободе (сплошная линия). Кроме того, *глупые - dull*s крысы сделали больше всего ошибок, далее (в порядке убывания ошибок) *средние - mixed* и *умные - bright* крысы.



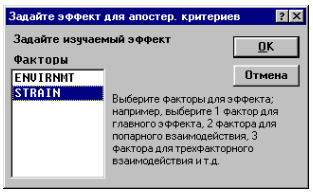
Апостериорные сравнения средних

Может возникнуть вопрос: значимо ли отличаются по количеству сделанных ошибок *средние - mixed* крысы от *глупых - dull* и *умных – bright*. Так как нет ни одной априорной гипотезы по этому вопросу, то нужно использовать *апостериорные* сравнения для проверки разности средних между типами крыс (см. *Вводный обзор*, стр. 1724, для объяснения основ апостериорных критериев).

Выбор *апостериорного* критерия. После возвращения в окно *Результаты дисперсионного анализа* нажмите кнопку *Апостериорные*

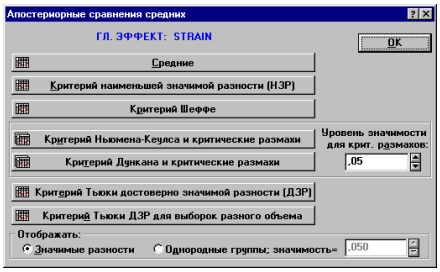
StatSoft®
Copyright © StatSoft, 1995

сравнения. Откроется окно, позволяющее выбирать эффекты для *апостериорных* критериев.



Для этого примера выберите фактор *Tun - Strain* для того, чтобы сравнить (невзвешенные) маргинальные средние этого фактора.

Выбор критерия. Теперь щелкните *OK* в этом окне, и откроется окно *Апостериорные сравнения средних*.

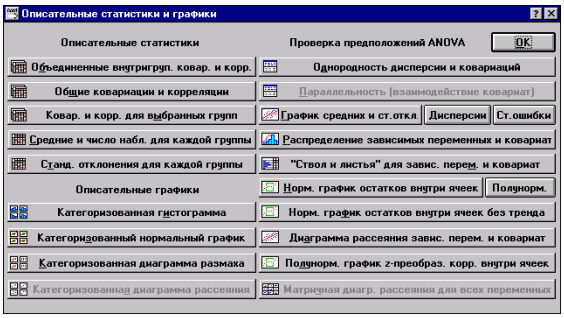


Различные *апостериорные* критерии в этом окне в некоторой степени “защищают” от случайных результатов (о природе *апостериорных* сравнений см. на стр.1724). Все эти критерии позволяют сравнивать средние при отсутствии априорной гипотезы относительно этих средних. Более подробно *апостериорные* критерии описаны на стр. 1769. Теперь нажмите кнопку *Критерий Шеффе*.

Эта таблица показывает статистическую значимость разностей между всеми парами средних. Из таблицы видно, что статистически значима только разность между группой 1 (умные крысы) и группой 3 (глупые крысы) на уровне *p<.05*. Из таблицы следует также, что, что глупые крысы сделали гораздо большее количество ошибок, чем умные, в то время как средние крысы сделали такое количество ошибок, которое не сильно отличается от количества ошибок, сделанных крысами двух остальных типов.

Проверка предположений

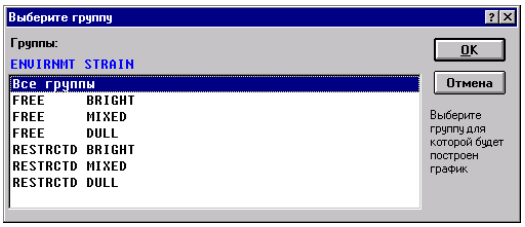
В части *Вводный обзор* обсуждались предположения, лежащие в основе метода дисперсионного анализа (см. стр.1730). Теперь посмотрим на данные с точки зрения этих предположений. Вернемся к диалогу *Результаты дисперсионного анализа.* Нажмите на кнопку *Описательные статистики и графики.* Диалоговое окно *Описательные статистики и графики* предлагает большое количество различных графиков и критериев. Некоторые из них применимы только к более сложным планам. Эти возможности описаны на стр. 1771.



Распределение зависимой переменной. В одномерном дисперсионном анализе предполагается, что распределение зависимой переменной (внутри групп) является нормальным. Для того чтобы посмотреть, как распределена зависимая переменная, нужно нажать на кнопку *Распределение зависимых переменных и ковариат.* Можно просматривать

Copyright © StatSoft, 1995
StatSoft®

общее распределение для всех групп вместе или же распределение для одной выбранной группы, выбирая группу в окне *Выберите группу*.



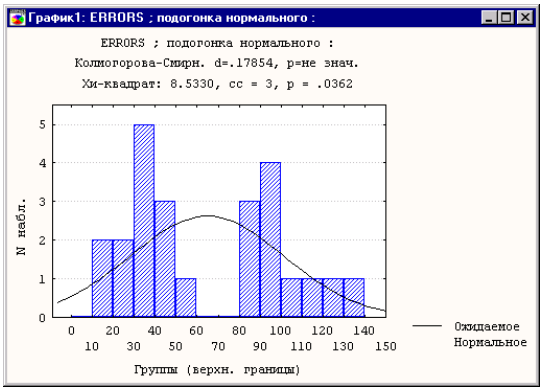
Для примера щелкните OK для выбора всех групп (по умолчанию выбрано значение *Все группы*). Появится таблица результатов и гистограмма частот.

ERRORS ; подогонка нормального : (rats.sta)

ДИСПЕРС. АНАЛИЗ / Верхняя Граница	Колмогорова-Смирн. d=.17854, p=не знач. Хи-квадрат: 8.5330, cc = 3, p = .0362							
	Наблюд. Частота	Кумулят. Наблюд.	Процент Наблюд.	Кумул. % Наблюд.	Ожидаем. Частоты	Кумулят. Ожидаем.	Процент Ожидаем.	Кумул. % Ожидаем.
10.	0	0	0.000	0.00	.7300	.730	3.042	3.042
20.	2	2	8.333	8.33	1.0968	1.827	4.570	7.612
30.	2	4	8.333	16.67	1.5296	3.356	6.373	13.985
40.	5	9	20.833	37.50	1.9799	5.336	8.250	22.235
50.	3	12	12.500	50.00	2.3788	7.715	9.912	32.146
60.	1	13	4.167	54.17	2.6527	10.368	11.053	43.199
70.	0	13	0.000	54.17	2.7458	13.114	11.441	54.640
80.	0	13	0.000	54.17	2.6379	15.752	10.991	65.631
90.	3	16	12.500	66.67	2.3523	18.104	9.801	75.433
100.	4	20	16.667	83.33	1.9470	20.051	8.112	83.545
110.	1	21	4.167	87.50	1.4958	21.547	6.232	89.778
120.	1	22	4.167	91.67	1.0666	22.613	4.444	94.222
130.	1	23	4.167	95.83	.7059	23.319	2.941	97.163
140.	1	24	4.167	100.00	.4336	23.753	1.807	98.970

Таблица показывает распределение частот для переменной *Error*. Очевидно, общее для всех групп распределение частот не соответствует нормальному распределению, хотя критерий хи-квадрат значим на уровне .05. Однако предположение о нормальности оправдано по отношению к *внутригрупповым* распределениям, а не к общему распределению, и с этой точки зрения, можно не беспокоится. Более того (как описано во введении—стр.1730), небольшие отклонения от нормальности не сильно влияют на *F* критерий.

Теперь посмотрим на гистограмму распределения.



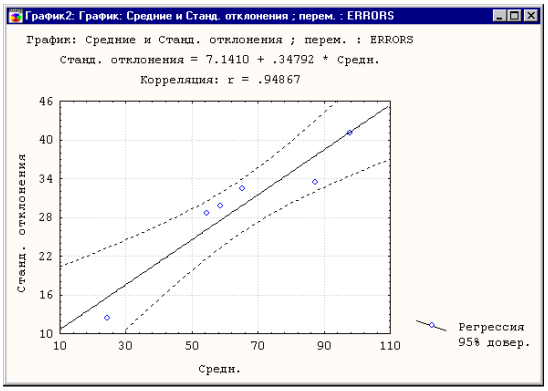
Видно, что общее распределение мультимодально, так как оно имеет более одного “пика”. Можно теперь предвидеть, что будут найдены значимые главные эффекты. Если нужно проверить предположение об однородности более основательно, можно посмотреть на распределения внутри каждой группы. Сейчас будет исследовано более серьезное нарушение предположений дисперсионного анализа.

Корреляция между средним и стандартным отклонением. Как уже отмечено во введении (см. стр.1730), отклонение от нормальности— не главный “враг”. Наиболее вероятной “ловушкой” в таких исследованиях - это обоснование каких-либо интерпретаций эффекта, основанного на “экстремальном” среднем в ячейке, сильно отличающемся от среднего в других ячейках. Другими словами, когда средние и стандартные отклонения коррелируют по ячейкам плана, тогда применение *F* критерия может приводить к ошибкам. Можно отклонить основную гипотезу при $p<.05$, когда реальное значение p может быть равно $.50$!

Теперь посмотрим, как связаны средние и стандартные отклонения по ячейкам в изучаемом плане. Можно выбрать график средних по одной оси, стандартные отклонения или дисперсии, или

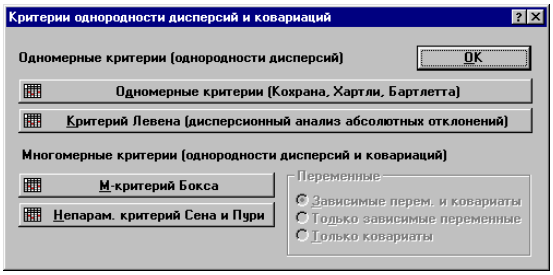
StatSoft®
Copyright © StatSoft, 1995

стандартные ошибки—по другой. Для этого нужно нажать на соответствующую кнопку (*График средних: со стандартным отклонением; с дисперсией; со стандартной ошибкой*, соответственно) в окне *Описательные статистики и графики*. Для рассматриваемого примера нажмите на кнопку *График средних и стандартных отклонений*.



Видно, что средние и стандартные отклонения заметно коррелируют. Если на результатах этого эксперимента должно основываться какое-либо важное решение, нужно перепроверить значимость главных эффектов, используя, например, какую-нибудь непараметрическую процедуру (см. главу 11, *Непараметрическая статистика и распределения*). Непараметрические процедуры, в основном, используют не числовые значения (и дисперсии), а ранги. В любом случае, к результатам о значимости эффектов нужно относится с осторожностью.

Однородность дисперсии. Теперь, рассмотрим критерий однородности дисперсии. Нажмите на кнопку *Однородность дисперсии/ковариаций* в окне *Описательные статистики и графики* и откроется диалоговое окно доступных процедур.



Можно использовать стандартный *одномерный* критерий однородности дисперсии или критерий *Левена*, но в данном примере ни один из них не получит статистически значимого результата.

Резюме

Кроме демонстрации главных функциональных аспектов модуля *Дисперсионный анализ* примера показал, как важно быстро представлять данные графически (например, строить диаграмму рассеяния средних и стандартных отклонений). Если доверять только F-критериям значимости и стандартным критериями однородности дисперсии, то не всегда можно обнаружить потенциально серьезное отклонение предположений, которое было обнаружено в диаграмме рассеивания средних вместе со стандартными отклонениями. В рассматриваемом примере, по-видимому, можно сделать вывод, что фактор условий среды и генетический фактор (*Strain*) оба вносят (аддитивный) вклад на поведение крыс в Т-лабиринте. Однако следует продолжить анализ данных с использованием непараметрических методов для того, чтобы гарантировать, что уровень значимости (*p)* статистических критериев дисперсионного анализа не слишком завышен.

Copyright © StatSoft, 1995
StatSoft®

Пример 2: План 2 x 3 x 3 (межгрупповой с повторными измерениями)

Обзор

Следующий пример служит демонстрацией того, как создавать план с более чем одним фактором повторных измерений. Будет показано, как контрастный анализ может помочь в интерпретации эффектов взаимодействия. Проблема *сферичности* будет проиллюстрирована следующим образом: многомерный критерий для взаимодействий не оказывается статистически значимым, в то время как обычный одномерный критерий указывает на статистическую значимость. Результаты, полученные методами Гринхауса - Гейзера и Хюнха - Фельдта, сравниваются с результатами, полученными при использовании многомерного критерия. Наконец, будет объяснена природа и причина возникновения отклонения от предположения о сферичности.

Проблема исследования

Обзор. Этот пример основан на (фиктивных) данных, напечатанных в Winer (1962, стр. 324). Предположим, нужно изучить, как отличаются влияния *смыслового -meaningful* и *белого – white* (*non-meaningful)* шумовых фонов на способность людей выполнять работу, требующую внимания. Операторы сложных машин (например, операторы АЭС) постоянно должны считывать (и обрабатывать) данные с различных измерительных приборов и производить настройку (наборных дисков) соответственно этим показаниям. Было бы интересно узнать, как на поведение людей в этом случае влияет белый шум (белый шум похож на звук радио, настроенного на диапазон без станций), и как влияет смысловой шум (если радио настроено на какую-либо станцию).

Будет “проведен” эксперимент, в котором субъекты должны настраивать один из трех различных наборных дисков до тех пор, пока соответствующий индикатор показывает значительное отклонение от нужной позиции диска. Некоторые субъекты будут проводить настройку, слушая белый шум, другие – смысловой шум (например, новости по радио). Зависимая переменная - количество ошибок, сделанное данным субъектом (не настройка в течение допустимого промежутка времени), в течение 3-х последовательных 10-минутных интервалов.

Резюме плана. Планом этого эксперимента является план одномерного дисперсионного анализ с тремя факторами: фактор *Шум – Noise* с 2 уровнями, фактор *Время – Time* с 3 уровнями и фактор *Dials, - Dial* с 3 уровнями. Последние два фактора - *внутригрупповые* факторы или факторы *повторных измерений,* так как представляют повторные измерения на одном субъекте. Первый фактор - это *межгрупповой* фактор, так как выбор субъектов для работы под тем или иным шумовым фоном является случайным.

Задание файла данных. Настройки для анализа плана с повторными измерениями очень просты. Межгрупповой фактор (*Noise)* выбирается так же, как и раньше, то есть, устанавливается переменная, содержащая коды, которые однозначно определяют, к какому значению (уровню) фактора принадлежит данный субъект. Все повторные измерения будут записаны как разные переменные.

Порядок расположения этих переменных не важен, так как модуль *Дисперсионный анализ* дает большую свободу в выборе уровней факторов повторных измерений. В рассматриваемом примере 6 субъектов (по 3 в каждой группе) и переменные расположены так, что для каждого последовательного периода

StatSoft®

Copyright © StatSoft, 1995

времени (настройки) измерения для каждого наборного диска записаны один за другим. Ниже приведен файл данных *Dials.sta*.

Данные: DIALS.STA 10п * 6н

Example data file for repeated measures ANOVA

	1 NOISE	2 TIM1_DL1	3 TIM1_DL2	4 TIM1_DL3	5 TIM2_DL1	7 TIM2_DL3	8 TIM3_DL1	9 TIM3_DL2	10 TIM3_DL3
1	MEANGFLL	45	53	60	40	57	28	37	46
2	MEANGFLL	35	41	50	30	47	25	32	41
3	MEANGFLL	60	65	75	58	70	40	47	50
4	WHITE	50	48	61	25	51	16	23	35
5	WHITE	42	45	55	30	43	22	27	37
6	WHITE	56	60	77	40	57	31	29	46

Выбор плана

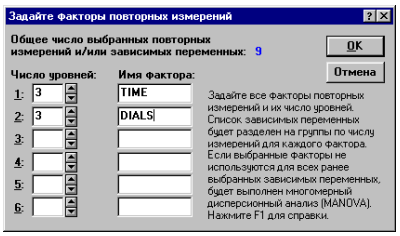
После запуска модуля *Дисперсионный анализ* откройте файл данных *Dials.sta*. Нажмите на кнопку *Переменные* стартовой панели модуля для выбора переменной (*Noise*) как независимой переменной и переменных 2-10 (*Tim1_Dl1* to *Tim3_Dl3*) как зависимых переменных. Затем щелкните на кнопке *Коды* и выберите коды (*1-Белый* и *2-Смысловой)* для независимой переменной.

Задание факторов повторных измерений. Факторы повторных измерений в модуле *Дисперсионный анализ* всегда задаются следующим способом. Во-первых, задается количество факторов повторных измерений; во-вторых, задаются имена этим факторам и их уровни. Затем программа разделит зависимую переменную по уровням факторов повторных измерений. Группы, на которые разбивается зависимая переменная, описываются уровнями факторов повторных измерений. См. также стр.1760 и 1790 для дополнительной информации о том, как задавать факторы планов повторных измерений в модуле *Дисперсионный анализ.*

В этом примере имеется два фактора повторных измерений: (1) - последовательные 10-минутные периоды *времени* и (2) - 3 *наборных диска.* Для задания этих факторов щелкните на кнопке *План повторных измерений* и появится окно *Задайте факторы повторных измерений.* В нем задаются факторы повторных измерений и их уровни.

Глядя на файл данных, можно заметить, что в обозначении переменных медленнее всего меняется тот индекс, который задает уровни для фактора *Time.* Для того чтобы понять, что такое термин “индекс”, посмотрим на таблицу из файла данных. Проведем следующую процедуру: начнем проводить пальцем по списку, начиная с переменной *Tim1_Dl1,* и будем произносить “один, два, три”. Каждый раз произнесенное слово “один” будет означать, что в данном наблюдении соответствующая переменная содержит данные для первого наборного диска(*Dl1*), слово “два”—для второго(*Dl2)*, “три”—для третьего(*Dl3*). Следовательно, быстрее всего изменяется индекс, относящийся к фактору *Dials*, и его нужно задавать последним. Первым должен быть задан фактор *Time.*

Когда вы щелкнете на кнопке *План повторных измерений,* в стартовой панели появится окно *Задайте факторы повторных измерений.* В этом окне нужно ввести имена факторов и количество уровней для плана с повторными измерениями. Для рассматриваемого примера нужно ввести 3 - число уровней первого фактора повторных измерений (*Time).* Число уровней для второго фактора повторных измерений *(Наборные диски)* также равно 3.



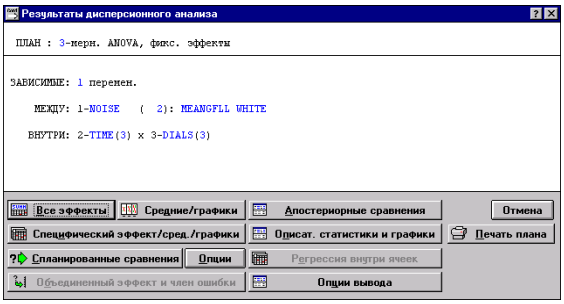
Теперь задана внутригрупповая (“внутрисубъектная”, с повторными измерениями) часть плана. Нажмите OK, чтобы вернуться к стартовой панели.

Copyright © StatSoft, 1995

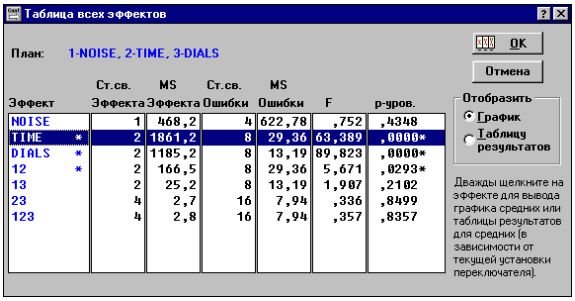
StatSoft

Анализ результатов

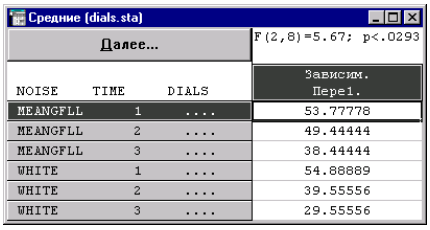
Теперь щелкните *OK* в стартовой панели для начала анализа. Через несколько секунд появится окно *Результаты дисперсионного анализа.*



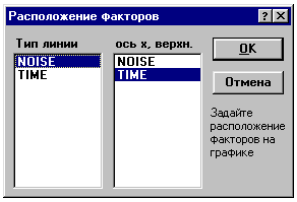
Вначале исследуем таблицу всех эффектов.



Нажмите на кнопку *Средние/графики* для просмотра таблицы. Из таблицы видно, что имеются 3 статистически значимых (на уровне *p*) эффекта (отмечены *): главный эффект фактора *Time*, главный эффект фактора *Dials* и фактора взаимодействия между *Noise* и *Time*. Теперь посмотрим на взаимодействие более подробно. Сначала выделим в этом диалоговом окне взаимодействие 12 (взаимодействие фактора 1 *(Noise)* и фактора 2 *(Time)*). Затем установим опцию *Таблицу результатов* в группе *Отобразить*. Нажмите *OK* для вывода таблицы результатов с маргинальными средними для эффекта взаимодействия.



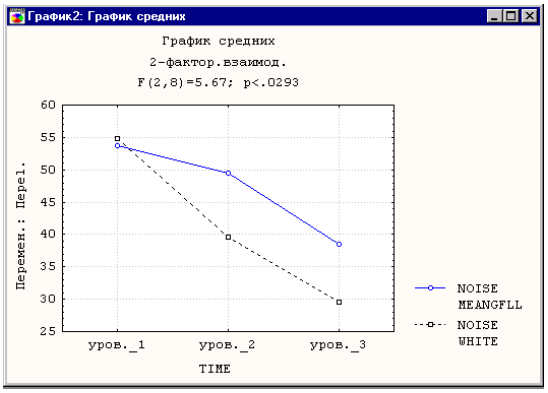
Самый простой (и иногда единственный) путь для того, чтобы понять характер взаимодействия - это построить график. Поэтому щелкните на кнопке *Далее* в этой таблице результатов для возвращения в диалоговое окно *Таблица всех эффектов* и поменяйте опцию *Отобразить* на *График*. Нажмите *OK,* и откроется окно *Расположение факторов*.



По желанию можно поменять порядок факторов на графике, для этого же примера просто щелкните *OK* для принятия установок по умолчанию.

Видно, что при разных видах шума испытуемые делают все меньше ошибок в течение последовательных 10-минутных временных промежутков. Однако два типа шума начинают по-разному действовать со второго временного промежутка (*level_2*). Видно, что субъекты сделали меньше ошибок при наличии *белого шума,* чем субъекты, работавшие при *смысловом* шумовом фоне во время 2 и 3 временного интервала. Во время первого временного интервала тип шума не повлиял на качество работы.

StatSoft®
Copyright © StatSoft, 1995

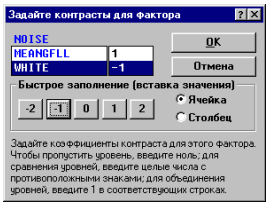


Главный эффект. Можно было бы построить график для главного эффекта фактора *Dials.* При этом можно было бы увидеть, что субъекты сделали больше ошибок, настраивая третий диск, чем при настройке второго, и при настройке второго больше, чем при настройке первого. Этот эффект отражает различия между различными наборными дисками, использованными в этом эксперименте. Впрочем, это не представляет теоретического интереса в нашем исследовании. Вернемся к эффекту взаимодействия.

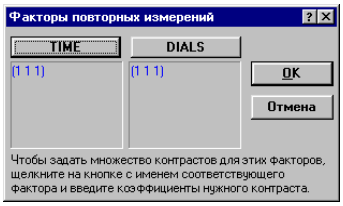
Исследование взаимодействия: анализ контрастов. Перейдем к дальнейшему исследованию взаимодействий помощью анализа контрастов. Возвратимся к диалоговому окну *Результаты дисперсионного анализа* и нажмем на кнопку *Спланированные сравнения.* Общие правила задания контрастов объяснены в деталях на стр. 1723, и в разделе *Примечания* (глава I), начиная со страницы 1785. Следует помнить, что контрасты задаются в виде набора положительных и отрицательных целых чисел. При этом сравниваются (сопоставляются) ячейки плана, которым присвоены целые числа различных знаков.

Задание контраста. Заметим, что совокупность контрастов также может быть задана одной длинной строкой, состоящей из контрастных коэффициентов (вместо отдельного ввода контрастов для каждого фактора, см. ниже). Для этого нужно щелкнуть на кнопке *Опции* (рядом с кнопкой *Спланированные сравнения*) в диалоговом окне *Результаты дисперсионного анализа* и установить опцию *Вводить контрасты для всех факторов вместе (в виде вектора контрастов).*

Для этого примера будет исследована разность между двумя уровнями фактора *Noise* (отдельно для временных промежутков 2 и 3). Нажмите на кнопку *Спланированные сравнения,* и откроется диалоговое окно *Задайте контрасты для этого фактора.* В этом окне сначала нужно задать контраст для фактора *Noise.* Введем 1 и –1 соответственно (для сравнения первого и второго уровня фактора *Noise*). Заметим, что вместо ввода контрастных коэффициентов, можно использовать метод *Вставка значений*, который позволяет вводить коэффициенты (для каждой ячейки или для целых столбцов ячеек), нажимая соответствующие числовые кнопки.

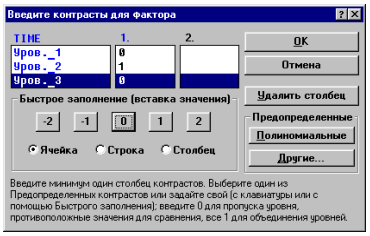


Нажмем на *OK,* и появится следующее окно. Оно содержит список всех факторов повторных измерений.

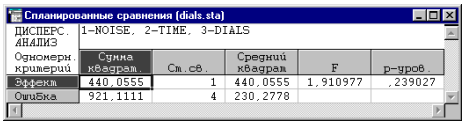


Для того чтобы ввести контрасты для фактора *Time* щелкните на кнопке *Time.*

Copyright © StatSoft, 1995
StatSoft®

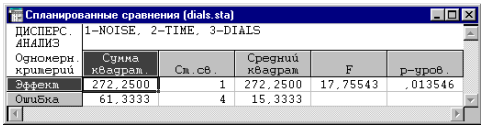


Затем в окне *Введите контрасты для этого фактора* введем коэффициенты 0,1,0 соответственно. (Напомним, что можно использовать опцию *Вставка значения*). Нажмем кнопку *OK* для возвращения в диалог *Факторы повторных измерений* и аналогично, щелкнув на кнопке *Dials,* вводим *1*, *1*, и *1* – контрастные коэффициенты для этого фактора. Этим выбором коэффициентов все уровни фактора *Dials* объединяются. Так как коэффициенты (1,1,1) заданы по умолчанию, можно просто щелкнуть *OK* в диалоге *Ввести контрасты для этого фактора*. Затем щелкните *OK* в диалоге *Повторные измерения* для просмотра таблицы результатов.



Видно, что разность между двумя средними фактора *Noise* - (*Белый, Смысловой*) для временного промежутка 2 статистически не значима. Для сравнения этих же двух средних на 3 временном промежутке, для фактора *Time* нужно ввести коэффициенты *0, 0, 1,* а все остальные введенные параметры оставить теми же самыми. Соответствующая разность также статистически не значима.

Задание контраста частичного взаимодействия. Итак, нельзя сказать, что разность ошибок для разных типов шума во 2 и 3 промежутке времени статистически значима. Вернемся к графику взаимодействия этих двух факторов. Из этого графика видно, что изменение количества ошибок от времени 1 к времени 2 — это главное, из-за чего взаимодействие между этими двумя факторами значимо. Исследуем значимость частичного взаимодействия между факторами *Noise* и *Time* (игнорируя третий уровень фактора *Time).* Для того чтобы сделать это, в очередной раз нужно нажать кнопку *Спланированные сравнения* и задать контрастные коэффициенты для переменной *Noise* равными *1* и *-1* в диалоговом окне *Задать контрасты для этого фактора.* Затем нужно ввести коэффициенты *1*, *-1*. *0*, соответственно для фактора *Time* и коэффициенты *1*, *1*, *1* для фактора *Dials* в окне *Факторы повторных измерений* (см. выше). Можно заметить, что сопоставляются первые два уровня фактора *Noise* и первые два уровня фактора *Time*. Теперь щелкните *OK* для просмотра таблицы результатов.



Этот контраст статистически значим. При наличии статистической значимости взаимодействия между факторами *Noise* и *Time* видно, что основной вклад во взаимодействие вносят различия в количестве ошибок между 1 и 2 временными интервалами для разных типов шумов. Эта разность больше для тех, кто работал при *белом* шуме и меньше для работавших при *смысловом* шуме.

Многомерный подход

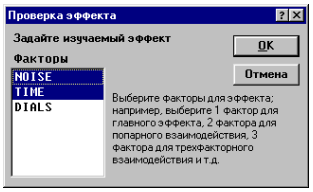
Во введении обсуждались специальные предположения одномерного дисперсионного анализа при повторных измерениях (см. стр. 1730). Во многих научных дисциплинах многомерный дисперсионный анализ становится единственно приемлемым способом анализа планов с повторными измерениями. Это связано с тем, что многомерный подход не ограничен

StatSoft®
Copyright © StatSoft, 1995

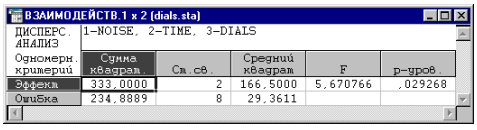
предположениями сферичности или сложной симметрии.

Короче говоря, одномерный дисперсионный анализ при повторных измерениях предполагает, что изменения по уровням факторов повторных измерений некоррелированы. В большинстве случаев это очень сильное предположение. В рассматриваемом примере совершенно ясно, что люди, участвующие в эксперименте, во втором временном интервале почти достигли предела своих возможностей в точности выполнения задания. Уменьшение числа ошибок между 2 и 3 временными интервалами меньше, чем между 1 и 2. Это наводит на мысль, что предположение сферичности одномерного дисперсионного анализа в данном примере не выполняется. Посмотрим на многомерные статистики.

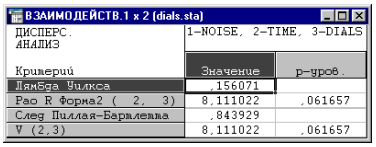
Выбор эффекта взаимодействия. Из диалогового окна *Результаты* щелкните на кнопке *Специфический эффект/Средние/Графики* для того, чтобы открыть диалог *Проверка эффекта,* с помощью которого можно исследовать отдельный эффект в деталях. В этом окне выделим факторы *Noise* и *Time* для того, чтобы посмотреть на эффект взаимодействия.



Сначала одномерные результаты будут отображены в таблице.



Как и раньше, одномерный *F* критерий значим (это, конечно, совпадает с результатом, запрошенным через кнопку *Все эффекты.* Теперь, посмотрим на многомерные результаты.



Существует большое количество многомерных критериев. В данном случае, λ - критерий Уилкса, критерий Рао, след Пилая - Бартлетта и *V*-критерий дают один и тот же результат: взаимодействие не значимо на уровне *.05*. Следовательно, здесь мы имеем случай, когда нарушение предположения о сферичности привело к неправильной интерпретации эффекта взаимодействия (получилось, что взаимодействие статистически значимо).

Технические замечание: определение причины нарушения. Для тех, кто намерен пока отложить рассмотрение вопроса о сферичности, можно рекомендовать пропустить этот параграф и перейти к чтению следующего. Заметим, что, не вдаваясь в детали, можно определить, где нарушается предположение о сферичности.

Для этого, запросите внутригрупповую ковариационную матрицу разностей (в диалоговом окне *Результаты дисперсионного анализа* нажмите кнопку *Опции вывода* и в выданном диалоге нужно отметить опцию *Вывод матриц сумм квадратов*). Затем нужно нажать на кнопку *Спланированные сравнения* и потом выбрать два независимых контраста для фактора *Time,* например, (1) *-2*, *1*, *1* и (2) *0*, *1*, *-1*. Все другие контрастные коэффициенты положим равными *1*.

Опишем сейчас две независимые гипотезы, определяемые заданными контрастами, для фактора *Time.* Гипотеза, соответствующая первому контрасту, сравнивает среднее в 1

Copyright © StatSoft, 1995
StatSoft®

временном интервале с суммой средних 2 и 3 временного интервала. Гипотеза, соответствующая второму контрасту, сравнивает средние 2 и 3 временного интервала. Ковариационная матрица ошибок *SSCP: Ошибка* содержит дисперсии и ковариации контрастов. Она выглядит так::

SSCP:Ошибка (dials.sta)

ДИСПЕРС. АНАЛИЗ	1	2
1	146,5556	-81,5988
2	-81,5988	88,3333

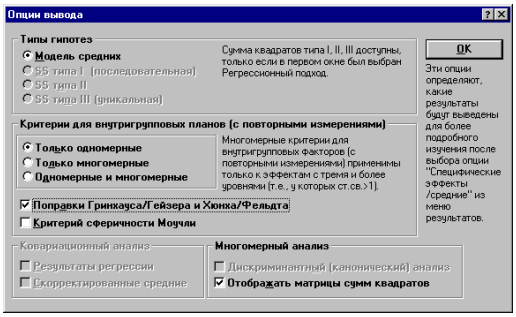
Внедиагональный элемент этой матрицы равен -*81.6*. Для вычисления корреляции между контрастами нужно разделить эту величину на квадратный корень из произведения диагональных элементов, т.е.

- `81.6/√(146.6*88.33) = -.72.`

Итак, получено, что коэффициент корреляции между контрастными коэффициентами равен -.72. Применение *F* критерия обоснованно, если выполнены предположения сферичности. При этом предположении коэффициент корреляции должен быть близок к нулю. Отрицательность коэффициента корреляции между контрастами означает, что если разность между средними в первом временном интервале и суммой средних в интервалах 2 и 3 велика (значение контраста (1)), то разность в средних 2 и 3 временного интервала мала (значение контраста (2)). Это еще раз подтверждает первоначальное предположение, что те, кто вначале резко улучшает качество работы, в остальные промежутки времени работают примерно одинаково.

Скорректированные одномерные критерии. Перед тем, как многомерный подход получил распространение, некоторыми авторами были предложены поправки одномерного *F* критерия (на случай нарушения предположений *F* – критерия): поправка Гринхауса-Гейзера (Greenhouse и Geisser, 1958, 1959) и поправка Хюнха-Фельдта (Huynh and Feldt, 1976).

Заметим, что это приближенные поправки и, если возможно, нужно использовать многомерный подход. Однако в некоторых научных дисциплинах эти поправки по-прежнему обычно используются. Для вычисления этих поправок вернемся в окно *Результаты дисперсионного анализа* и нажмем на кнопку *Опции вывода.*



Вызов поправок Гринхауса - Гейзера и Хюнха - Фельдта. В представленном выше окне содержатся некоторые специализированные опции вывода. Некоторые из этих опций не применимы в рассматриваемом примере и поэтому недоступны. Выберем опцию *Поправки Гринхауса - Гейзера и Хюнха - Фельдта* и затем нажмем *OK.* Затем в диалоговом окне *Результаты дисперсионного анализа* нужно нажать на кнопку *Специфический эффект/средние/графики.* В открывшемся окне нужно выделить факторы *Шум – Noise*, *Время - Time* и нажать кнопку *OK.* Сначала появится таблица коэффициентов *эпсилон*.

Эпсилон Гринхауса/Гейз. и Хюнха/Фельдта (dials.sta)

ДИСПЕРС. АНАЛИЗ	ВЗАИМОДЕЙСТВ. 1 x 2
	Эпсилон
Эпсилон Гринхауса-Гейзера	,647598
Эпсилон Хюнха-Фельдта	1,000000
Эпсилон нижней границы	,500000

Анализ результатов. В представленной таблице содержатся поправки для одномерных критериев. В частности, имеются поправочные множители для степеней свободы. Соответствующие значения *эпсилон* умножаются

StatSoft®
Copyright © StatSoft, 1995

на степени свободы, соответствующие средним квадратам эффектов и ошибок. Имеется также значение обычного одномерного *F* критерия с числом степеней свободы 2 для MS_{effect} и 8 степеней свободы для MS_{error}. Таблица с поправками для *F* критерия представлена ниже.

Видно, что поправка Гринхауса - Гейзера хорошо работает, защищая от ошибочного признания эффекта взаимодействия статистически значимым на уровне .05. На самом деле, полученный здесь уровень (.063) очень близок к уровню многомерного критерия (*p* =.062). Поправка Хюнха - Фельдта не очень помогает в нашем случае.

Одномерный критерий со скоррект. степ. свободы (dials.sta)

ДИСПЕРС. АНАЛИЗ Одномерн. критерий: F = 5,670766, ВЗАИМОДЕЙСТВ.1 x 2

	Нескорр.	Гринхс.-Гейзера	Хюнха-Фельдта	Нижняя Граница
Эпсилон		,647598	1,000000	,500000
См.св. 1	2,000000	1,295195	2,000000	1,000000
См.св. 2	8,000000	5,180781	8,000000	4,000000
р-уров.	,029268	,056940	,029268	,075879

Резюме

Анализируя этот пример, находим, что взаимодействие между факторами *Noise* и *Time* проявляется только между первыми двумя 10-минутными временными интервалами. Испытуемые, работающие под действием белого шума, быстрее улучшают свою квалификацию (делают меньше ошибок), чем те, кто работает при смысловом шуме. Однако разность в числе ошибок (межгрупповая) в любом временном интервале статистически незначима. При дальнейшем анализе были обнаружены серьезные отклонения от предположений сферичности, что поставило под сомнения первоначальные выводы. Многомерный критерий проверки значимости эффекта взаимодействия между факторами *Noise* и *Time,* а также одномерный критерий с поправкой Гринхауса - Гейзера, признали эффект взаимодействия незначимым на уровне .05. На этом примере видно преимущество многомерного подхода по сравнению с одномерным дисперсионным анализом для планов с повторными измерениями с более чем двумя уровнями (у факторов повторных измерений).

Пример 3: Многомерный ковариационный анализ

Обзор

В этом примере будет определен план многомерного дисперсионного анализа с несколькими (фиксированными) ковариатами. Будет исследована гипотеза о параллельности и проведен стандартный многомерный дисперсионный анализ.

Файл данных

Этот пример содержит данные, сообщенные в книге Finn (1974). Четыре группы, по 12 человек в каждой, должны были отсортировать 50 слов (каждое слово написано на отдельной карточке) по категориям.

Экспериментальные группы различались инструкциями, которые ими были получены относительно числа категорий представленных слов и числом “подкатегорий”. В частности, каждая из групп получила следующие инструкции:

(1) первой группе было предложено сортировать слова на 5 категорий по 2 подкатегории в каждой (условие *M5_S2*);

(2) второй группе было предложено сортировать слова просто на 10 категорий без всяких подкатегорий (условие *M10_S0*);

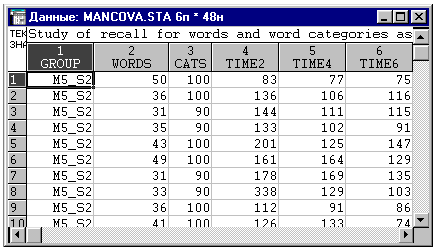
(3) третья группа должна была сортировать слова на 5 категорий без подкатегорий (условие *M5_S0*);

Copyright © StatSoft, 1995

(4) третья группа должна была рассортировать слова, как и группа 3 на 5 категорий, но список слов не содержал никаких заметных или специально подобранных экспериментатором структур (контрольная группа; *M5_S0_C*).

Основными зависимыми переменными были: (1)количество слов, которые запомнили испытуемые после 6 попыток сортировки и (2) процент категорий, воссозданных субъектами при сортировке, относительно числа категорий, предусмотренных экспериментатором. В качестве ковариат в план входит количество времени (в секундах), затраченное субъектом при 2ой, 4ой, и 6ой сортировках.

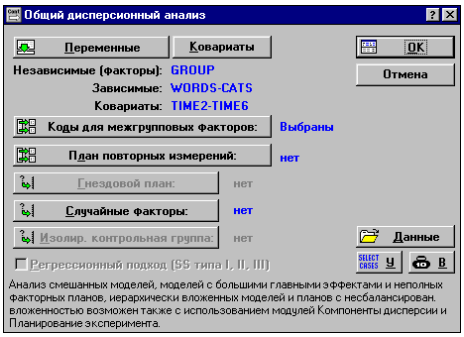
Эти данные содержатся в файле *Mancova.sta*. Часть этого файла представлена ниже.



Переменная 1 (*Группа- Group)* содержит коды, определяющие группу, к которой принадлежит субъект. Переменная 2 *(Слова - Words),* содержит число запомненных слов; переменная 3 (*Кат - Cats*) содержит процент категорий, воспроизведенных субъектами; переменные от 4 до 6 (*Time2* to *Time6*) содержат количество времени, затраченное субъектами на 2ую, 4ую и 6ую сортировку соответственно.

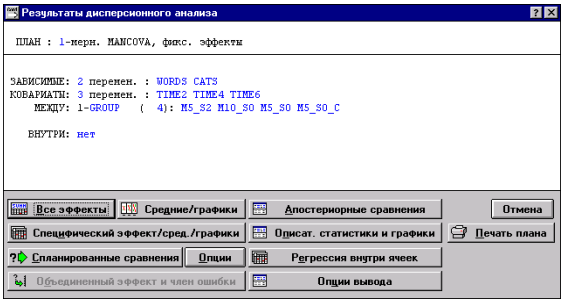
Задание плана

После открытия модуля *Дисперсионный анализ*, нужно открыть файл данных *Mancova.sta*. В стартовой панели нажмите на кнопку *Ковариаты* и задайте переменную 1 (*Группа - Group*) как независимую переменную, переменные 2 и 3 (*Слова - Words* и *Кат - Cats*) как зависимые переменные, и переменные 4, 5 и 6 (*Время2 - Time2*, *Время4 - Time4*, *Время6 - Time6*) как ковариаты. Затем выберите коды для независимой переменной (нажмите на кнопку *Коды для межгруппового фактора,* и в появившемся окне нажмите на кнопку *Все;* таким образом, будут заданы все четыре кода, задающие группы: *M5_S2*, *M10_S0*, *M5_S0*, and *M5_S0_C*). Теперь план полностью определен.



Анализ результатов

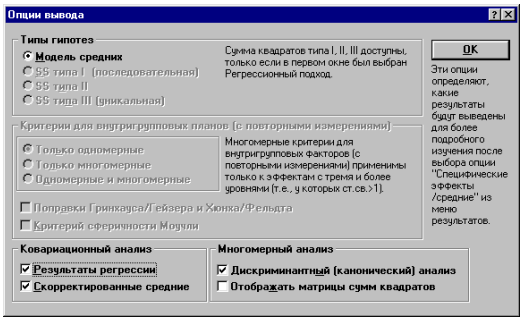
После нажатия в стартовой панели кнопки *OK* появляется окно *Результаты дисперсионного анализа*.



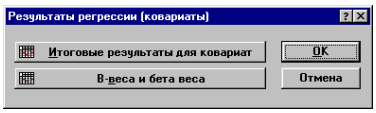
В этом плане имеется только один фактор (еще есть ковариаты) и нужно исследовать вклад этого фактора подробно перед использованием опции *Все эффекты*.

StatSoft®
Copyright © StatSoft, 1995

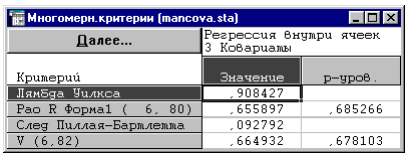
Сначала откроем окно *Опции вывода* (нажав на кнопку *Опции вывода*) и выберем *Результаты регрессии* (для проведения вычислений регрессионного анализа для ковариат), *Дискриминантный (канонический) анализ* (для проведения вычислений дискриминантного анализа) и *Скорректированные средние* (для вычисления скорректированных средних).



Теперь нажмем на кнопку *OK* и вернемся в окно *Результаты дисперсионного анализа.* Нажмем на кнопку *Специфический эффект/средние/графики,* и откроется диалоговое окно *Результаты регрессии (ковариаты).*



Результаты регрессии (ковариаты). Так как имеется несколько зависимых переменных, можно исследовать, значимы или нет канонические корреляции (внутри групп) между всеми зависимыми переменными и всеми ковариатами. Нажмите на копку *Итоговые результаты для ковариат*.

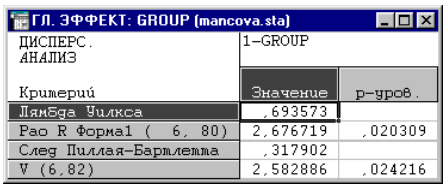


Из этой таблицы можно заключить, что ковариаты не связаны значимо с зависимыми переменными. В этот момент нужно было бы вернуться в диалоговое окно *Результаты регрессии* (нажать на кнопку *Далее* на таблице) и выбрать опцию *B-веса и бета - веса* для просмотра множества коэффициентов корреляции для каждой зависимой переменной. Но так как общий критерий незначим, то этот шаг можно пропустить.

Одномерные критерии. После возвращения в окно *Результаты регрессии,* нажмите на кнопку *OK* для вывода на экран таблицы одномерных *F* критериев для каждой зависимой переменной.

Ни один из одномерных критериев не дает статистической значимости главного эффекта фактора *Группа – Group.* Результат для зависимой переменной *Words* числа отсортированных слов, близок к статистической значимости (на уровне *.10)*.

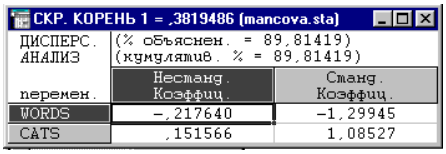
Многомерные критерии. Многомерные критерии для нашего примера значимы. Два общепринятых критерия приведенных здесь *R* – критерий Рао и *V* –критерий показывают статистическую значимость главного эффекта на уровне значимости *p<.03.*



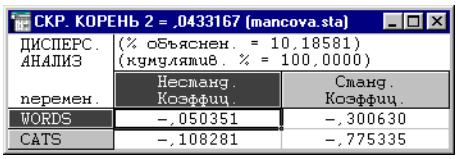
Канонический анализ. Теперь рассмотрим таблицу результатов анализа дискриминантной функции. Модуль *Дисперсионный анализ* будет циклически обрабатывать все канонические корни. Для каждого корня на экране появятся

Copyright © StatSoft, 1995
StatSoft®

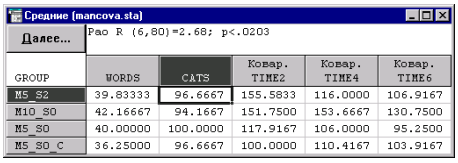
стандартизованные и нестандартизованные параметры дискриминантной функции.



На первый корень приходится почти 90% общей дисперсии. Общая дисперсия - это та дисперсия, которая может быть объяснена межгрупповыми факторами. Нужно помнить, что первый канонический корень извлекает большую часть дисперсии, за счет которой становятся значимыми главные многомерные эффекты фактора *Группа - Group*. Из таблицы видно, что знак коэффициента для переменной *Words* отрицательный, а для переменной *Cats* положительный. Это означает, что эти переменные входят с противоположным знаком в линейную функцию, максимизирующую дисперсию, объясняющую главный эффект. Эта информация будет использована позднее для интерпретации графиков канонических средних (см. ниже). Теперь посмотрим на таблицу параметров для второго корня. На второй корень приходится только 10% общей дисперсии.

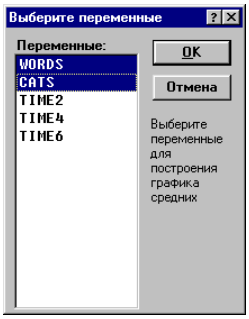


Анализ средних. Перейдем теперь к таблице средних.

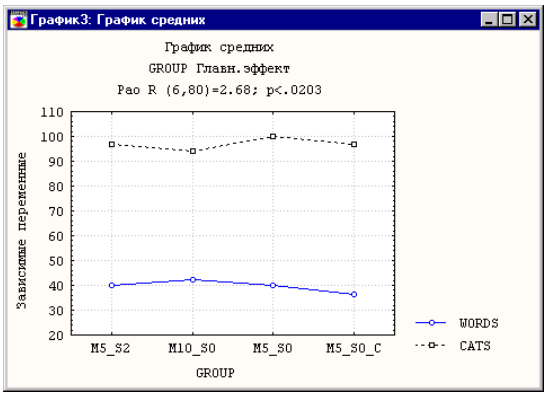


Как в любой таблице маргинальных средних в модуле *Дисперсионный анализ* можно построить график одной или нескольких переменных, нажав правую кнопку мыши и выбрав опцию *Быстрые статистические графики.* В появившемся меню нужно выбрать опцию *График средних*.

Построение графиков средних. После выбора опции *Графики средних* из выпадающего меню можно затем в диалоговом окне *Выбор переменных для графика* выбрать как зависимые переменные, так и ковариаты.



Для нашего примера выберем только переменные *Слова - Word* и *Cats*, выделяя их в этом окне. Теперь, для построения графика, нажмем кнопку *OK*.



Полученный график в некоторой степени неудачен, так как переменные *Cats* и *Words* были измерены в разных шкалах. Тем не менее, видно, что в группе 2 (*M10_S0*) вспомнили наибольшее число слов, но получили наименьшее число

StatSoft®
Copyright © StatSoft, 1995

категорий; характеристики других групп сразу не заметны. Вернемся теперь к таблице средних и нажмем на кнопку *Далее.* Появится таблица скорректированных средних. Эта таблица содержит средние для зависимых переменных, подправленные ковариатами. Поступая как в предыдущем случае, можно построить график этих средних (нажимая на таблице на правую кнопку мыши и т.д.). График будет очень похож на график нескорректированных средних.

Анализ результатов для дискриминантных функций. Теперь, нажав кнопку *Далее* в таблице скорректированных средних получим таблицу средних для дискриминантных функций.

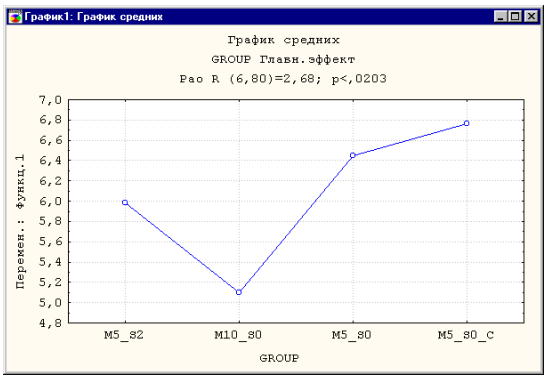
Средние для дискрим. функций (mancova.sta)

Далее... | Pao R (6,80)=2,68; p<,0203

GROUP	Дискрим. функц. 1	Дискрим. функц. 2
M5_S2	5,982027	-12,4729
M10_S0	5,095286	-12,3197
M5_S0	6,450972	-12,8422
M5_S0_C	6,761903	-12,2924

Напомним, что на первую функцию приходится наибольшая часть общей (для переменных *Words* и *Cats)* дисперсии, отвечающей за значимость главного эффекта фактора *Группа - Group*. Для построения графика средних для этой функции нужно на таблице нажать правую кнопку мыши и выбрать опцию *Быстрые статистические графики*. Затем выбрать *График средних* в выпадающем меню, после чего откроется окно *Выбор переменных для графика*. В этом окне нужно выбрать первую дискриминантную функцию (*Функц. 1*) и нажать кнопку *OK*.

На графике яснее проявляются разные типы групп. Очевидно, что на группе (*M10_S0*) первая дискриминантная функция имеет самое маленькое значение, на контрольной группе (*M5_S0_C*) - самое большое. Значения функции в остальных группах лежат между ними. Напомним, что в эту дискриминантную функцию переменная *Words* входит с отрицательным коэффициентом, а переменная *Cats* - с положительным. Поэтому наименьшее значение первой дискриминантной функции соответствует группе с наибольшим числом отсортированных слов и наименьшим числом восстановленных категорий. В целом, этот график согласуется с приведенным ниже графиком средних.



Выводы. Из проведенного анализа можно заключить, что в данном эксперименте существует главный значимый эффект. В частности, субъекты из группы 2 (*M10_S0*), которые сортировали слова в 10 категорий (без подкатегорий), отсортировали больше число слов, чем субъекты из других групп, но на меньшее число категорий. Ковариаты оказались не связанными с зависимыми переменными. В связи с этим можно было бы продолжить анализ с помощью контрастов, для дальнейшего объяснения значимости главного эффекта. Но вернемся теперь в диалоговое окно *Результаты дисперсионного анализа* для рассмотрения влияния ковариат.

Исследование гипотезы параллельности.

Одно из предположений стандартного ковариационного анализа состоит в том, что линии регрессии имеют одинаковый угол наклона (параллельны) в каждой группе, или другими словами нет попарных взаимодействий,

Copyright © StatSoft, 1995
StatSoft®

включающих ковариаты. В многомерном случае (многомерные зависимые переменные) предполагается, что регрессионные плоскости – параллельны. Можно проверить это предположение. Нужно нажать на кнопку *Описательные статистики и графики* в диалоговом окне *Результаты дисперсионного анализа.* В диалоговом окне *Описательные статистики и графики* нажмите на кнопку *Параллельность (взаимодействия ковариат)*. Сначала появится таблица с одномерными критериями для каждой зависимой переменной.

Одном. критерий параллельн. (mancova.sta)

ДИСПЕРС. АНАЛИЗ	cc=9,32			
зависим. перемен.	Сред.кв. Эффекта	Сред.кв Ошибки	F	p-уров.
WORDS	32,55085	36,51974	,891322	,543629
CATS	31,61927	56,79786	,556698	,821537

Затем будет построена таблица с многомерными результатами.

Многомерн.критерии параллельности (mancova.sta)

ДИСПЕРС. АНАЛИЗ	Значение	p-уров.
Лямбда Уилкса	,740759	
Рао R Форма1 (18, 62)	,557589	,915886
След Пиллая-Бартлетта	,273932	
V (18,64)	,564276	,911941

Ни один из этих критериев незначим. Поэтому нет причин для отклонения гипотезы о параллельности плоскостей.

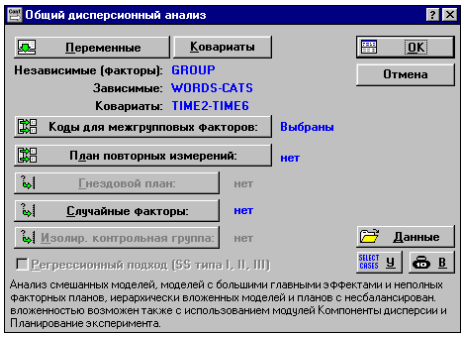
StatSoft®

Copyright © StatSoft, 1995

ДИАЛОГОВЫЕ ОКНА, ОПЦИИ, СТАТИСТИКИ

Стартовая панель

После запуска модуля *Дисперсионный анализ* открывается стартовая панель модуля.



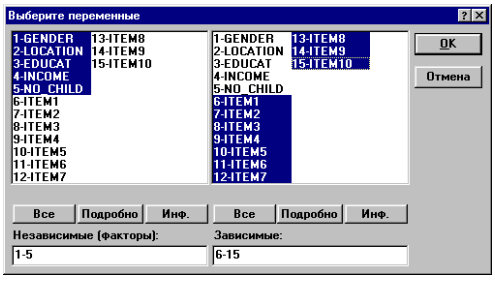
Межгрупповой план. Для задания межгруппового плана в этом диалоге должна быть выбрана хотя бы одна независимая или группирующая переменная (межгрупповой фактор). Также должно быть задано не менее двух кодов (через опцию *Коды*) для каждого межгруппового фактора. Заметим, что если коды не вводятся в явной форме, то программа будет использовать (по умолчанию) в качестве кодов все значения, встречающиеся в заданных независимых переменных.

Планы с повторными измерениями. Для того чтобы задать план, содержащий факторы повторных измерений, в открывшемся окне нужно задать хотя бы две зависимые переменные. План с повторными измерениями определяется через одноименную опцию стартовой панели. Множество зависимых переменных, которые не могут быть проинтерпретированы программой (с помощью плана, который был задан) как уровни факторов, повторных измерений, интерпретируются в плане многомерного дисперсионного анализа как множество зависимых переменные. Это может произойти, например, если выбраны две или более зависимых переменных, и они не определены как повторные измерения; или выбрано больше зависимых переменных, чем позволяют заданные факторы повторных измерений и их уровни.

Планы с пропущенными ячейками. Модуль *Дисперсионный анализ* автоматически обрабатывает планы с пропущенными ячейками. Для анализа *латинских квадратов*, *греко-латинских квадратов* или других сбалансированных неполных планов нужно задать их так, как если бы они были полными факторными планами. Затем, используется опция *Объединенный эффект и член ошибки* в диалоговом окне *Результаты дисперсионного анализа* для построения соответствующего члена ошибки при исследовании эффектов. Планы со сложной структурой пропущенных ячеек должны быть проанализированы с помощью специально спланированных сравнений.

Переменные (зависимые и независимые)

При выборе этой опции, появится диалоговое окно для выбора списка независимых переменных (межгрупповых факторов) и списка зависимых переменных.



Независимые (или группирующие) переменные - это переменные, содержащие коды, которые однозначно определяют к какой группе (ячейке)

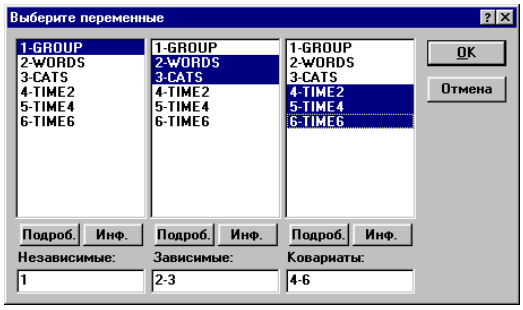
Copyright © StatSoft, 1995

StatSoft®

в межгрупповом плане принадлежит данный случай. Может быть задано до 8 переменных. Если план не имеет межгрупповых факторов, нужно оставить первый список пустым. Если требуется провести анализ повторных измерений, задайте как зависимые переменные *все* переменные, которые содержат данные для фактора (факторов) повторных измерений. Принадлежность переменной к уровням фактора (ов) повторных измерений определяется *порядком, в котором зависимые переменные расположены в списке зависимых переменных*, а не порядком, в котором эти переменные были введены в файл.

Ковариаты

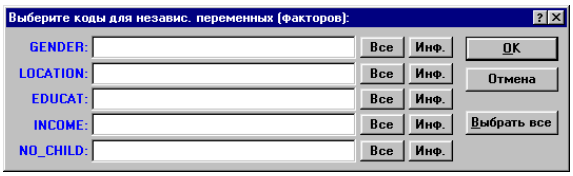
После выбора этой опции выводится три одинаковых списка для выбора переменных. В первом списке выбираются независимые или группирующие переменные, во втором – зависимые переменные (описанные выше). В третьем списке выбираются ковариаты (см. стр. 1721).



При выборе безразлично являются ли ковариаты фиксированными (постоянными) (каждая ковариата была измерена один раз для каждого субъекта) или переменными (каждая ковариаты была измерена при каждом повторном измерении зависимой переменной). Если число ковариат кратно общему числу уровней факторов повторных измерений плана, тогда после нажатия на кнопку *OK* и после выхода из этого окна, будет задан пользователю вопрос: являются ли ковариаты переменными или фиксированными.

Коды для межгрупповых факторов.

После нажатия на эту кнопку, откроется диалоговое окно выбора кодов для независимых переменных. Коды для независимых переменных могут быть либо числовыми, либо текстовыми и необходимы для определения групповой принадлежности каждого наблюдения (например, можно ввести 1или *Мужчина,* если субъект, участвующий в эксперименте, был мужчиной, ввести 2 или *Женщина,* если субъект - женщина). Для того чтобы увидеть, как коды используются для определения межгрупповых планов дисперсионного анализа, можно посмотреть раздел *Примеры* (стр. 1738).



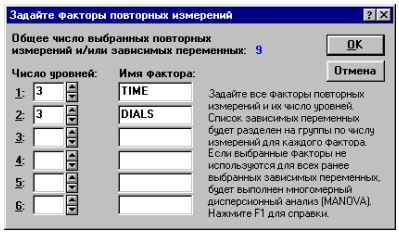
Если коды не заданы, то после нажатия на кнопку *OK* в стартовой панели программа будет автоматически использовать в качестве кодов все числовые значения соответствующей независимой переменной.

Планы с повторными измерениями (внутригрупповая сумма квадратов)

Эта опция служит для задания части плана, связанной с повторными измерениями. После выбора этой опции появится диалоговое окно для задания факторов повторных измерений.

StatSoft®

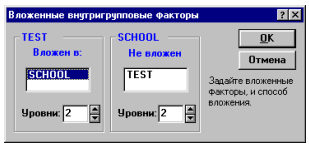
Copyright © StatSoft, 1995



Для каждого фактора задается имя, содержащее не более 8 знаков, и число уровней. Для получения более подробной информации о планах с повторными измерениями см. *Задание планов с повторными измерениями*, стр. 1760 и 1790.

Гнездовые (вложенные) планы

Эта опция служит для задания гнездовых планов. В гнездовых планах уровни некоторого фактора располагаются внутри уровней другого фактора. Например, если проверяются четыре различных теста в четырех разных классах высшей школы (т.е. межгрупповой фактор имеет четыре уровня), при этом два из четырех классов принадлежат высшей школе A, а другие два школе B. Тогда первый фактор (четыре разных теста) будет располагаться внутри второго фактора (две различные высшие школы).



Заметим, что не все гнездовые планы должны (или даже могут) задаваться с помощью этой опции (см. ниже).

После выбора этой опции появится окно *Вложение межгрупповых факторов* (см. выше), в котором пользователь должен выбрать:

(1) факторы, которые вкладываются,

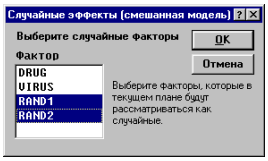
(2) факторы, в которые они вкладываются, и

(3) число уровней для каждого вложенного фактора, внутри уровней соответствующего над - фактора.

Факторы повторных измерений могут быть вложены только в факторы повторных измерений, межгрупповые факторы только в межгрупповые.

Случайные факторы

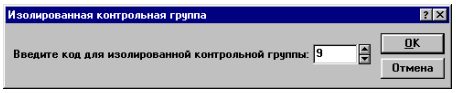
После выбора этой опции пользователь может задать факторы, которые рассматриваются как случайные.



Для получения более подробной информации о выборе случайных факторов см. *Случайные эффекты (смешанные модели)*, стр. 1801.

Изолированная контрольная группа

Эта опция позволяет задавать код независимой переменной для идентификации отдельной контрольной группы.



Подробнее об изолированной контрольной группе см. *Задание изолированной контрольной группы*, стр. 1788.

Регрессионный подход

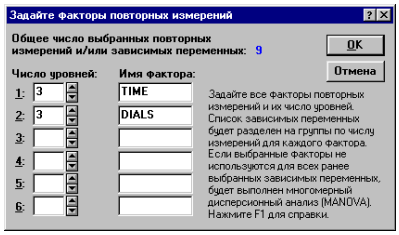
Если установлена опция *Регрессионный подход,* то программа будет автоматически выполнять дополнительные вычисления при считывании данных. Благодаря этому, впоследствии, пользователь сможет провести проверку всех

Copyright © StatSoft, 1995

эффектов или конкретных гипотез, а также объединенных эффектов, используя суммы квадратов типа *I*, *II* или *III*. (Эти опции описываются позднее, на стр. 1763, 1767, и 1778, соответственно.)

Задание планов с повторными измерениями

Для определения фактора повторных измерений необходимо ввести названия уровней и число уровней для этого фактора (см. также стр. 1790). Программа будет интерпретировать список зависимых переменных как уровни фактора повторных измерений. Поэтому число зависимых переменных должно быть кратно общему числу факторов повторных измерений. Если существует больше зависимых переменных (скажем, в два или три раза) общего числа уровней факторов повторных измерений, то соседние зависимые переменные в списке зависимых переменных будут интерпретироваться как многомерные зависимые переменные в многомерном дисперсионном анализе.



Задание более одного фактора повторных измерений

Если план содержит более одного фактора повторных измерений, то порядок, в котором они задаются, будет определять распределение зависимых переменных по уровням факторов повторных измерений. Программа, проходя по списку зависимых переменных, будет "циклически" приписывать переменной соответствующий уровень фактора повторных измерений. В частности, самое быстрое изменение уровня будет у фактора, который был задан последним, затем будет меняться уровень фактора, заданного предпоследним и т.д. Если были заданы *переменные* ковариаты (ковариаты, которые были замерены на каждом уровне фактора повторных измерений), списку ковариат будут приписаны уровни факторов повторных измерений тем же самым способом.

Пример. Если были выбраны 12 зависимых переменных, и нужно задать 2 x 2 x 3 план с повторными измерениями, то уровни будут приписываться следующим образом:

Фактор повторных измерений	Уровни факторов повторных измерений											
Первый	1	1	1	1	1	1	2	2	2	2	2	2
Второй	1	1	1	2	2	2	1	1	1	2	2	2
Третий	1	2	3	1	2	3	1	2	3	1	2	3
Зав.пер.	1	2	3	4	5	6	7	8	9	10	11	12

Заметим, что номера в последней строке этого примера характеризуют порядковое число переменных в списке зависимых переменных заданное пользователем, а не номера этих переменных в файле данных (например, число 1 в последней строке может соответствовать номеру 256 в файле данных).

Многомерные планы с повторными измерениями

Если бы в приведенном выше примере плана 2 x 2 x 3 с повторными измерениями, зависимая переменная состояла из 24 зависимых переменных, то был бы выполнен *многомерный* дисперсионный анализ (M*ANOVA*) (с двумя зависимыми переменными на каждом уровне факторов повторных измерений).

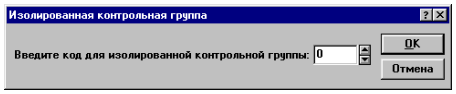
StatSoft®

Copyright © StatSoft, 1995

При циклическом просмотре списка зависимых переменных, быстрее всего меняющиеся уровни (для нескольких соседних переменных в списке) представляют многомерные зависимые переменные. Более детальное объяснение, как программа разбивает переменные из списка зависимых переменных по уровням факторов повторных измерений, содержится в Примере 2 (стр. 1744) и в разделе *Задание плана с повторными измерениями* (стр. 1790) (см. также стр. 1760).

Задание кода для изолированной контрольной группы

Задаваемый здесь код, используется в независимой (группирующей) переменной для определения наблюдений, которые принадлежат изолированной контрольной группе.



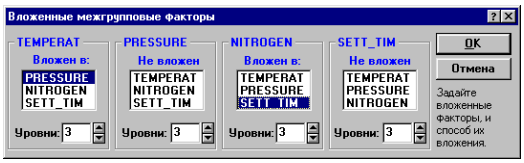
План эксперимента – это обычно план с полностью пересекающимися факторами (факторный план), т. е. каждый уровень каждого фактора появляется вместе с каждым уровнем другого фактора. Например, в плане 2 (типы лекарства) x 2 (типы вируса), каждый тип лекарства будет применен для каждого типа вируса. Однако экспериментальный план иногда требует добавление еще одной группы, т. е. *изолированной контрольной группы* (или “висячей” контрольной группы) (например, группы здоровых животных, не подвергнутых воздействию никакого вируса и никакого лекарства).

В модуле *Дисперсионный анализ* предполагается, что изолированная контрольная группа определяется *одним общим кодовым значением* - для *всех* независимых (группирующих) переменных. В качестве кода может быть введено любое целое число между -999 и +999. Позднее, в вычислениях сумм квадратов модуль *Дисперсионный анализ* будет учитывать изолированную контрольную группу при подсчете суммы квадратов ошибок.

С помощью опции *Спланированные сравнения* в диалоговом окне *Результаты дисперсионного анализа* могут быть заданы специфические сравнения, включающие изолированную контрольную группу. Для детального обсуждения анализа планов с изолированной контрольной группой нужно посмотреть раздел *Задание кодов для изолированной контрольной группы* (стр. 1788).

Задание гнездовых (вложенных) планов

Гнездовые (вложенные) планы задаются в диалоговом окне *Группировка межгрупповых факторов*.



Когда открывается это окно, появляется несколько столбцов (по одному на каждый фактор) с названием фактора наверху и с названиями остальных факторов ниже. Для задания вложенного фактора:

(1) выберите столбец с именем фактора (в заголовке столбца), который будет определен как вложенный;

(2) в этом столбце выберите фактор (факторы), в которых этот фактор будет вложен (выделив имена этих факторов, щелкнув на них). Теперь первый фактор (с именем в заголовке

Copyright © StatSoft, 1995

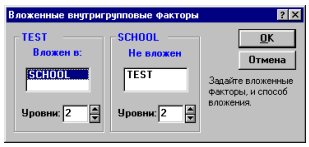
столбца) будет вложен в выделенный фактор (факторы);

(3) задайте число уровней этого фактора внутри каждого фактора (факторов), в который он вкладывается;

(4) повторите эту процедуру для каждого из вложенных факторов (в соответствующем столбце).

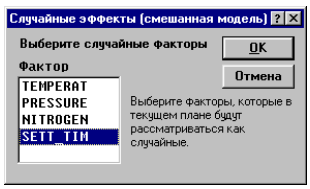
Гнездовые (вложенные) планы

В гнездовых планах уровни фактора группируются внутри уровней другого фактора. Например, проверялись четыре разных теста в четырех разных классах высшей школы (межгрупповой фактор с 4 уровнями) и два из этих классов принадлежат школе A, два других – школе B. Тогда уровни первого фактора (4 разных теста) будут сгруппированы во втором факторе (2 разных высших школы).



Задание случайных факторов (эффектов)

В диалоговом окне *Случайные эффекты (смешанная модель)* выбираются факторы, которые рассматриваются как случайные факторы в плане.



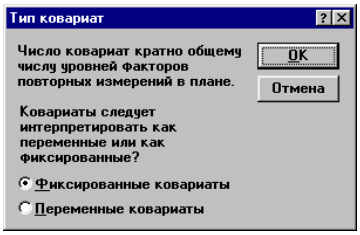
Фиксированные и случайные эффекты отличаются друг от друга тем, что уровни для фиксированных эффектов, главным образом, выбираются или “фиксируются” исследователем (т.е. задаются в эксперименте), в то время как уровни случайных эффектов выбираются исследователем *случайно* из совокупности возможных уровней (например, выбор трех различных школ для эксперимента).

Планы с несколькими случайными эффектами

Если в плане дисперсионного анализа содержится несколько случайных факторов или случайный фактор и факторы повторных измерений, то суммы квадратов (и F значения) для некоторых эффектов не определены и не могут быть оценены.

Фиксированные и переменные ковариаты

Если число ковариат в плане нацело делится на общее число уровней факторов повторных измерений, тогда список переменных, выбранных как ковариаты, может рассматриваться как фиксированные или переменные ковариаты.



Переменные ковариаты

Фиксированные ковариаты обычно обсуждаются в учебниках по одномерному дисперсионному анализу, в то время как переменные ковариаты обсуждаются гораздо реже. Когда имеются повторные измерения, можно интересоваться различием в повторных измерениях на одних и тех же объектах. Часто представляет интерес значимость *изменений* при разных измерениях.

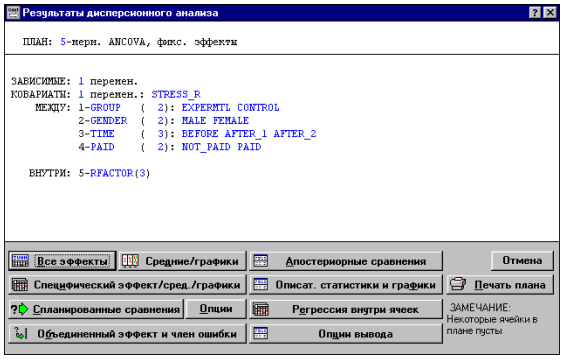
StatSoft®

Copyright © StatSoft, 1995

Если имеются ковариаты, то они также измеряются в тех же точках, что и зависимые переменные. Поэтому можно вычислить корреляцию между изменением ковариат и изменением зависимой переменной. Например, можно изучать интерес к математике и математические навыки в начале и в конце семестра. Было бы интересно узнать, коррелированы ли изменения в интересе к математике с изменениями в математических навыках в течение семестра.

Диалоговое окно *Результаты*

После выбора плана в стартовой панели нужно нажать кнопку *OK* и появится диалоговое окно *Результаты дисперсионного анализа*.



В этом окне можно выбрать различные опции вывода специальных результатов.

Все эффекты

Эта опция вычисляет полную таблицу всех эффектов и выводит эту таблицу на экран. В этой таблице значимые эффекты подсвечиваются, уровень значимости *альфа* может быть изменен с помощью кнопки *Сервис* на панели инструментов таблицы.

Средние/графики

Эта опция будет также вычислять полную таблицу всех эффектов. Диалоговое окно, которое появится затем на экране, позволяет выбирать отдельные эффекты и выводить на экран или график средних (взаимодействий), или таблицу средних. Эта опция позволяет самым быстрым способом описать характер значимых эффектов в сложных планах одномерного дисперсионного анализа (дополнительные сведения даны на стр. 1767).

Специфический эффект/средние/графики

После нажатия на эту кнопку откроется окно, в котором можно выбрать эффекты для более подробного их исследования. Для выбора результатов представленных здесь, нужно нажать на кнопку *Опции вывода* (см. ниже) и задать конкретное представление данных. Правило задания эффектов описано на стр. 1767.

Спланированные сравнения

Эта опция позволяет исследовать специфические контрасты (*анализ контрастов*). Тип контрастов (отдельный или совместный), выбираемый в окне *Опции для спланированных сравнений (*она доступна с помощью кнопки *Опции*, см. ниже), определяет, какое диалоговое окно откроется после нажатия на кнопку *Спланированные сравнения.* Если было выбрано *Вводить контрасты отдельно для каждого фактора,* то будет предложено вводить коэффициенты контрастов одного фактора за другим в диалоговом окне *Задание контрастов.* Если было выбрано *Вводить контрасты для всех факторов вместе*, тогда коэффициенты контрастов нужно будет вводить в диалоговом окне *Контрасты для межгрупповых (внутигрупповых) факторов.* Общие правила построения контрастов описаны на стр. 1785 как

Copyright © StatSoft, 1995

для межгрупповых планов, так и для планов с повторными измерениями. В диалоговых окнах используется определение контрастов, описанное на стр. 1781.

Опции

После нажатия на кнопку *Опции спланированных сравнений* откроется диалоговое окно, в котором можно выбрать способ, каким будут вводиться контрасты в опцию *Спланированные сравнения,* о которой говорилось выше. Контрасты могут вводиться отдельно для каждого фактора (если выбрана опция *Вводить контрасты отдельно для каждого фактора*), или для всех факторов *вместе*, , если выбрана опция *Вводить контрасты для всех факторов вместе*). В многомерных планах можно также выбрать ввод контрастов для зависимых переменных. Для получения дополнительных сведений см. стр. 1781.

Объединенный эффект и член ошибки

После выбора этой опции, пользователь может задать эффекты, которые будут объединены, и эффекты, которые будут отнесены к ошибкам в диалоговом окне *Объединенные эффекты и члены ошибки*. Эта опция особенно полезна при исследовании эффектов в неполных планах (см. стр. 1778).

Апостериорные сравнения

После выбора этой опции появляется окно *Задайте эффект для апостериорных критериев,* в котором задаются эффекты для *апостериорных* сравнений. Условия задания эффектов в этом окне те же, что и при выборе опции *Специфический эффект/средние/графики* (см. выше). *Апостериорные* критерии основываются на маргинальных средних соответствующих эффектов. Среди доступных *апостериорных* критериев – критерий *НЗР* (наименьшей значимой разности), Ньюмена - Кеулса, критерий множественного сравнения Дункана, критерий Шеффе, критерий Тьюки *ДЗР* (достоверно значимой разности) и *ДЗР* критерий Спйотвола – Столайна – обобщение критерия Тьюки на выборки неравного объема. Эти критерии могут быть выбраны в диалоговом окне *Апостериорные сравнения средних,* которое появляется после задания эффекта (дополнительная информация дана на стр. 1769).

Описательные статистики и графики

Эта опция предоставляет диалоговое окно *Описательные статистики и графики* с большим набором опций вычисления описательных статистик, исследования предположений одномерного дисперсионного анализа и различных графического отображения различных графических аспектов изучаемых данных. Для получения дополнительной информации см. стр. 1771.

Регрессия внутри ячеек

В этой опции выполняется регрессионный анализ для ковариат. Результаты могут быть выведены отдельно для каждой зависимой переменной (одномерной), или для всех зависимых переменных одновременно (многомерной переменной). Заметим, что в этой опции будет производиться анализ всегда отдельно для каждой зависимой переменной из списка зависимых переменных даже, если эти переменные являются уровнями фактора повторных измерений. Если в плане с повторными измерениями нужно просмотреть результаты для зависимой переменной, которая содержит уровней, то можно использовать *Опции вывода* (см. ниже), чтобы потребовать *Результаты регрессии*. Затем, используя опцию *Специфический эффект/средние/графики* (см.

StatSoft®

Copyright © StatSoft, 1995

выше), можно рассмотреть интересующие эффекты (см. также стр. 1767).

Опции вывода

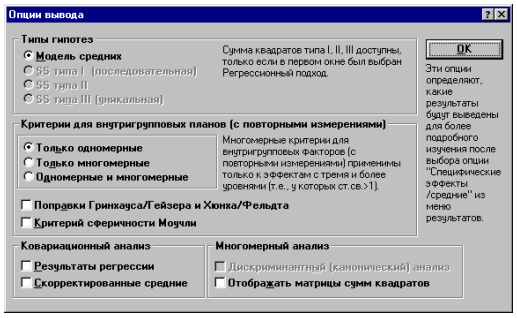
Опция открывает диалоговое окно *Опции вывода* для выбора специфических опций вывода, которые появятся при исследовании специфических эффектов с помощью опции *Специфический эффект/средние/графики,* описанной выше. Заметим, что к некоторым планам не все опции применимы и при установке этих опций они будут игнорироваться.

Печать плана

После нажатия на эту кнопку, план дисперсионного анализа будет либо распечатан на принтере, либо занесен в выходной файл (как определено в диалоговом окне *Параметры страницы/вывода*).

Опции вывода

В этом диалоговом окне выбираются конкретные результаты для вывода при исследовании специфических эффектов с помощью опции *Специфический эффект/средние/графики*.



Заметим, что не все опции применимы ко всем планам.

Типы гипотез

Выбирается тип гипотез для анализа (см. ниже).

Модель средних. Если в стартовой панели модуля *Дисперсионный анализ* не установлена опция *Регрессионный подход*, то по умолчанию программа использует *Модель средних по ячейкам*. Точнее, суммы квадратов для различных эффектов вычисляются с помощью линейных комбинаций средних по ячейкам. Для полного факторного эксперимента это приводит к суммам квадратов, совпадающим с суммами квадратов, которые обсуждаются ниже, как суммы квадратов *типа III.* Используя опцию *Спланированные сравнения* в диалоговом окне *Результаты дисперсионного анализа*, пользователь может исследовать гипотезы о любых взвешенных или не взвешенных линейных комбинациях средних по ячейкам. Таким образом, пользователь может исследовать гипотезы не только *типа III*, но также любой тип гипотез (включая *Тип IV*). Этот общий подход оказывается особенно полезным при анализе планов с пропущенными ячейками (так называемыми неполными планами). Для получения дополнительной информации см. стр. 1729.

Гипотезы типа I, II, и III. Эти опции доступны, если в стартовой панели установлена опция *Регрессионный подход.* Вообще говоря, в линейных регрессионных моделях можно исследовать значимость частных коэффициентов корреляции эффекта, контролируя все другие эффекты, или вводить переменные пошаговым способом, контролируя все другие факторы, которые уже были введены, игнорируя те, которые еще не введены. В сущности, в этом и состоит различие между суммами квадратов *типа III* и *типа I*. Для дополнительной информации см. стр. 1726.

Copyright © StatSoft, 1995

Другая “промежуточная” стратегия состоит: в исследовании отдельного главного эффекта при контролировании всех других главных эффектов; в исследовании отдельного эффекта попарного взаимодействия при контролировании всех главных эффектов и других эффектов попарных взаимодействий; в исследовании отдельного взаимодействия трех факторов, при контролировании всех главных эффектов, всех эффектов попарных взаимодействий и всех остальных эффектов взаимодействия трех факторов и т.д. Суммы квадратов для эффектов, вычисленные таким способом, называются суммами квадратов *типа II*. Таким образом, суммы квадратов *типа II* контролируют все эффекты того же порядка и ниже, что и исследуемый эффект, игнорируя эффекты более высокого порядка. Дополнительная информация дана на стр. 1726.

Критерии для внутригрупповых планов (планов с повторными измерениями)

Эти опции применимы лишь для планов с факторами повторных измерений, содержащими более двух уровней. В разделе *Вводный обзор* обсуждается преимущество многомерных критериев для исследования эффектов этих факторов по сравнению с традиционным одномерным подходом (стр. 1732). Короче говоря, важное предположение, лежащее в основе одномерных статистических критериев для факторов с повторными измерениями, так называемое предположение о сложной симметрии. Это предположение, которое часто нарушается, не требуется для многомерного подхода. Нужно установить соответствующие зависимые кнопки для вычисления *только одномерных* критериев, *только многомерных* критериев, или *одномерных и многомерных* (по умолчанию).

Критерии Гринхауса - Гейзера и Хюнха - Фельдта. Цель этих критериев скорректировать одномерные результаты при проведении дисперсионного анализа с повторными измерениями (для факторов с числом уровней больше, чем 2) в случае нарушения предположения о сложной симметрии. Это предположение, так же как и эти критерии, обсуждаются в разделе *Вводный обзор* (стр. 1732).

Критерий сферичности Моучли. Этот критерий применим только для планов с факторами повторных измерений, с числом уровней больше 2. Предположение *сферичности* связано с предположением *сложной симметрии* и оба этих предположения касаются условий, при которых одномерный подход к планам с повторными измерениями приводит к статистически обоснованным критериям. Эти предположения обсуждались в разделе *Вводный обзор* (стр. 1732).

Критерий сферичности Моучли проверяет гипотезу о том, что выполнены предположения сферичности (нулевая гипотеза). Этот критерий либо принимает гипотезу сферичности, либо отвергает ее. Однако исследования с помощью метода Монте-Карло (см. работы Rogan, Keselman, and Mendoza, 1979; Keselman, Rogan, Mendoza, and Breen, 1980) показали, что критерий Моучли очень чувствителен относительно отклонений от многомерной нормальности. Даже незначительное нарушение предположений сферичности ведет к ошибочным выводам в одномерном дисперсионном анализе (см. Bock, 1975; Box, 1954; Kirk, 1982). Поэтому при повторных измерениях всегда рекомендуется вычислять многомерные статистики.

Ковариационный анализ

Следующие две опции применимы лишь в том случае, когда план содержит ковариаты и (в случае *Скорректированных средних*) межгрупповые факторы.

StatSoft®

Copyright © StatSoft, 1995

Результаты регрессии. Эта опция определяет, нужно ли выводить результаты регрессии для ковариат (чтобы включить вывод - щелкните на ней мышкой).

Скорректированные средние. Чтобы на экран выводились скорректированные средние (скорректированные ковариатами), щелкните на этой опции.

Многомерный дисперсионный анализ

Следующие две опции доступны при проведении многомерный дисперсионный анализ.

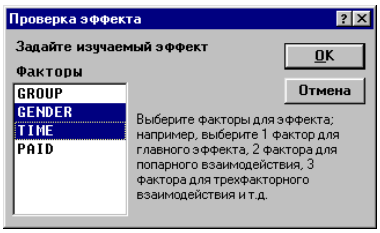
Дискриминантный (канонический) анализ. В многомерных планах с межгрупповыми факторами могут быть построены дискриминантные функции (см. также подробно об этой процедуре в разделе *Дискриминантный анализ*, том III). При выборе этой опции программа вычислит линейную комбинацию зависимых переменных, по значениям которой группы наиболее различимы. Можно вычислить также несколько таких линейных комбинаций (или корней). Число линейных комбинаций не больше минимума из двух чисел: (1)число групп минус один; (2) число зависимых переменных.

Отображать матрицы сумм квадратов. При установке этой опции на экран выводятся матрицы сумм квадратов гипотез и ошибок. Когда межгрупповой план многомерен (имеется несколько зависимых переменных), то на диагонали этих матриц стоят суммы квадратов, соответствующих гипотезам или ошибкам, а вне диагонали — суммы соответствующих попарных произведений.

Если в плане присутствует только одна зависимая переменная, но имеется фактор повторных измерений с более чем двумя уровнями, то при проверке гипотез, включающих повторные измерения, матрицы сумм квадратов будут содержать соответствующие суммы квадратов и попарных произведений (*SSCP*) для внутригрупповой модели.

Проверка эффекта

В этой опции определяется эффект, который нужно исследовать подробнее. Можно нажать на кнопку *Опции вывода* и задать, какие результаты нужно выводить для выбранных эффектов.



Правила задания эффектов просты:

- для исследования главного эффекта, выберите (выделите) соответствующий фактор;
- для исследования взаимодействия между двумя факторами, выделите два соответствующих фактора;
- для исследования взаимодействия трех факторов, выберите соответствующие три фактора, и т.д.

Заметим, что выбор нескольких факторов осуществляется при нажатой клавише CTRL с помощью левой кнопки мыши.

Результаты регрессии (ковариаты)

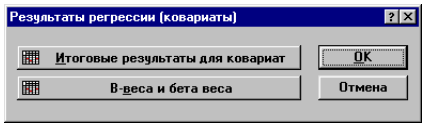
Это диалоговое окно доступно лишь для планов, содержащих ковариаты, и появляется на экране в следующих двух случаях:

(1) если пользователь выбрал опцию *Регрессия внутри ячеек* в диалоговом окне *Результаты*;

Copyright © StatSoft, 1995

(2) если пользователь выбрал опцию *Результаты регрессии* в диалоговом окне *Опции вывода* и исследуются *специфический эффект* (с помощью опции *Специфический эффект/средние/графики*).

Ковариаты и повторные измерения. Результаты, выводимые на экран в двух случаях, описанных выше, различаются между собой, если план содержит факторы повторных измерений. Если пользователь выбирает опцию *Регрессия внутри ячеек* из диалогового окна *Результаты*, то результаты, выводимые на экран, касаются регрессии ковариат на каждую переменную из списка зависимых переменных, даже если эти переменные определяют уровни фактора повторных измерений. Во втором случае (при использовании опции *Специфический эффект*), осуществляется регрессия ковариат на соответствующую "логическую" зависимую переменную. Если ковариаты *фиксированы* (*постоянны)*, то "логическая" зависимая переменная - это всегда сумма по всем уровням всех факторов повторных измерений. Если ковариаты являются *переменными*, то результаты регрессии относятся к регрессии ортонормальной матрицы эффекта/ошибки для ковариат на ортонормальную матрицу эффект/ошибка для зависимых переменных.



Итоговые результаты для ковариат

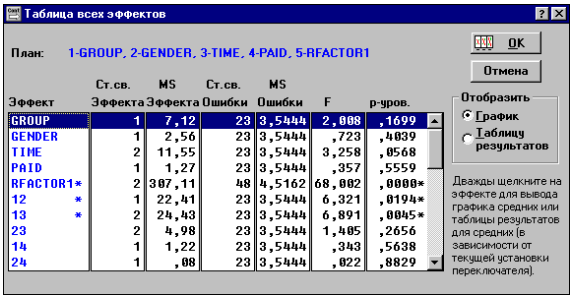
Если имеется только одна зависимая переменная, то на экран в таблице одномерного дисперсионного анализа будут выведены результаты для всех ковариат. В многомерном случае будут выведены соответствующие многомерные критерии.

B-веса и бета-веса

Эта опция позволяет пользователю просмотреть простые и стандартизованные коэффициенты регрессии (*B-* и *бета - веса*, соответственно) и их статистическую значимость для каждой ковариаты и для каждой зависимой переменной.

Выбор эффектов для графиков средних (взаимодействий) или таблиц средних

После нажатия кнопки *Средние/графики* в диалоговом окне *Результаты*, появится диалоговое окно *Таблица всех эффектов,* в котором нужно выбрать переменные или их взаимодействия для построения графика средних или таблиц средних (нужно выбрать в опции *Отобразить* зависимую кнопку *График* или *Таблицу результатов*).



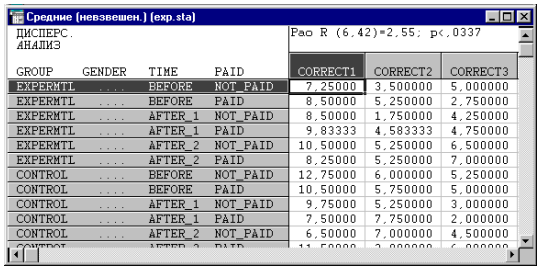
Это диалоговое окно предоставляет итоговую сводку результатов дисперсионного анализа и можно выбрать эффект для подробного рассмотрения. Например, можно рассмотреть значимые эффекты, помеченные звездочкой (*). Выбрав желаемый эффект или взаимодействие, нужно щелкнуть на имени этого эффекта или взаимодействия, подсветив его. Затем нажать *OK* для вывода на экран графика или таблицы.

StatSoft®

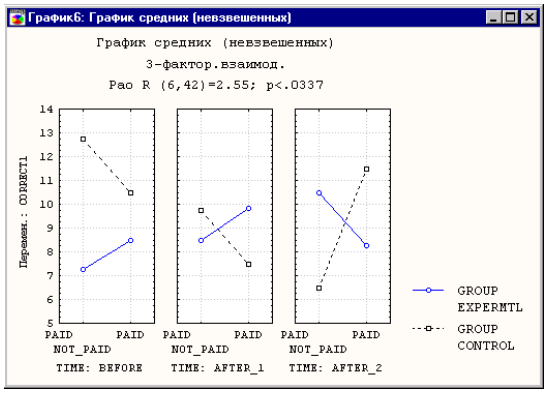
Copyright © StatSoft, 1995

Отобразить

Таблицы результатов. При выборе этой опции и нажатия *OK* в этом диалоговом окне на экране появится таблица средних для подсвеченных главных эффектов или взаимодействий.

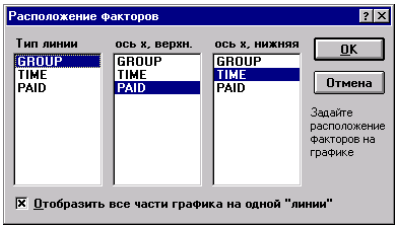


Графики. Когда выбрана эта опция, то после нажатия кнопки *OK* в диалоговом окне появляется на экране график средних для подсвеченных главных эффектов и взаимодействий.



Назначение факторов осям графика

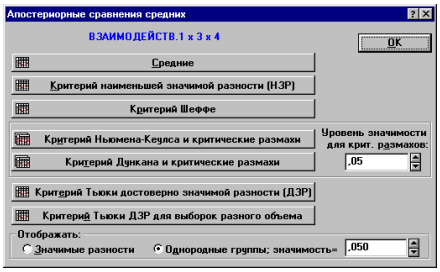
Когда строится график взаимодействий (см. выше *Средние/графики*), можно выбрать сопоставление факторов типу линии, верхней *X* – оси и нижней *X* – оси в диалоговом окне *Расположение факторов*.



Для взаимодействия трех или большего числа факторов имеется опция для показа всех графиков или в одной строке (эта опция *Отобразить все части графика на одной линии,* показывающая все графики один за другим в одной строке). В противном случае все графики располагаются более чем в одной строке. Если нужно построить графики более чем трех факторов, тогда соответствующие графики строятся при фиксированных уровнях всех других (не выбранных) факторов.

Апостериорные сравнения

Это диалоговое окно позволяет выбирать различные *апостериорные* процедуры. Можно обратиться к разделу *Вводный обзор* (стр. 1724), для обсуждения базисной логики, стоящей за этими критериями. Обсуждение *апостериорных* процедур содержится также в Winer (1962), Hays (1988), или Milliken and Johnson (1984). Эти критерии будут вычислены для маргинальных средних выбранных эффектов.



Например, если были выбраны главные эффекты, то *апостериорные* критерии будут выполнены

Copyright © StatSoft, 1995

StatSoft®

для маргинальных средних на каждом уровне соответствующего фактора. Если требуется проверить *апостериорные* критерии для эффектов взаимодействий 2 x 2, то *апостериорные* критерии будут выполнены для четырех средних, которые образуют таблицу взаимодействий 2 x 2.

Выбор члена ошибки. *Апостериорные* критерии основаны на членах ошибки из соответствующего общего анализа. *Апостериорные* критерии могут применяться как для межгрупповых факторов, так и для факторов повторных измерений.

Средние

В этой опции строится таблица с маргинальными средними для соответствующего эффекта. Как и для других таблиц маргинальных средних в модуле *Дисперсионный анализ,* эти средние могут быть представлены графически с помощью опции *Быстрые статистические графики* из выпадающего меню (появляющегося после нажатия правой кнопки мыши).

Критерий наименьшей значимой разности (НЗР)

Этот критерий эквивалентен *t*-критерию для независимых или зависимых выборок (см. также *Основные статистики и таблицы*, глава 9), основанному на числе наблюдений *n,* в соответствующих ячейках плана, включенных в сравнение. Критерий предлагает некоторую защиту от завышенного уровня ошибки первого рода (*альфа*), возникающую из-за многократных *апостериорных* сравнений. После выбора этой опции, в таблице выводится матрица *p*-значений. Эти *p*-значения указывают *апостериорную* значимость уровней для соответствующих пар средних.

Критерий Шеффе

Эта опция строит таблицу с *апостериорными p*-уровнями критерия Шеффе. Критерий Шеффе обычно более осторожный, чем критерий Ньюмена - Кеулса или критерий Дункана (см. Winer, 1962).

Критерий Ньюмена - Кеулса и критические размахи

Этот критерий основан на статистике *стьюдентизированного* размаха. Программа сначала упорядочивает средние значения по возрастанию. Затем последовательно (начиная с макимально удаленной пары, т.е. для пары с максимальным размахом) оценивается вероятность при нулевой гипотезе (нет различия между средними) того, что разность средних будет не меньше полученного значения, при соответствующем числе выборок. Этот критерий на самом деле является критерием значимости *размахов*, при соответствующем числе выборок. Заметим, что модуль *Дисперсионный анализ* не просто выдает значения критерия, но вычисляет вероятности, основанные на распределении статистики *стьюдентизированного размаха.*

После выхода из таблицы, содержащей матрицу *p*-значений для пар средних, программа будет вычислять *критические размахи* между упорядоченными средними, при соответствующем уровне α (по умолчанию $p<.05$). Можно изменить уровень α, используя опцию *Уровень* значимости *для критических размахов*.

Критерий множественных сравнений Дункана и критические размахи

Этот критерий основан на той же логике, что и процедура Ньюмена-Кеулса; однако в нем используется менее консервативный

StatSoft®

Copyright © StatSoft, 1995

проверочный критерий (см., например, Milliken and Johnson, 1984).

Критерий Тьюки достоверно значимой разности (ДЗР)

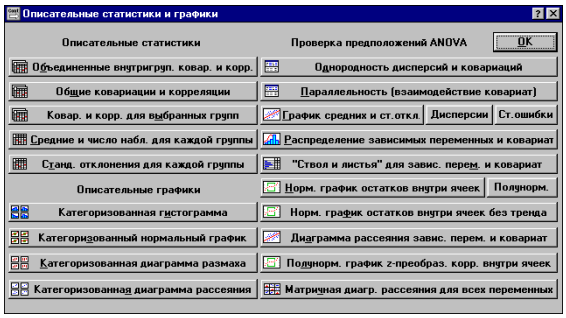
Критерий *ДЗР* Тьюки располагается между критериями Ньюмена-Кеулса и Шеффе с точки зрения консервативности. (Подробное описание этого критерия приведено в Главе 9.)

Критерий ДЗР Тьюки для выборок разного объема (критерий Спйотвола и Столайна)

Этот критерий обобщает критерий Тьюки на выборки неравных объемов (см. Spjotvoll and Stoline, 1973, стр. 975).

Описательные статистики и проверка предположений дисперсионного анализа

При нажатии на кнопку *Описательные статистики и графики* в диалоговом окне *Результаты дисперсионного анализа* откроется одноименное диалоговое окно.



Описательные статистики

В разделе *Описательные статистики* доступны следующие опции.

Объединенные внутригрупповые ковариации и корреляции. С помощью этой опции на экран можно вывести объединенную внутригрупповую ковариационную и корреляционную матрицу зависимых переменных и ковариат. Сначала будет выведена ковариационная матрица, затем объединенная внутригрупповая корреляционная матрица.

Общие ковариации и корреляции. Эта опция выводит на экран общую ковариационную и корреляционную матрицу зависимых переменных и ковариат.

Ковариации и корреляции для выбранных групп. Эта опция полезна при исследовании ковариационной матрицы для выбранных групп (ячеек в межгрупповом плане). После выбора этой опции появляется на экране список ячеек в межгрупповом плане. Из этого списка выбираются ячейки, для которых затем на экран выводятся ковариационная и корреляционная матрицы зависимых переменных и ковариат.

Средние и число наблюдений для каждой группы. Эта опция выводит таблицу средних каждой зависимой переменной и ковариаты, а также число наблюдений (величину *n*) для каждой ячейки межгруппового плана. Выводится также общее среднее и общее число наблюдений *n*.

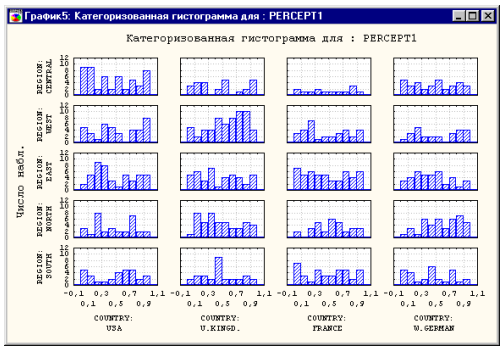
Стандартные отклонения для каждой группы. Эта опция выводит на экран стандартные отклонения для каждой зависимой переменной и ковариаты, а также число наблюдений (величину *n*) для каждой ячейки межгруппового плана. Также выводятся общее стандартное отклонение и общее число наблюдений *n*.

Copyright © StatSoft, 1995

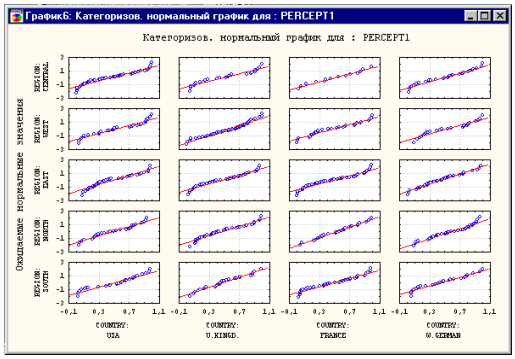
Описательные графики

Следующие опции позволяют строить конкретные описательные графики. После выбора одной из этих опций открывается диалоговое окно выбора “эффекта”, для которого строится график. График будет категоризован относительно выбранных факторов, без учета факторов, которое не были выбраны.

Категоризованные гистограммы. Эта опция позволяет визуально исследовать распределение текущей выделенной переменной категоризованной по уровням выбранных перед этим группирующих факторов.

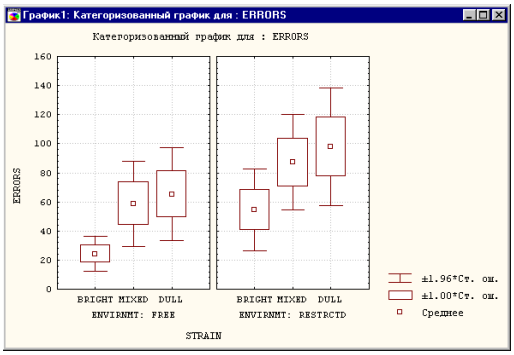


Категоризованный нормальный вероятностный график. При нажатии на эту кнопку на экран выводятся нормальные вероятностные графики, отдельно для всех групп, определяемых уровнями факторов, выбранных заранее.



Эти графики дают представление о нормальности распределения выбранного фактора по ячейкам выбранных группирующих факторов. Предположение о нормальности (данные выбраны из совокупности с нормальным распределением) является очень важным предположением в статистике (см. *Элементарные понятия статистики*, глава 8), и нормальные вероятностные графики часто используются для проверки этого предположения.

Категоризованные диаграммы размаха. Можно выбрать один из четырех типов категоризованных диаграмм размаха для представления распределения категоризованного значения выбранного фактора.

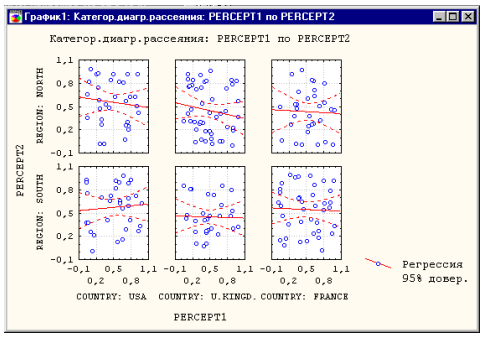


Графики этого типа дают представление о расположении “центра” и о разбросе для выбранного фактора в каждой категории, определяемой заранее выбранными группирующими факторами. В качестве “центра” можно выбрать среднее значение или медиану в каждой категории, а в качестве меры разброса – стандартное отклонение или стандартную ошибку. Все это можно выбрать в диалоговом окне *Тип диаграммы размаха.* Заметим, что специальный метод вычисления медиан и квартилей может быть выбран в диалоговом окне *Параметры по умолчанию: общие настройки*, доступном при выборе опции *Общие настройки* из выпадающего меню *Сервис (*см. главу 3 для

StatSoft®
Copyright © StatSoft, 1995

дополнительных сведений; см. также *Приложение 2* о деталях вычислений).

Категоризованные диаграммы рассеяния. При нажатии на эту кнопку строятся диаграммы рассеяния двух факторов, для всех ячеек, образованных уровнями группирующих факторов, выбранных заранее.



Категоризованная диаграмма рассеяния дает наглядное представление о взаимосвязи между двумя переменными *X* и *Y* (например, вес и рост). Каждое отдельное наблюдение из множества данных изображается точкой в двумерном пространстве, на координатных осях отложены значения переменных. Две координаты (*X* и *Y*), определяющие положение каждой точки, равны значению двух переменных.

Проверка предположений дисперсионного анализа

Следующие опции используются для исследования некоторых предположений. Обсуждение их дано в разделе *Предположения и эффекты нарушения предположений* (стр. 1730).

Однородность дисперсии и ковариаций. Одним из предположений одномерного дисперсионного анализа является предположение о равенстве (однородности) дисперсии по всем ячейкам межгруппового плана. В многомерном дисперсионном анализе (*MANOVA*) это предположение относится к ковариационной матрице зависимых переменных (и ковариат). Эти предположения и эффекты их нарушений обсуждались во *Вводном обзоре* (стр. 1730). После выбора этой опции, появляется диалоговое окно *Критерии однородности дисперсии/ковариаций.* В этом окне можно выбрать одномерный (*С-критерий Кохрена*, *критерий Хартли*, *Бартлетта* или *Левена*) или многомерный критерий (*М-критерий Бокса, непараметрический критерий Сена и Пури*).

Параллельность регрессионных линий и плоскостей (взаимодействие ковариат). Одним из обычных предположений ковариационного анализа является предположение о равенстве регрессионных моделей для разных ячеек межгруппового плана. С помощью этой опции может быть исследовано это предположение. При этом сначала вычисляется общая регрессионная модель и остаточная ковариационная матрица общей модели в каждой ячейке. Затем вычисляется регрессионная модель для каждой ячейки и опять считается для каждой ячейки остаточная ковариационная матрица. По существу, исследование предположения о параллельности регрессионных линий (плоскостей) равносильно сравнению остаточных ковариационных матриц каждой ячейки при разных способах подсчета регрессионных моделей.

Если имеется несколько межгрупповых факторов, то пользователь может задать факторы, которые будут игнорироваться (в соответствующих ячейках плана регрессионная модель будет одинакова) и факторы, в соответствующих ячейках которых будут отдельно строиться регрессионные модели. Таким образом, пользователь может отдельно исследовать *взаимодействие,* между ковариатами и межгрупповыми факторами.

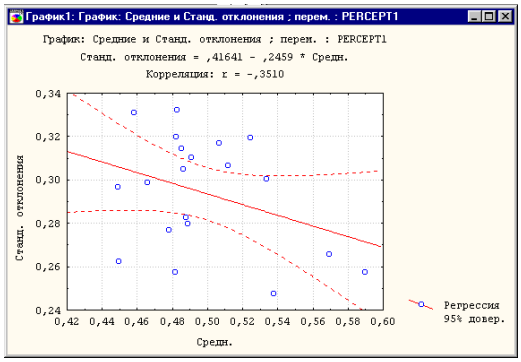
Если в плане имеется несколько межгрупповых факторов, то после выбора этой опции откроется

Copyright © StatSoft, 1995

StatSoft®

диалоговое окно для выбора факторов. По умолчанию, задаются все факторы. Если выбраны все факторы, то проверяется различие регрессионных линий или плоскостей для всех ячеек межгруппового плана. Если зависимая переменная многомерна, то будут выданы как одномерные (для каждой зависимой переменной), так и многомерные результаты.

График средних и стандартных отклонений/дисперсий/стандартных ошибок. Эти опции позволяют строить график средних с соответствующими мерами разброса (можно выбрать или *стандартное отклонение* или *дисперсию (вариацию)* или *стандартную ошибку* по всем ячейкам межгруппового плана.



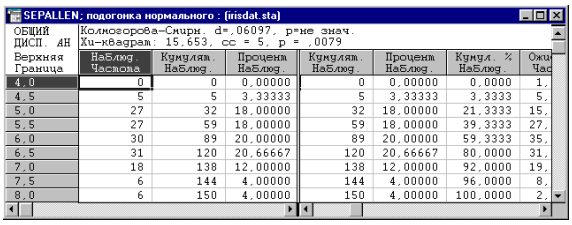
В разделе *Предположения и эффекты нарушения предположений* (стр. 1730) обсуждался случай, когда общий *F* критерий может давать ошибочные результаты, если средние по ячейкам коррелированы с дисперсией (вариацией). Короче говоря, если в ячейке с экстремальным средним (большим или маленьким) дисперсия (вариация) велика, то в этой ячейке имеется выброс (резко выделяющееся наблюдение). В результате доверительные интервалы соответствующего среднего гораздо больше доверительных интервалов для других ячеек и при применении в этой ситуации методов дисперсионного анализа, все доверительные интервалы увеличиваются, что делает результаты статистически более значимыми. В реальных исследованиях это случается довольно часто, поэтому перед принятием окончательных решений рекомендуется строить описываемые здесь графики. Если средние по ячейкам коррелированы с разбросом, то можно:

(1) пытаться выявить выбросы и исключить их из анализа,

(2) использовать непараметрические методы, или

(3) использовать преобразования (такие как логарифмическое *log* преобразование) к данным для того, чтобы “отбросить длинные хвосты” распределений с большим разбросом.

Распределение зависимых переменных и ковариат. Эта опция позволяет получать гистограммы для выбранных зависимых переменных и ковариат или для групп наблюдений или для всех данных. Сначала появляется таблица наблюдаемых частот и ожидаемых частот при нормальном распределении.

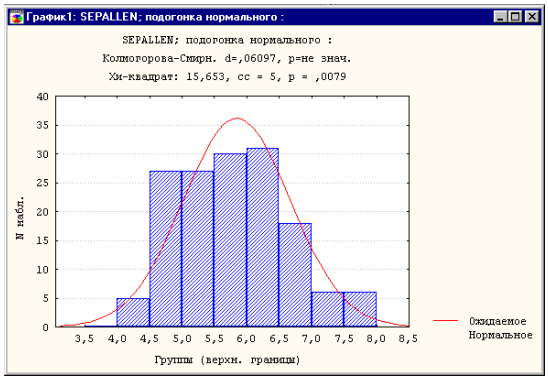


Здесь же подсчитывается значения критерия Колмогорова-Смирнова и критерия хи-квадрат.

Затем появляется график гистограммы с нормальной кривой, подобранной по наблюдаемым данным.

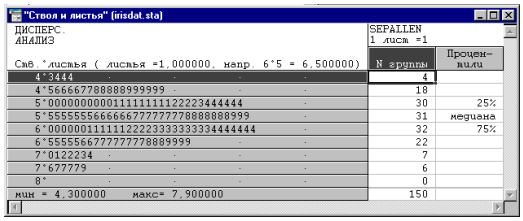
StatSoft®

Copyright © StatSoft, 1995



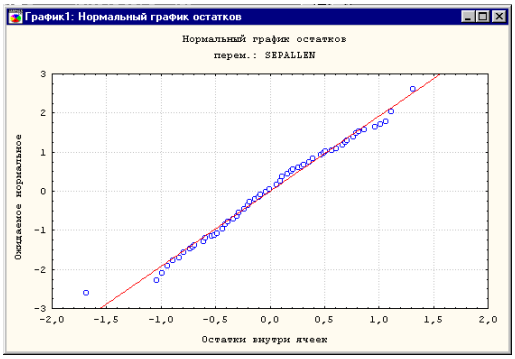
Обычно только при значительном отклонении от нормальности возникают проблемы с обоснованностью *F* статистики. Линдман (Lindman 1974, стр. 32) обсуждал влияние отклонения от нормальности, рассматривая *эксцесс* распределения. Исследования методом Монте-Карло позволили ему сделать вывод, что в этом смысле *F* статистика устойчива. Однако дополнительные “возвышенности” на графике могут быть очень полезны для выявления выбросов или мультимодальности (нескольких экстремумов) распределения.

График ствол и листья для зависимых переменных и ковариат. Этот график – альтернатива гистограмме (см. выше). Как и гистограмма, график ствол и листья строится для выбранных переменных или только для выбранных групп или для всех наблюдений сразу.



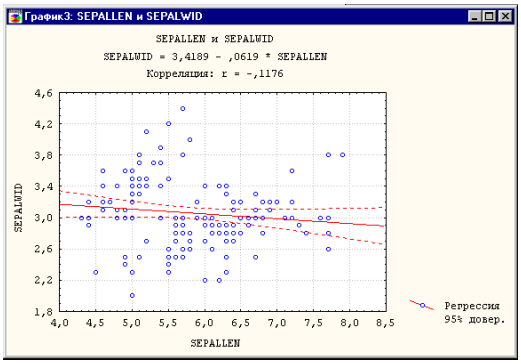
На графике ствол и листья каждый ствол изображает интервал, так же как в обычной гистограмме. Однако в отличие от гистограммы, где длина вертикальной полосы равна числу наблюдений, попавших в этот интервал (действительные значения строятся как *листья ствола*).

Нормальные и полунормальные вероятностные графики. Эти опции позволяют строить для отклонений внутри ячеек от соответствующих средних *нормальные* вероятностные графики (щелкните на кнопку *Нормальный график остатков внутри ячеек*), *полунормальные* вероятностные графики (щелкните на кнопку *Полунормальный*) и др.



Эти опции предоставляют еще одну возможность визуальной проверки предположений нормальности.

Диаграмма рассеяния зависимых переменных и ковариат. Эта опция позволяет строить диаграмму рассеяния для зависимых переменных и ковариат.



Copyright © StatSoft, 1995

StatSoft®

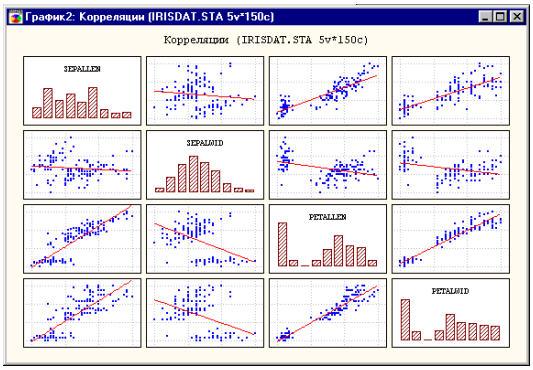
Эти диаграммы можно строить для всех наблюдений вместе или, разбив их на группы, для каждой группы наблюдений отдельно. Программа сначала предложит выбрать две переменные для графика, а затем спросит, строить график для групп по отдельности или все группы объединить в одну.

Полунормальный вероятностный график для z-преобразованных корреляций внутри ячеек. Эта опция доступна лишь при наличии нескольких зависимых переменных или ковариат в плане. Опция полезна для исследования распределения корреляций между переменными внутри ячеек.

Если корреляции однородны по группам (предположение многомерного дисперсионного анализа), то можно было бы ожидать, что z – преобразованные значения корреляции по группам будут нормально распределены.

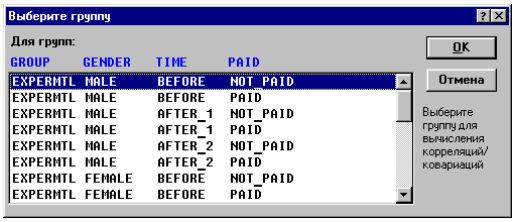
Для z – преобразованных значений корреляций строится полунормальный вероятностный график.

Матрица графиков рассеяния для всех переменных. Эта опция выводит на экран матрицу диаграмм рассеяния для всех зависимых переменных и ковариат плана.



Выбор группы для таблицы или графика

Это окно содержит все группы исследуемого межгруппового плана. Для выбора группы щелкните на ней мышью.



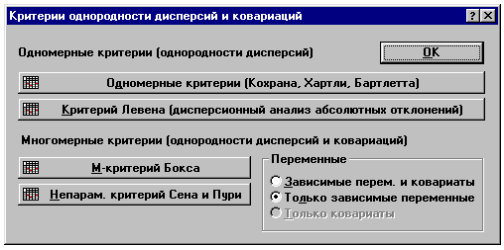
Если выбрана опция *Все группы*, то таблица или график будут построены для всех наблюдений (все группы объединяются).

Критерии однородности дисперсии

Одним из предположений одномерного дисперсионного анализа является равенство дисперсий (вариаций) по ячейкам межгруппового плана. В многомерном случае (*MANOVA*) это предположение применяется к ковариационной матрице зависимых переменных (и ковариат). Эти предположения и эффекты их нарушения обсуждены во *Вводном обзоре* (стр. 1730).

StatSoft®

Copyright © StatSoft, 1995



Одномерные критерии. Эта опция отображает на экране таблицу значений *критерия F-макс Хартли*, *С-критерия Кохрена* и *критерия хи-квадрат Бартлетта* (с соответствующими степенями свободы и *p*-уровнями). Все эти критерии исследуют предположения об однородности дисперсии (вариации) в одномерном случае. Если в плане имеется несколько независимых переменных или ковариат, то эти критерии выводятся на экран для каждой зависимой переменной и для каждой ковариаты. Все эти критерии описаны в большинстве стандартных учебников по дисперсионному анализу (например, см. Winer, 1962, стр. 94). Как описано во *Вводном обзоре* (стр. 1730), последствия даже довольно значительных отклонений от однородности дисперсий (вариаций) не очень опасны. Линдман (Lindman 1974, стр. 33), суммируя различные исследования этого вопроса, показал, что только при очень серьезных нарушениях предположения об однородности следует беспокоиться о надежности *F* статистики. Однако очень важно само по себе исследование корреляций между средними и переменными и дисперсиями (вариациями) (или стандартными отклонениями). Можно использовать опцию *График средних и стандартных отклонений,* чтобы увидеть эти корреляции.

Критерий Левена. Критерий Левена для проверки однородности дисперсий (вариаций) равносилен применению одномерного дисперсионного анализа к абсолютным отклонениям от среднего значения в ячейках. Смысл критерия в том, что самая большая дисперсия в ячейке соответствует большим абсолютным отклонениям в этой же ячейке. Этот критерий обсуждается в Milliken and Johnson (1984).

Многомерные критерии

М-критерий Бокса. Это многомерный критерий однородности дисперсий (вариаций) и ковариаций для нескольких зависимых переменных или ковариат. Используя зависимую кнопку *Переменные,* можно либо включать в этот критерий ковариаты, либо нет. М-критерий Бокса очень чувствителен к отклонениям от нормальности, поэтому к использованию этого критерия нужно относиться с осторожностью. Если критерий значим, то ковариационные матрицы в разных ячейках плана значимо отличаются друг от друга (см. Anderson, 1958). В этом случае, можно было бы, вероятно, исследовать уже неоднородность внутригрупповых ковариационных матриц, однако отклонения от предположений об однородности ковариационных матриц обычно не столь угрожающи.

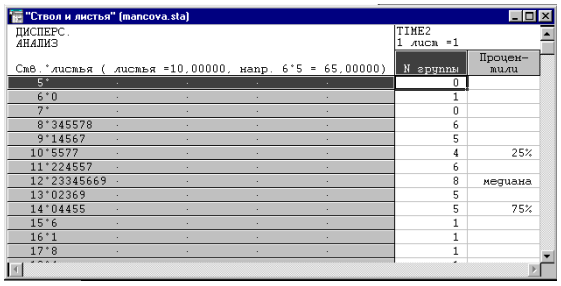
Непараметрический критерий Сена и Пури. Этот параметрический критерий совсем недавно приобрел популярность и является альтернативой М-критерию Бокса (Sen and Puri, 1968). Критерий основан только на информации о рангах наблюдений, и поэтому не подвержен влиянию отклонения от нормальности. Однако этот критерий также не очень чувствителен.

Сжатие и расширение графика ствол и листья

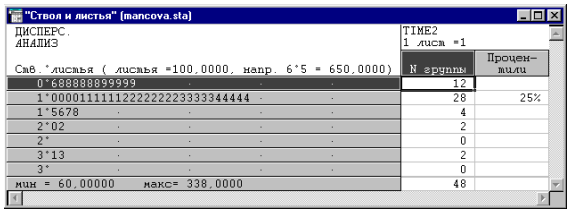
Эта опция изменяет число интервалов в графике ствол и листья. Если щелкнуть на кнопку *"Ствол и листья" для зависимых переменных и ковариат* в диалоговом окне *Описательные*

Copyright © StatSoft, 1995

статистики и графики, то появится график ствол и листья.

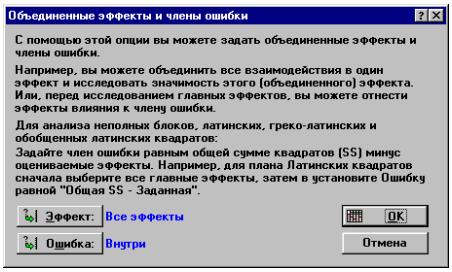


Если нажать на панели инструментов кнопку *Параметры* и выбрать опцию *Сжать/растянуть график*, график ствол и листья будет сжат в меньшее число интервалов (или если он уже был сжат, то будет расширен до первоначального числа интервалов), см. график ниже.



Объединенные эффекты и члены ошибки

Это диалоговое окно позволяет задать собственные эффекты и ошибки. В некоторых случаях требуется изучить все главные эффекты. Другая общая практика – относить к ошибкам взаимодействия высоких порядков, которые не имеют особого практического или теоретического интереса. С помощью этого метода могут анализироваться различные неполные планы, включая латинские квадраты (см. ниже). Заметим, что в этом разделе речь идет только о межгрупповых факторах.



Эффект

После нажатия на эту кнопку, появляется диалоговое окно *Межгрупповой эффект* (см. ниже) со всеми межгрупповыми эффектами плана. Выберите в этом диалоговом окне все эффекты, которые можно объединить и нажмите *OK*.

Ошибка

Эта кнопка позволяет переключиться на один из трех типов ошибки: *Внутри*, *Заданная (пользователем)*, и *Общая SS – Заданная*.

Внутри. При первой установке используется стандартное внутригрупповое значение ошибки для исследования объединенного эффекта, выбранного с помощью опции *Эффект*. Для выполнения вычислений нужно нажать *OK* в этом диалоговом окне (сразу после выбора объединенных эффектов).

Заданная. Чтобы определить пользовательское (объединенное) значение ошибки, нужно нажать на кнопку *Ошибка* (при установке *Внутри*). Откроется диалоговое окно *Межгрупповой эффект.* В этом окне нужно выбрать эффекты, которые будут отнесены к ошибке (например, взаимодействие трех факторов и внутренняя ошибка) точно так же, как делалось выше в опции *Эффект.* После того, как закроется это окно опция *Ошибка* примет значение *Заданная*. Для выполнения вычислений нужно нажать кнопку *OK* в этом окне.

StatSoft®
Copyright © StatSoft, 1995

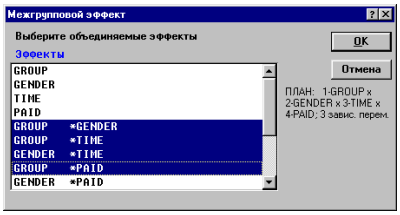
Общая *SS* - Заданная. Сразу после выбора эффектов в диалоговом окне, *Межгрупповые эффекты,* которые будут отнесены к значениям ошибок, нажмите на *OK,* выйдите из этого окна. После этого, нажмите на кнопку *Ошибки*, изменив установку на *Общая SS - Заданная*. При этой установке член ошибки, используемый в анализе – это общая сумма квадратов минус сумма квадратов, соответствующая эффектам, объединенным в пользовательскую ошибку (как выбрано в установке переключателя). Эта последняя установка переключателя особенно полезна при анализе планов типа латинских квадратов или рандомизированных блочных планов.

Латинские квадраты. При проведении анализа планов типа латинский квадрат, греко-латинский квадрат, (одномерных или многомерных), гипер-греко-латинский квадрат и т.п., сначала определяется план, как если бы он был полным факторным планом. Затем нужно перейти в диалоговое окно *Объединенные эффекты и члены ошибки* и выбрать изучаемые главные эффекты (с помощью опции *Эффект*). В качестве члена ошибок выбирается общая сумма квадратов минус сумма квадратов, соответствующая всем главным эффектам (нужно нажать на кнопку *Ошибка*, выбрать все эффекты, снова нажать на кнопку *Ошибка*, изменив член ошибки на *Общая SS - Заданная*). Полученные при этом результаты, будут соответствовать критериям значимости для главных эффектов.

Отметим, что всеобъемлющая реализация латинских, греко-латинских и гипер-греко-латинских квадратов возможна в модуле *Планирование эксперимента* (том IV).

Задание объединенных эффектов (объединение эффектов)

Выберите эффекты, которые объединяются.

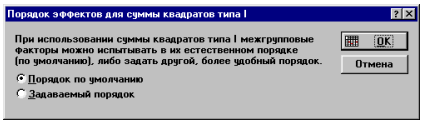


Для того, чтобы выделить эффект в этом диалоговом окне, щелкните на нем мышкой. Для того чтобы подсветить группу эффектов, нужно двигать мышку по этим эффектам, при нажатой левой кнопке мыши. Для того чтобы выбрать или отменить выбор эффекта, нужно нажать CTRL и щелкнуть мышкой на соответствующем эффекте.

Порядок эффектов в последовательных (тип I) суммах квадратов

Суммы квадратов *типа I* исследуют эффекты последовательно, подобно пошаговой многомерной регрессии (см. стр. 1724; см. также *Многомерная регрессия*, глава 12). Все эффекты, уже введенные в модель контролируются, а все не введенные игнорируются.

Если выбрать *суммы квадратов типа I* в диалоговом окне *Опции вывода* и затем нажать на кнопку *Все эффекты* в окне *Результаты,* откроется окно *Порядок эффектов для суммы квадратов типа I.*



Можно выбирать эффекты в порядке, заданном по умолчанию (нужно выбрать опцию *Порядок*

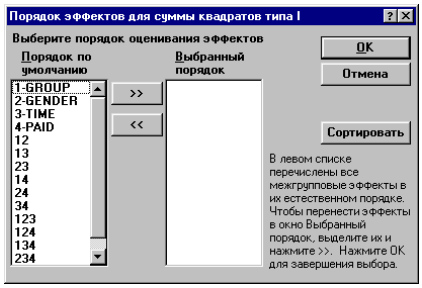
Copyright © StatSoft, 1995

StatSoft®

по умолчанию) или в порядке, заданном пользователем (опция *Задаваемый порядок*).

Переупорядочивание эффектов для сумм квадратов типа I

При выборе пользовательского порядка эффектов для сумм квадратов *типа I* в диалоговом окне *Порядок эффектов для сумм квадратов типа I* (см. выше) откроется диалоговое окно *Порядок эффектов для гипотез типа I.*

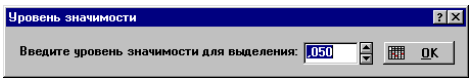


В этом диалоговом окне можно задать порядок, в котором эффекты исследуются (вводятся в модель). В левом столбце показаны эффекты в некотором порядке (по умолчанию). В этом столбце можно выбрать один или несколько эффектов стандартным способом, щелкнув по ним. После этого нажмите на кнопку >>,чтобы перенести выбранный эффект в правый столбец. В этом столбце эффекты располагаются в том порядке, в котором они будут исследоваться (вводиться в модель). Если в расположении эффектов в правом столбце была сделана ошибка, то можно выбросить некоторые эффекты из правого столбца.

Если не все эффекты были выбраны в правый столбец, то после нажатия на кнопку *OK*, все невыбранные эффекты будут включаться в модель после выбранных в порядке, заданном по умолчанию. Можно щелкнуть на кнопку *Сортировать* для того, чтобы упорядочить список эффектов в правом столбце в соответствии с порядком, заданным по умолчанию.

Уровень значимости, выделяемый в таблице всех эффектов

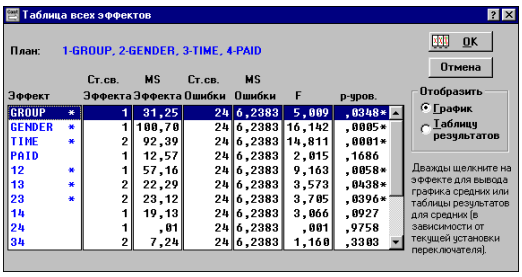
Эта опция позволяет задать уровень значимости (*альфа*) для выделения значимых эффектов в итоговой таблице.



По умолчанию значение *альфа* принято .05. Это значение используется для выделения статистически значимых значений (если *p*-значение меньше *альфа*, то соответствующий эффект выделяется как статистически значимый). В этом диалоговом окне можно изменить значение *альфа* (.0001 < *альфа* < .5) и, нажав на *OK,* сразу использовать измененное значение *альфа* в следующей копии текущей таблицы.

Таблицы и графики средних значений

Для того, чтобы построить график средних (график взаимодействий) или таблицу средних (в соответствие с установкой опции *Отобразить*), необходимо выделить в таблице главный эффект или взаимодействие и после этого нажать *OK,* для вывода графика или таблицы на экран.



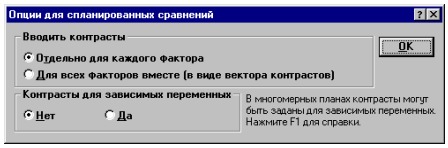
StatSoft®

Copyright © StatSoft, 1995

В это диалоговое окно можно вернуться, нажав кнопку *Далее* на таблице или графике выбранного эффекта.

Опции для спланированных сравнений

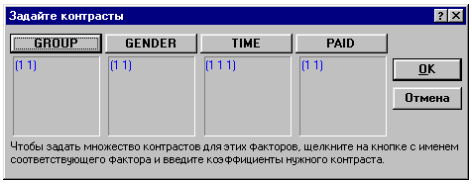
Контрасты можно вводить как для каждого фактора в отдельности, так и для всех факторов *вместе* (в виде *вектора* контрастов). В многомерном плане исследователь может также задать контрасты для зависимых переменных.



Для получения дополнительной информации о логике и интерпретации контрастов можно обратиться к разделу *Вводный обзор* (стр. 1723). Примеры задания контрастов обсуждаются также в разделе *Примечания* (стр. 1785), как для межгрупповых планов, так и для планов с повторными измерениями.

Ввод контрастов отдельно для каждого фактора

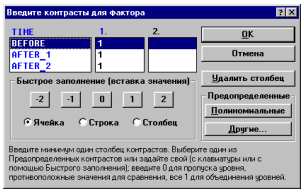
Если выбрать ввод контрастов отдельно для каждого фактора в диалоговом окне *Опции для спланированных сравнений* (см. выше), а затем нажать кнопку *Спланированные сравнения* в диалоговом окне *Результаты,* откроется окно *Задайте контрасты.* Для задания контрастов сначала в этом промежуточном окне, щелкните на нужном эффекте, а затем введите коэффициенты контрастов в диалоговом окне *Задайте контрасты для этого фактора*. Введенные контрасты показываются внизу под соответствующим фактором.



Для получения дополнительной информации о контрастах можно обратиться к *Вводному обзору* (стр. 1723). Примеры задания контрастов обсуждаются также в разделе *Примечания* (стр. 1785) как для межгрупповых планов, так и для планов с повторными измерениями.

Задание множества контрастов

Если нажать на кнопку с именем фактора в диалоговом окне *Задайте контрасты* (см. выше), откроется диалоговое окно *Задайте контрасты для этого фактора,* в котором можно задать контрасты для выбранного фактора.



Общие правила генерирования контрастных коэффициентов (подробнее о контрастном анализе см. *Вводный обзор*, стр. 1723 и *Примечания*, стр. 1785):

(1) Контрасты должны иметь столько же коэффициентов, сколько уровней у соответствующего фактора.

(2) Уровни, которые не учитываются в контрасте, имеют соответствующий коэффициент 0 (нуль).

(3) Уровни, которые сравниваются друг с другом, имеют коэффициенты

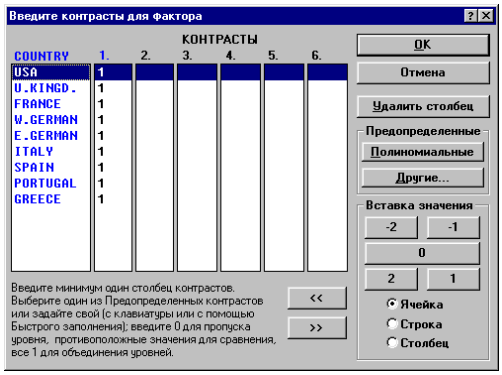
Copyright © StatSoft, 1995

противоположных знаков, сумма таких коэффициентов равна нулю.

(4) Уровни, которые объединяются, имеют равные коэффициенты (например, все равны 1).

Если выделенный фактор имеет более чем два уровня, то возможно использование некоторых предопределенных контрастов, включая контрастные коэффициенты для ортогональных полиномов (например, при анализе тренда).

Размер и расположение диалогового окна *Задайте контрасты для этого фактора* зависят от числа уровней соответствующего фактора (максимальное число векторов контрастов равно *k-1*, где *k* число уровней соответствующего фактора). Если факторы имеют много уровней, то можно просматривать одновременно и уровни и вектора контрастных коэффициентов.



Вставка значения

Для быстрого ввода контрастных коэффициентов служит опция *Вставка значения*. Если выбрать опцию *Ячейка* и нажать на кнопку с числом, то коэффициент в ячейке для выделенной переменной будет заменен этим числом (только в одной ячейке).

Если выбрана опция *Столбец*, то после нажатия на кнопку с числом, коэффициенты всех переменных будут заменены выбранным числом (столбец контрастных коэффициентов будет заменен этим числом).

При выборе опции *Строка* и последующем нажатии кнопки с числом, все коэффициенты переменной будут заменены соответствующим числом (вся строка контрастных коэффициентов будет заменена этим числом).

Предопределенные контрасты

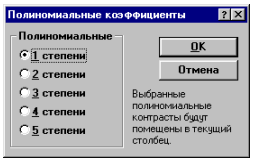
Кнопки *Полиномиальные* и *Другие* служат для выбора заранее определенных контрастов (см. ниже).

Типы предопределенных контрастов

После нажатия на кнопку *Полиномиальные* или на кнопку *Другие...* в диалоговом окне *Задайте контрасты для этого фактора* (см. выше), появляется либо диалоговое окно *Полиномиальные коэффициенты,* либо *Предопределенные контрасты* контрасты. В этих окнах можно выбрать предопределенное множество контрастов.

Полиномиальные коэффициенты

В этом окне выбирается степень полиномиальных контрастов (возможен выбор полинома до пятой степени).



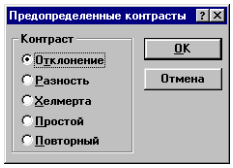
Для получения дополнительной информации об анализе полиномиальных контрастов см. *Примечания* стр. 1804.

StatSoft®

Copyright © StatSoft, 1995

Предопределенные контрасты

В этом диалоговом окне можно выбрать один из предопределенных типов контрастов.



Отклонение. Этот тип контраста сравнивает отклонения от общего среднего зависимой переменной. Например, для фактора с тремя уровнями будет создана следующая матрица контрастов.

```
 2 -1 -1
-1  2 -1
```

Разность. Этот тип контраста сравнивает уровни переменной со средними предыдущих уровней этой переменной. Например, для фактора с тремя уровнями будет создана следующая матрица контрастов.

```
-1  1  0
-1 -1  2
```

Хелмерта. Это контрастная матрица Хелмерта (Kirk, 1982). Например, для фактора с тремя уровнями будет создана следующая контрастная матрица.

```
1 -1 -1
0  2 -1
```

Простой. Этот тип контраста сравнивает каждый уровень с последним. Например, для фактора с тремя уровнями будет создана следующая контрастная матрица.

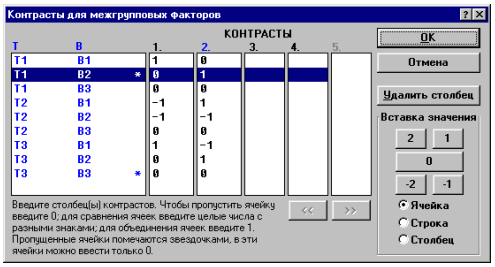
```
1  0 -1
0  1 -1
```

Повторный. Этот тип контраста сравнивает соседние уровни переменной. Например, для фактора с тремя уровнями будет создана следующая контрастная матрица.

```
1 -1  0
0  1 -1
```

Задание контрастных векторов

Если в диалоговом окне *Опции для спланированных сравнений* (см. выше), выбран ввод контрастов для всех факторов одновременно, то после нажатия на кнопку *Спланированные сравнения* в диалоговом окне *Результаты* откроется диалоговое окно *Контрасты для межгрупповых факторов.* В этом окне можно вводить контрастные коэффициенты для исследования уровней (ячеек межгрупповых или внутригрупповых планов). Заметим, что если в плане имеются пропущенные ячейки, то соответствующие строки помечены звездочкой, и в эти ячейки можно вводить только нули.



Общие правила для генерирования контрастных коэффициентов даны выше. Для более подробного изучения и интерпретации анализа контрастов см. *Вводный обзор* (стр. 1723) и *Примечания* (стр. 1785).

Copyright © StatSoft, 1995

StatSoft®

ПРИМЕЧАНИЯ, ЧАСТЬ I: ПРИМЕРЫ АНАЛИЗА ПЛАНОВ ОБЩЕГО ТИПА

Задание одномерных и многомерных межгрупповых планов

Однофакторный межгрупповой дисперсионный анализ и анализ контрастов

Резюме. *Межгрупповые планы* – это планы одномерного дисперсионного анализа с несколькими различными группами субъектов (или наблюдений). Например, в трех разных классах проводится обучение алгебры по трем разным учебникам, а затем на стандартных алгебраических тестах оценивается в некоторой шкале степень овладения предметом. Три разных класса образуют три разные группы, и для сравнения этих трех групп может быть использован однофакторный одномерный дисперсионный анализ.

Каждая группа в этом плане “демонстрирует” различие учебников по алгебре, поэтому тип учебника является *межгрупповым* фактором. Межгрупповой фактор содержит три уровня.

Задание файла данных. Файл данных в этом исследовании должен содержать, *по крайней мере, две переменные*. Одна переменная будет содержать оценки студентов, измеренные в некоторой шкале, по стандартным алгебраическим тестам. Эта переменная является *зависимой переменной*. Другая переменная будет содержать числовые или текстовые коды для задания групповой принадлежности каждого студента (каждого случая) в файле данных. Эта переменная является *независимой* или *группирующей* переменной.

Например, можно каждому студенту в группе обучающихся по первому учебнику поставить в соответствие текстовой код *Книга_1*, каждому студенту в группе обучающихся по второму учебнику поставить в соответствие код *Книга_2*, а всем студентам в третьей группе назначить код *Книга_3*. В результате файл данных может выглядеть так:

Независимая переменная	Зависимая переменная
Книга_1	23
Книга_3	48
Книга_3	26
Книга_2	27
Книга_1	43
Книга_1	39
Книга_2	30
...	...

В этом примере группирующая переменная содержит только три различных текстовых кода, которые будут введены в качестве кодов межгруппового фактора. Если в файле данных содержатся наблюдение с текстовым или числовым значением группирующей переменной, которые не содержатся в списке кодов, то они будут исключены из анализа.

Заметим, что при вводе данных в файл не важен порядок, в котором они вводятся.

Анализ плана. Для анализа этих данных нужно сначала нажать на кнопку *Переменные* в стартовой панели *Дисперсионный анализ* (стр. 1757) и задать в качестве межгруппового фактора или *независимой переменной* переменную с номером 1. Переменная *2* будет определена как *зависимая* переменная. После этого с помощью кнопки *Коды* задается независимая переменная, содержащая коды, как указывалось выше.

StatSoft®

Copyright © StatSoft, 1995

Заметим, что вовсе не обязательно вводить все коды в соответствующем окне. Можно просто нажать на *OK* и программа выберет по умолчанию все коды. После выбора переменных и кодов, нужно нажать на кнопку *OK* в стартовой панели и появится диалоговое окно *Результаты.*

Исследование спланированных сравнений. *Спланированные сравнения* полезны, если нужно сравнивать отдельные группы плана. Для этого нужно ввести *множество контрастов,* которые задают нужные сравнения. Если в настоящем примере желательно сравнить группу, обучающихся по первому учебнику (группу 1), с группой, обучающихся по третьему учебнику (группу 3), при этом вторая группа не интересует исследователя, то должно быть введено множество контрастных коэффициентов: *1*, *0*, *-1*.

Правила формирования контрастов. Правила формирования контрастов состоят в следующем:

(1) Контрасты должны иметь столько же коэффициентов, сколько уровней у соответствующего фактора.

(2) Группы, не участвующие в сравнении, имеют контрастные коэффициенты равные 0 (нулю).

(3) Группы, участвующие в сравнении друг с другом, имеют в качестве контрастных коэффициентов отрицательные или положительные целые значения, но главное, чтобы сумма этих коэффициентов была *равна нулю*.

(4) Группы, которые объединяются в анализе, имеют одинаковые контрастные коэффициенты (например, *1*).

В нашем примере, если нужно сравнить группу 1 с объединением групп 2 и 3, то множество контрастных коэффициентов может быть следующим: *2*, *-1*, *-1*. Если нужно сравнить группу 2 с объединением групп 1 и 3, то подходящими контрастными коэффициентами будут: *1*, *-2*, *1*. Если нужно сравнить группу 2 с группой 3 (группа 2 игнорируется), подходящими контрастными коэффициентами будут *0*, *1* и *-1*.

Многомерный дисперсионный анализ. Пока обсуждение было ограниченно одномерным дисперсионным анализом. Однако если выбрано несколько зависимых переменных в стартовом диалоговом окне модуля *Дисперсионный анализ*, то автоматически будет выполняться многомерный дисперсионный анализ (если только многомерные зависимые переменные не определяются как уровни факторов повторных измерений).

Если, в рассматриваемом примере, знания алгебры оценивались не одним, а тремя тестами, то файл данных выглядит так:

Независимая Переменная	Зависимые переменные 1	2	3
Книга_1	23	24	28
Книга_3	48	43	40
Книга_3	26	38	21
Книга_2	27	29	35
Книга_1	43	42	42
Книга_1	39	40	31
Книга_2	30	29	27
...	...	...	...

Для того чтобы был выполнен многомерный дисперсионный анализ (*MANOVA*) с тремя переменными, нужно ввести *все три переменные* в список зависимых переменных. Теперь при исследовании главных эффектов или спланированных сравнений будут вычисляться и многомерные критерии (для всех зависимых переменных одновременно) и одномерные (когда исследуется одна переменная).

Способ, каким задаются спланированные сравнения почти тот же, что и в одномерном случае. Однако при введении контрастных коэффициентов имеется возможность задавать

Copyright © StatSoft, 1995

StatSoft®

контрастные коэффициенты и для зависимых переменных. Для этого нужно нажать на кнопку *Опции,* которая следует за кнопкой *Спланированные сравнения* (см. диалоговое окно *Результаты,* стр. 1763). Если затем выбрать *Контрасты для зависимых переменных*, то можно будет задавать контрасты для зависимых переменных. Некоторые зависимые переменные могут при этом опускаться из сравнений, если соответствующие контрастные коэффициенты равны 0 (нулю).

Многофакторный межгрупповой дисперсионный анализ и анализ контрастов

Резюме. Если план содержит более чем один межгрупповой фактор, то этот план называется *многофакторным межгрупповым* планом. Например, как и раньше, можно сравнивать эффективность преподавания по трем разным учебникам (межгрупповой фактор номер 1) и, кроме того, ввести еще один фактор – пол студента (межгрупповой фактор номер 2). Зависимой переменной являются замеренные в некоторой шкале оценки по стандартному алгебраическому тесту.

Задание файла данных. Теперь в файле данных к зависимой переменной (оценкам по стандартному алгебраическому тесту) добавлены две *независимые или группирующие* переменные. Одна независимая или группирующая переменная содержит текстовые или числовые коды, которые однозначно определяют группу, в которой обучался студент (какой учебник использовался при обучении). Другая независимая или группирующая переменная содержит числовые или текстовые коды, определяющие пол студента. Файл данных может выглядеть следующим образом:

Независимые переменные		Зависимая
Текст	Пол	переменная
Книга_1	Муж	23
Книга_3	Жен	48
Книга_3	Муж	26
Книга_2	Жен	27
Книга_1	Муж	43
Книга_1	Муж	39
Книга_2	Жен	30
...	...	. ..

Первая переменная (первый столбец) содержит значения *Книга_1, Книга_2* и *Книга_3*. Эта независимая или группирующая переменная определяет, по какой книге обучался студент. Вторая переменная (второй столбец) содержит только значения *Муж* и *Жен.* Эта независимая или группирующая переменная определяет пол студента. Третий столбец – это оценки по стандартному алгебраическому тесту (замеренные в некоторой шкале).

Задание плана. Этот план будет задан как 3 (учебники) x 2(пол) межгрупповой план. В стартовой панели (стр. 1757), нужно задать переменные 1 и 2 как независимые переменные (межгрупповые факторы) и переменную 3 как зависимую переменную.

Исследование главных эффектов и взаимодействий. В этом плане имеются три типа эффектов, которые можно исследовать. Можно исследовать главные эффекты для разных учебников, главные эффекты для разных полов, и эффекты взаимодействия между типом учебника и полом студента. Если использовать опцию *Все эффекты* или опцию *Средние/графики* из диалогового окна *Результаты* (стр. 1763), то будут подсчитаны все эффекты. Для более детального исследования эффектов нужно воспользоваться опцией *Специфический эффект/средние* из диалогового окна результаты и задать нужные эффекты.

Исследование спланированных сравнений. Спланированные сравнения задаются точно так же, как и в однофакторном

StatSoft®

Copyright © StatSoft, 1995

дисперсионном анализе. Различие лишь в том, что множество контрастов задается не для одного, а для каждого из факторов, входящих в план.

К правилам задания контрастов, которые были введены при обсуждении планов однофакторного дисперсионного анализа (стр. 1785), нужно добавить, что в данном случае возможно (хотя и необязательно) вводить контрастные коэффициенты, сумма которых не равна 0 (нулю). Например, для того, чтобы задать множество контрастов, исследование которых эквивалентно исследованию главных эффектов фактора *Пол,* нужно задать следующее множество контрастов:

Для фактора 1(тип учебника):

```
1    1    1
```

Для фактора 2(пол):

```
1   -1
```

Тогда все уровни первого фактора (тип учебника) имеют одинаковый вес (все уровни объединяются в один), в то время как контрастные коэффициенты второго фактора сопоставляют два уровня (мужчин и женщин).

Можно задать довольно много различных сравнений. Например, можно исследовать главные эффекты фактора 1 (тип учебника) *только для мужчин* (уровень 1 фактора *Пол)*. Для этого нужно ввести следующее множество контрастов:

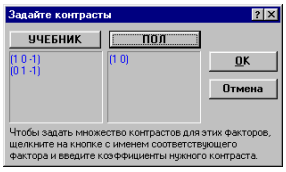
Для фактора 1 (тип учебника) два множества контрастов:

```
1     0    -1
0     1    -1
```

Для фактора 2 (пол):

```
1     0
```

Диалоговое окно *Задайте контрасты* выглядит при этом следующим образом:



Для фактора 1 введены два независимых контраста, которые будут исследоваться *одновременно*. Эти два контраста представляют полное множество контрастов, в том смысле, что они сравнивают одновременно все уровни фактора 1 (универсальное множество контрастов). Для второго фактора коэффициент, соответствующий второму уровню фактора (женщины), равен 0 (нулю). Это означает, что женщины будут исключаться из исследования, и главные эффекты первого фактора будут исследоваться *только для мужчин*.

Аналогичным способом можно определять и более сложные контрасты в планах трехфакторного, четырехфакторного и т. п. дисперсионного анализа (например, можно изучать внутри уровня одного фактора эффект взаимодействия других факторов). Для более подробной информации о сложных контрастах и о контрастных коэффициентах (как изучать простые эффекты, линейные и нелинейные эффекты и взаимодействия) можно обратиться к специальным учебникам по дисперсионному анализу (например, Lindman, 1974; Keppel, 1973; Winer, 1962, 1971). Заметим также, что в диалоговом окне, в котором вводятся контрастные коэффициенты, содержатся опции, позволяющие использовать предопределенные контрасты. В частности, важной является опция *Полиномиальные* контрасты (стр. 1782), которая позволяет исследовать линейный и нелинейный тренды по уровням соответствующего фактора.

Copyright © StatSoft, 1995
StatSoft®

Многомерный многофакторный дисперсионный анализ (M*ANOVA*). Задание плана для многофакторного межгруппового дисперсионного анализа с несколькими зависимыми переменными (M*ANOVA*) проводится так же, как и для одномерного анализа. Должны быть заданы независимые или группирующие переменные и соответствующие коды, однозначно определяющие групповую принадлежность каждого наблюдения в файле данных. Если задано несколько зависимых переменных, то автоматически будет выполнен многомерный дисперсионный анализ (за исключением случая, когда зависимые переменные могут быть проинтерпретированы как уровни факторов повторных измерений). По умолчанию предполагается, что одномерные тесты также проводятся для каждой зависимой переменной.

Способ задания спланированных сравнений практически аналогичен одномерному случаю. Однако при вводе контрастных коэффициентов для спланированного сравнения у пользователя имеется возможность ввода контрастов для зависимых переменных. Нужно нажать на кнопку *Опции* (она следует за кнопкой *Спланированные сравнения*; см. диалоговое окно *Результаты* на стр. 1763) и выбрать кнопку *Контрасты для зависимых переменных*. После этого программа позволит задавать контрасты для зависимых переменных. Например, пользователь может выбирать зависимые переменные, которые будут игнорироваться при сравнении, с помощью приписывания им *0* (нулей) при задании контрастов.

Задание кодов для изолированной контрольной группы

Иногда экспериментальные планы нуждаются во включении контрольной группы, которая не подвергнута ни какой обработке. Например, представим себе экспериментальный тест эффективности двух различных лекарств (фактор 1) при заражении двумя различными типами вируса (фактор 2). В эксперименте две группы лабораторных животных были заражены одним из двух типов вирусов, к ним было применено лечение одним из двух типов лекарств. Первичной зависимой переменной является поведение животных во время обучения. То есть, сколько потребуется попыток, чтобы животное научилось точно отделять различные геометрические формы.

В этом эксперименте можно сравнивать обучаемость животных при разных экспериментальных условиях с обучаемостью животных, которые не были заражены никаким вирусом, и которые не подвергались действию никаких лекарств. Без этого сравнения эксперимент описывается как 2 (тип лекарства) x 2 (тип вируса) экспериментальный план. Добавление в этот экспериментальный план группы животных, которые не соответствуют никакому уровню ни одной из независимых переменных, означает, что добавлена *изолированная контрольная группа.* Этот план будет описываться как 2 x 2 план с изолированной контрольной группой.

Установка файла данных. План, содержащий экспериментальные данные в этом примере, может быть описан следующим образом (заметим, что для простоты в качестве кодов используются лишь числовые значения):

Независимые переменные		Зависимая
Лекарство	Вирус	переменная
1	1	45
2	1	29
9	9	35
2	2	38
2	1	20
9	9	26
1	1	25
2	1	39
1	2	37
1	2	34
9	9	25

Первая переменная (столбец 1) содержит коды, которые определяют тип использованного

StatSoft®

Copyright © StatSoft, 1995

лекарства (коды 1 и 2). Вторая переменная (столбец 2) содержит коды, определяющие один из типов вирусов, которым животное было инфицировано (коды 1и 2). *Обе переменные* содержат кодовое число 9, которое указывает, что соответствующее животное принадлежит контрольной группе.

Во всех случаях изолированная контрольная группа должна задаваться единственным кодовым числом, которое входит во *все независимые переменные*, для указания того, что соответствующее наблюдение принадлежит изолированной контрольной группе.

Задание плана. Когда для этого примера задается план в стартовой панели (стр. 1757), то сначала задается, как обычно 2 x 2 межгрупповой план. Для независимых переменных должны быть введены коды 1 и 2. Затем нужно нажать на кнопку *Изолированная контрольная группа* для задания числового кода, который будет определять принадлежность к изолированной контрольной группе (в нашем примере, это
число 9).

Исследование главных эффектов и взаимодействий будет проходить как для обычного простого 2 x 2 плана дисперсионного анализа. Но сумма квадратов ошибок (и степень свободы) будет вычисляться и для всех четырех групп и для изолированной группы (т. е. суммарная внутригрупповая дисперсия вычисляется для *всех* групп).

Для того чтобы подсчитать среднее (средние) контрольной группы нужно воспользоваться опцией *Специфический эффект* в диалоговом окне *Результаты* (стр. 1763) и выбрать исследование взаимодействий самого высокого порядка (в нашем примере взаимодействие второго порядка). После вывода на экран таблицы для части плана с полностью пересекающимися факторами, программа выведет на экран среднее (средние) для изолированной контрольной группы.

Спланированные сравнения Для оценки различия между контрольной группой и другими группами плана нужно ввести подходящее множество контрастных *векторов,* т. е. прежде, чем вводить контрасты отдельно для каждого фактора нужно ввести контрастные коэффициенты для всех групп *вместе*. Нужно нажать на кнопку *Опции* расположенную за кнопкой *Спланированные сравнения* (см. диалоговое окно *Результаты*, стр. 1763), и в открывшемся диалоговом окне установить зависимую кнопку в положение *Для всех факторов вместе (в виде вектора контрастов)*. Затем нужно выбрать *Спланированные сравнения* и в открывшемся окне вводить контрастные коэффициенты. Последним вводится коэффициент для контрольной группы.

Если в нашем примере нужно сравнить изолированную контрольную группу со всеми остальными группами плана, соответствующее множество контрастов (введенное либо как один вектор, либо как множество, т. е. все коэффициенты введены вместе) должно быть таким:

```
Фактор 1, уровень:        1  1  2  2
Фактор 2, уровень:        1  2  1  2
                          ------------
Введите эти
коэффициенты контрастов: 1  1  1  1  -4
```

Если нужно сравнить обучаемость животных, подвергнутых лечением первым лекарством (уровень 1 первой независимой переменной) с обучаемостью контрольной группы, то множество коэффициентов должно выглядеть примерно так:

```
Фактор 1, уровень:        1  1  2  2
Фактор 2, уровень:        1  2  1  2
                          ------------
Введите эти
коэффициенты контрастов: 1  1  0  0  -2
```


Copyright © StatSoft, 1995

StatSoft®

Если нужно сравнить обучаемость животных, инфицированных первым вирусом (уровень 1 второй независимой переменной) и подвергнутых лечению вторым лекарством (уровень 2 первой независимой переменной) с обучаемостью животных в контрольной изолированной группе, то множество контрастных коэффициентов может выглядеть примерно так:

```
Фактор 1,уровень:          1  1  2  2
Фактор 2,уровень:          1  2  1  2
                           -------------
Ввести эти
Коэффициенты контрастов: 0  0  1  0  -1
```

В общем случае, когда все контрастные коэффициенты вводятся одновременно, программа будет назначать коэффициенты последовательно всем ячейкам (уровням факторов). При этом быстрее всего будут меняться уровни фактора, который был введен последним, следующим по скорости изменения уровней фактора будет фактор, который был введен предпоследним, и т. д. Коэффициент, который относится к изолированной контрольной группе, всегда вводится последним.

Задание внутригрупповых (с повторными измерениями) одномерных и многомерных планов

Однофакторный внутригрупповой (с повторными измерениями) дисперсионный анализ

Довольно часто нужно проводить исследования на одних и тех же субъектах или через определенные промежутки времени или при разных условиях. В частности, можно исследовать вопрос об изменениях, происходящих с каждым субъектом, насколько субъект “достигает совершенства” в чем-либо со временем. Такие планы относятся к *внутригрупповым* или к *планам с повторными измерениями*.

Например, предположим, что нужно проконтролировать, как повышается уровень знаний по алгебре в течение трех месяцев обучения. Для контроля давались стандартные алгебраические тесты после первого месяца обучения (уровень 1 фактора повторных измерений), после двух месяцев обучения (уровень 2 фактора повторных измерений) и, затем, после трех месяцев обучения (уровень 3 фактора повторных измерений). Таким образом, имеется фактор повторных измерений (*Time)* с тремя уровнями.

Задание файла данных. Для анализа таких данных используется внутригрупповой (с повторными измерениями) дисперсионный анализ. Сначала должен быть создан файл данных, содержащий оценки для всех студентов, который, который будет выглядеть следующим образом:

	Контрольный тест после 1 месяца	2 месяцев	3 месяцев
Наблюдение 1	24	26	28
Наблюдение 2	30	30	32
Наблюдение 3	30	29	28
Наблюдение 4	30	29	34
Наблюдение 5	35	36	32
...	..	..	..

Первая переменная (столбец 1) содержат оценки (баллы) студентов по алгебре после одного месяца обучения, вторая переменная (столбец 2) содержит оценки студентов после двух месяцев обучения и третья переменная (столбец 3) содержит соответствующие оценки после трех месяцев обучения.

Задание плана. Для того чтобы анализировать эти данные, нужно сначала записать переменные 1 – 3 в списке зависимых переменных (оставив список независимых переменных пустым). После этого нужно задать план повторных измерений. Для этого нужно нажать на кнопку *План*

StatSoft®

Copyright © StatSoft, 1995

повторных измерений и задать один фактор повторных измерений с тремя уровнями (в нашем примере по умолчанию).

Исследование эффектов. Эффекты, включающие фактор повторных измерений, исследуются точно так же, как и эффекты в межгрупповых планах дисперсионного анализа. Если фактор повторных измерений имеет более 2 уровней, существуют два альтернативных способа оценки значимости эффектов, включающих этот фактор. Традиционный способ (см. Winer, 1962, 1971) состоит в использовании одномерных критериев. В последнее время стал широко использоваться многомерный дисперсионный анализ при исследовании таких планов. Преимущество многомерного подхода обсуждалось в разделе *Вводный обзор* (см. на стр. 1720 и 1790). Кратко можно сказать, что многомерный подход основан на менее ограничительных предположениях. На практике эти два подхода обычно приводят к одним и тем же результатам, кроме тех случаев, когда изменения (разности) по уровням фактора повторных измерений коррелированы внутри группы. В рассматриваемом примере успехи студентов между 1 и 2 месяцем обучения коррелированы с успехами между 2 и 3 месяцем обучения.

В любом случае следует с осторожностью относиться к одномерным и многомерным результатам. Второй пример этого раздела (стр. 1744) содержит данные с повторными измерениями, для которых многомерный и одномерный подходы дисперсионного анализа приводят к несогласованным результатам. Там же обсуждается источник этой несогласованности.

Исследование спланированных сравнений. Исследовать спланированные сравнения различных уровней фактора повторных измерений можно точно так же, как и спланированные сравнения в межгрупповых планах. После выбора опции *Спланированные сравнения* из диалогового окна *Результаты* нужно будет ввести множество контрастных коэффициентов.

Правила генерирования контрастных коэффициентов следующие:

(1) Число контрастных коэффициентов должно быть равно числу уровней соответствующего фактора.

(2) Уровни, которые опускаются при сравнении, имеют соответствующий коэффициент 0 (нуль).

(3) Уровни, которые сравниваются друг с другом, имеют своими коэффициентами положительные или отрицательные целые значения, сумма которых равна нулю.

(4) Коэффициенты уровней, которые объединяются при сравнении, имеют одинаковые целые значения (например, 1).

Вернемся к нашему примеру. Если нужно сравнить степень освоения студентами алгебры после одного месяца обучения с результатами после трех месяцев обучения, то могут быть использованы следующие контрастные коэффициенты:

```
1    0    -1
```

Если нужно сравнить результаты после трех месяцев обучения с результатами после одного и после двух месяцев обучения, то нужно ввести следующие коэффициенты:

```
-1    -1    2
```

В *Примере 2* раздела *Примеры* (стр. 1744) показано, как задаются и интерпретируются контрастные коэффициенты.

Copyright © StatSoft, 1995

StatSoft®

Однофакторный внутригрупповой дисперсионный анализ с многомерными повторными измерениями (*MANOVA*)

Если в примере, описанном выше, оцениваются два различных алгебраических теста после каждого месяца обучения, то естественно использовать многомерный дисперсионный анализ для повторных измерений.

Задание файла данных. Файл данных может быть организован следующим образом:

Контрольный тест после					
1 Месяца		2 Месяцев		3 Месяцев	
Тест		Тест		Тест	
1	2	1	2	1	2
24	30	26	34	28	28
30	32	30	28	32	34
30	27	29	28	28	33
30	28	29	26	34	35
35	25	36	27	32	36
..	..	..	..	..	..

В этом примере первая переменная (столбец 1) содержит оценки первого теста после первого месяца обучения, вторая переменная (столбец 2) содержит оценки второго теста после первого месяца обучения. Третья переменная содержит оценки по первому тесту после двух месяцев обучения, четвертая переменная содержит оценки по второму тесту после двух месяцев обучения и т. д.

Задание плана. Для того чтобы исследовать эти данные с помощью однофакторного многомерного (два теста) дисперсионного анализа с повторными измерениями нужно все переменные (от 1 до 6) выбрать как зависимые переменные (с помощью кнопки *Переменные* в стартовой панели (стр. 1757)). Затем нужно нажать кнопку *План повторных измерений*, и задать один фактор повторных измерений с тремя уровнями, как и раньше.

Программа циклически проходит по списку зависимых переменных, пытаясь разделить его поровну. Отнесение переменных к уровням фактора повторных измерений определяется *порядком, в котором зависимые переменные входят в список зависимых переменных*, а не порядком, в котором они содержатся в файле данных. Например, предположим, что в нашем примере данные были введены в файл следующим образом:

Тест 1			Тест 2		
Контроль после месяца			Контроль после месяца		
1	2	3	1	2	3
24	26	28	30	34	28
30	30	32	32	28	34
30	29	28	27	28	33
30	29	34	28	26	35
35	36	32	25	27	36
..	..	..	..	..	..

Теперь данные введены так, что первая переменная содержит оценки по первому тесту после первого месяца обучения, вторая переменная содержит оценки по первому тесту после двух месяцев обучения, третья переменная содержит оценки по первому тесту после трех месяцев обучения, четвертая переменная содержит оценки по второму тесту после первого месяца обучения, пятая переменная содержит оценки по второму тесту после двух месяцев обучения и, наконец, шестая переменная содержит оценки по второму тесту после трех месяцев обучения. При задании списка зависимых переменных, нужно переупорядочить переменные, то есть в лист зависимых переменных нужно ввести переменные *1 4 2 5 3 6* именно в таком порядке. Затем внутригрупповой план определяется как раньше.

Многофакторные внутригрупповые (с повторными измерениями) планы дисперсионного анализа

Резюме. В экспериментальных планах довольно часто включается более одного фактора

StatSoft®

Copyright © StatSoft, 1995

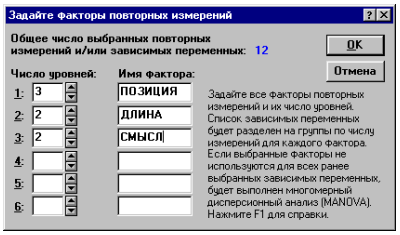
повторных измерений. В *Примере 3* в разделе *Примеры* (стр. 1751) описан как раз такой случай. Например, в этом эксперименте можно сравнивать, как запоминались слова из начала списка, из середины и из конца (первый фактор повторных измерений с тремя уровнями). Можно было бы сравнить, как запоминались слова, содержащие четыре или пять букв (второй фактор повторных измерений с двумя уровнями). Наконец, слова могли либо иметь смысл в английском языке либо нет (третий фактор повторных измерений с двумя уровнями). В результате имеется 3 (расположение слов) x 2 (длина слова) x 2 (осмысленные либо бессмысленные) план однофакторного дисперсионного анализа, все факторы которого – факторы повторных измерений.

Задание файла данных. Предположим, что файл данных выглядит следующим образом:

Факторы повторных измерений:	Уровень:											
1 (Расположение)	1				2				3			
2 (Длина)	1		2		1		2		1		2	
3 (Значение)	1	2	1	2	1	2	1	2	1	2	1	2
Субъект 1	9	4	3	4	3	4	5	4	3	4	5	4
Субъект 2	7	4	3	4	6	4	3	7	6	5	4	3
Субъект 3	6	5	7	8	7	6	8	5	6	4	5	2
.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.

Первая переменная содержит данные для первого уровня всех факторов повторных измерений. Вторая переменная содержит данные для второго уровня третьего фактора повторных измерений и первого уровня первого и второго фактора повторных измерений и т.д.

Задание плана. Для того чтобы задать план в этом примере, сначала нужно все переменные внести в список зависимых переменных в том самом порядке, в котором они записаны в файле данных (в этом примере переменные не переупорядочиваются), т.е. нужно ввести 1-12 в окно для списка зависимых переменных. Затем после нажатия кнопки *План повторных измерений* нужно задать факторы повторных измерений следующим образом. Сначала нужно задать фактор *Расположение* с тремя уровнями, затем фактор *Длина* с двумя уровнями и затем фактор *Значение* с двумя уровнями.



Программа будет относить значение переменных к уровням факторов повторных измерений следующим образом. Программа будет циклически проходить по списку зависимых переменных, и назначать переменным последовательно уровни факторов повторных измерений. При этой процедуре быстрее всего будут изменяться уровни фактора, который был введен последним, следующим по скорости изменения уровней идет фактор введенный предпоследним и т.д.

Когда программа приписывает каждому значению уровни факторов повторных измерений, то при этом важно, в каком порядке эти зависимые переменные расположены в списке зависимых переменных (не важно, в каком порядке эти зависимые переменные записаны в файле данных).

Например, если задан 2 x 2 x 3 план повторных измерений дисперсионного анализа, программа будет назначать зависимым переменным уровни факторов повторных измерений следующим образом:

Copyright © StatSoft, 1995
StatSoft®

```
Факторы                    Уровень:
повторных
измерений:
---------------------------------------
1 Фактор:        1                2
              ---------        ----------
2 Фактор:    1        2       1         2
            -----    -----    -----    --------
3 Фактор:  1 2 3    1 2 3    1 2 3    1  2  3
           - - -    - - -    - - -    -- -- --
Положение
В списке
зависимых
перемен.:  1 2 3    4 5 6    7 8 9    10 11 12
```

Заметим, что последняя строка определяет положение зависимой переменной в списке зависимых переменных (а не положение зависимой переменой в файле данных). Следовательно, при задании сложных многофакторных внутригрупповых планов, порядок, в котором каждой зависимой переменной присваиваются уровни факторов повторных измерений, определяется двумя моментами: (1) порядком, в котором задаются факторы повторных измерений и (2) порядком, в котором переменные заданы в списке зависимых переменных.

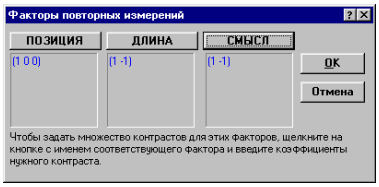
Исследование специфических эффектов. Специфические эффекты (опция *Специфический эффект/средние* в диалоговом окне *Результаты*) задаются обычным способом. Вернемся к нашему примеру. Чтобы задать взаимодействие всех трех факторов нужно нажать на кнопку *Специфический эффект/средние/графики* и выбрать все три фактора. Для того чтобы задать взаимодействие двух факторов (первого и третьего) нужно выбрать эти два фактора. Для исследования главного эффекта второго фактора нужно выбрать только этот фактор и т.д. Как и раньше, когда оцениваются эффекты, включающие факторы повторных измерений с более чем двумя уровнями, пользователь может выбирать, вычислять ли одномерный или многомерный критерий, или и тот и другой (при этом нужно использовать *Опции вывода*, для вывода нужных результатов).

Исследование спланированных сравнений. Спланированные сравнения можно осуществлять точно тем же способом, что и для межгрупповых планов. То есть нужно ввести множество контрастных коэффициентов. Обычно, легче вводить эти контрастные коэффициенты для каждого фактора отдельно, а не для всех эффектов вместе.

Правила введения контрастных коэффициентов для многофакторного плана повторных измерений, в общем, те же, что и для однофакторных планов повторных измерений. Однако здесь допускается (часто это и необходимо) введение контрастных коэффициентов, сумма которых не равна нулю (0). Вернемся к примеру 3 x 2 x 2 плана повторных измерений. Если нужно оценить взаимодействие второго и третьего факторов, но только на первом уровне первого фактора, могут быть введены следующие коэффициенты:

```
Фактор 1 (Расположение):  1  0  0
Фактор 2 (Длина)       :  1 -1
Фактор 3 (Смысл)       :  1 -1
```

Появятся следующие контрасты:



В этом случае, второй и третий уровни первого фактора повторных измерений не рассматривается, и взаимодействие между 2 и 3 фактором будет исследоваться только для первого уровня первого фактора повторных измерений. Если нужно исследовать главный эффект фактора *Смысл* (третий фактор) внутри первого уровня первого фактора, то должны быть введены следующие коэффициенты:

```
Фактор 1 (Расположение):  1  0  0
Фактор 2 (Длина)       :  1  1
Фактор 3 (Смысл)       :  1 -1
```


StatSoft®

Copyright © StatSoft, 1995

Вновь второй и третий уровни первого фактора повторных измерений игнорируются. Первый и второй уровни второго фактора повторных измерений берутся с равными весами (объединяются), а два уровня третьего фактора повторных измерений сравниваются друг с другом.

Если нужно оценить взаимодействие первых двух факторов с первым уровнем третьего фактора, то соответствующие коэффициенты выглядят так:

```
Фактор 1 (Расположение):  1  0 -1
и                      :  0  1 -1
Фактор 2 (Длина)       :  1 -1
Фактор 3 (Смысл)       :  1  0
```

Заметим, что для первого фактора задано два множества контрастных коэффициентов (обобщенное множество контрастов) для двух независимых сравнений. Детальное обсуждение анализа контрастов и независимости (ортогональности) выходит за рамки настоящего введения. Пользователь может обратиться к учебникам по одномерному дисперсионному анализу (например, Lindman, 1978; Winer, 1962, 1971) для информации как установить контрасты для проверки специфических гипотез (например, линейного или нелинейного тренда, с равными или неравными интервалами между наблюдениями и т.д.). Заметим, что диалоговое окно для задания контрастных коэффициентов содержит опции для предопределенных контрастов, в частности, так называемых, *Полиномиальных* контрастов (стр.1782), которые позволяют исследовать линейные и нелинейные тренды по уровням соответствующего фактора.

Многомерные многофакторные внутригрупповые (с повторными измерениями) планы

Как описывалось ранее, программа будет циклически проходить по списку зависимых переменных и назначать уровни факторов повторных измерений. Если в списке зависимых переменных число переменных больше общего числа уровней всех факторов повторных измерений, то автоматически будет выполняться многомерный дисперсионный анализ (см. также стр. 1760). Программа будет назначать соответствующему числу зависимых переменных последовательно уровни факторов повторных измерений. Это число зависимых переменных равно частному от деления числа зависимых переменных в списке зависимых переменных на произведение числа уровней всех факторов повторных измерений.

Например, если нужно задать 2 x 3 план повторных измерений дисперсионного анализа с 2 зависимыми переменными, то программа разбросает переменные по уровням следующим образом:

```
Фактор
повторных                уровень:
измерений:
-------- ------------------------------
Первый:          1                  2
         --------------  --------------
Второй:    1    2    3     1     2    3
          ---  ---  ---   ---  ---- -----
Зав. Пер. 1 2  1 2  1 2   1 2  1 2  1  2
          - -  - -  - -   - -  - -- -- --
Положение
в списке
зависимых
перемен.: 1 2  3 4  5 6   7 8  9 10 11 12
```

Снова заметим, что здесь важен порядок в списке зависимых переменных, а не в файле данных.

Меж- и внутригрупповые одномерные и многомерные планы

Резюме. В экспериментальном исследовании часто назначают различные препараты различным группами субъектов (межгрупповые факторы). Затем проводится мониторинг зависимой переменной по времени или по каким-либо другим условиям (внутригрупповой фактор или фактор повторных измерений). Этот тип

Copyright © StatSoft, 1995

плана сочетает в себе факторы повторных измерений и межгрупповые факторы.

Например, исследователь может заразить лабораторных животных двумя типами вирусов (межгрупповой фактор 1 с двумя уровнями), а затем снабдить их двумя типами различных лекарств (межгрупповой фактор 2 с двумя уровнями). Затем исследователь может наблюдать активность животных после 1 дня, затем после 2,3 дней (фактор повторных измерений *Время— Time* с тремя уровнями). Получившийся план это 2 x 2 x 3 план дисперсионного анализа, где третий фактор - фактор повторных измеренийr.

Задание файла данных. Для задания этого плана пользователь должен отдельно задать межгрупповую часть плана и отдельно внутригрупповую часть плана. Например, представим себе, что данные были введены в файл данных следующим образом:

Переменная:	1	2	3	4	5
	-	-	--	--	--
Наблюдение1	1	1	23	34	32
Наблюдение2	1	2	40	43	45
Наблюдение3	2	2	23	24	23
Наблюдение4	2	1	54	34	54
Наблюдение5	1	2	23	24	25
Наблюдение6	2	1	34	23	34
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.

Первые две переменные (столбцы) - это независимые или группирующие переменные (для упрощения здесь показаны числовые, а не текстовые коды), которые однозначно определяют к какой экспериментальной группе “принадлежит” данное наблюдение (лабораторное животное). Переменные от 3 до 5 - это зависимые переменные, то есть, замеренные активности животных после одного, двух или трех дней. Следовательно, переменная 3 содержит данные для первого уровня фактора повторных измерений, переменная 4 содержит данные для второго уровня фактора повторных измерений, а переменная 5 — для третьего уровня фактора повторных измерений.

Задание плана. Для того чтобы задать меж-внутригрупповой план 2 x 2 x 3, нужно использовать кнопку *Переменные* (в стартовой панели, см. стр.1757), затем выбрать переменные 1 и 2 как независимые переменные, переменные 3, 4 и 5 - как зависимые переменные. Потом, нужно выбрать коды для независимых переменных. В этом примере, коды 1 и 2 были использованы, чтобы определить, каким типом вируса было заражено данное лабораторное животное (независимая переменная 1), и какой тип лекарства получало это животное (независимая переменная 2).

Межгрупповая часть плана полностью задана. Для задания внутригрупповой части нужно с помощью опции *План повторных измерений,* задать один фактор повторных измерений (можно называть его *Время—Time*) с тремя уровнями.

В общем, можно задавать межгрупповую и внутригрупповую часть плана дисперсионного анализа любой сложности с помощью отдельного задания отдельного задания этих частей.

Исследование специфических эффектов и проверка спланированных сравнений В диалоговом окне результаты можно выбрать частные эффекты (*Специфический эффект/средние*) и *Спланированные сравнения* также как и в многофакторных межгрупповых или внутригрупповых планах. Межгрупповые факторы и внутригрупповые факторы (факторы повторных измерений) располагаются в списке последовательно.

StatSoft®

Copyright © StatSoft, 1995

Многомерный дисперсионный анализ зависимых измерений

Модуль *Дисперсионный анализ* будет автоматически проводить многомерный дисперсионный анализ, когда заданы несколько зависимых переменных (т.е. зависимые переменные не являются уровнями фактора повторных измерений). Например, если пользователь задал межгрупповой план с тремя зависимыми переменными, то будет автоматически проведен многомерный дисперсионный анализ (M*ANOVA*) с тремя зависимыми переменными. Однако если еще задан внутригрупповой фактор с тремя уровнями, то будет проведен одномерный дисперсионный анализ, так как три зависимые переменные будут интерпретированы, как уровни фактора повторных измерений. Если было задано шесть зависимых переменных, то многомерный дисперсионный анализ будет проведен еще раз, с двумя зависимыми переменными на каждом уровне фактора повторных измерений.

Модуль *Дисперсионный анализ* будет делить число заданных зависимых переменных, на общее число уровней всех факторов повторных измерений. Если результат этого деления больше 1, будет проведен многомерный дисперсионный анализ (если результат деления не является целым числом, появится сообщение об ошибке).

Например, если задан план с 2 факторами повторных измерений, первый с тремя уровнями и второй — с двумя. Количество заданных зависимых переменных должно быть равным, по крайней мере, шести (так как имеется 2 x 3 = 6 уровней). Однако если пользователь задаст 12 зависимых переменных, то программа будет проводить многомерный дисперсионный анализ с двумя зависимыми переменными (так как 12 поделенное на 6 дает 2). Во всех других случаях, анализ планов с многомерными зависимыми переменными (т.е. подсчет главных эффектов и спланированных сравнений) проходит аналогично одномерному случаю. Например, при задании контраста, пользователь может присваивать зависимым переменным нули (тогда они будут игнорироваться при сравнении). Для задания контрастов для зависимых переменных, нужно установить соответствующую кнопку в диалоге *Опции для спланированных сравнений*. Можно обратиться к стр.1760, где на примерах показано, как задаются многомерные планы повторных измерений.

Гнездовые планы

Простые межгрупповые гнездовые планы

В некоторых исследованиях невозможно осуществить полный факторный план, т. е. план, в котором содержатся измерения при каждом уровне каждого фактора пересеченного с каждым уровнем другого фактора. Например, нужно оценить влияние 4 различных удобрений (*Удобрение* - межгрупповой фактор 1 с 4 уровнями) на рост зерна. Предположим, что используются по два различных вида удобрений на двух различных полях (*Поле* - межгрупповой фактор 2 с двумя уровнями). Здесь мы имеем дело с планом 4 (*Удобрения*) x 2 (*Поля*). Однако так как только два уровня первого фактора находятся внутри каждого из двух уровней второго фактора, то имеется гнездовой (вложенный) план 4 (вкладывается внутри уровней фактора 2) x 2. Вообще, план является гнездовым (вложенным), когда несколько уровней одного фактора содержатся внутри другого фактора. При этом уровни одного фактора размещаются внутри уровней другого фактора как яйца в гнезде. Вернемся к нашему примеру. Данные для этого плана можно ввести следующим образом:

Copyright © StatSoft, 1995

Удобрение	Поле	Рост (зависимая переменная)
1	1	24
1	1	34
2	1	25
2	1	28
3	2	45
3	2	42
4	2	33
4	2	31

Здесь предполагается, что с каждого поля выбирается две выборки зерна, и измеряются их размеры (в некоторой шкале). Каждая строчка в этом файле данных представляет результат этих измерений. Первая переменная (столбец 1) содержит числовые коды, определяющие вид удобрений, при которых выросли замеренные зерна, т.е. это независимая или межгрупповая переменная. Вторая переменная (столбец 2) содержит числовые коды, определяющие поле, на котором выросли замеренные зерна, т.е. это вторая независимая или межгрупповая переменная. Заметим, что уровни 1 и 2 первого межгруппового фактора (*Удобрения)* появляются только внутри первого уровня второго межгруппового фактора *(Поле)*, в то время как уровни 3 и 4 первого фактора появляются только внутри второго уровня второго межгруппового фактора. Следовательно, первый фактор *(Удобрения)* группируется внутри второго фактора *(Поле).*

Задание плана. Для того чтобы задать этот план, нужно, как обычно, в стартовой панели (стр. 1757) сначала выбрать зависимую переменную (переменную 3) и независимые переменные (переменные 1 и 2). Для задания кодовых значений первого межгруппового фактора, нужно ввести 1-4, то есть ввести те самые значения, которые использовались в файле данных. Затем нужно нажать на кнопку *Гнездовой план*. Для каждого фактора нужно задать: (1) будет или нет этот фактор сгруппирован; (2) внутри каких других факторов он будет сгруппирован; (3) сколько уровней фактора располагается в уровнях факторов, в которых он группируется. В нашем примере первый фактор группируется только внутри одного другого фактора, а именно внутри фактора 2. Внутри каждого уровня второго фактора содержится 2 уровня другого фактора 1, поэтому в соответствующее поле нужно ввести 2.

Исследование главных эффектов и взаимодействий. Вследствие группировки, взаимодействие между сгруппированным фактором и фактором, в котором он группируется, не может быть оценено. Так в нашем примере, мы не можем оценить влияние взаимодействия типа удобрения и особенностей поля на размер зерен. Невозможно определить из наших данных, какой был бы объем зерен, выращенных на поле 1 (первый уровень фактора *Поле)* под воздействием третьего и четвертого вида удобрений (уровень 3 и 4 фактора *Удобрения).*

Простые внутригрупповые (с повторными измерениями) планы

В некоторых случаях, нужно исследовать планы, в которых фактор повторных измерений группируется внутри других планов повторных измерений. Например, на субъекте может быть проведено 4 различных теста памяти (фактор 1 повторных измерений с 4 уровнями) в течение двух дней (фактор 2 повторных измерений с 2 уровнями). Если два теста памяти были проведены в течение первого дня и два других во второй день, то можно рассматривать 4 уровня первого фактора повторных измерений сгруппированными в 2 два уровня второго фактора повторных измерений. В настоящем примере файл данных может выглядеть следующим образом:

StatSoft®

Copyright © StatSoft, 1995

Фактор повторных измерений:				
Фактор 1 (Тест), Уровень	1	2	3	4
Фактор 2 (День), Уровень	1	1	2	2
	--	--	--	--
Наблюдение1	23	24	23	25
Наблюдение2	25	23	25	28
Наблюдение3	29	38	37	29
Наблюдение4	48	47	48	47
...	..	..	..	..

Первая переменная (столбец 1) содержит данные для первого уровня первого фактора повторных измерений и первого уровня второго фактора повторных измерений. Вторая переменная (столбец 2) содержит данные для второго уровня первого фактора повторных измерений и первого уровня второго фактора повторных измерений и т. д. Заметим, что в этом примере, 1 и 2 уровни первого фактора повторных измерений входят только с уровнем 1 второго фактора повторных измерений, уровни 3 и 4 первого фактора повторных измерений входят только с уровнем 2 второго фактора повторных измерений. Этот план является гнездовым (вложенным) планом повторных измерений.

Для задания этого плана (в стартовой панели, стр. 1757), сначала нужно ввести все переменные в список зависимых переменных в том же порядке, как они заданы в файле данных, т. е. нужно ввести 1 – 4 в список зависимых переменных соответствующего диалогового окна. Затем нужно нажать на кнопку *План повторных измерений.* После этого нужно сначала задать первый фактор повторных измерений (*Тест).* Задав этот фактор, содержащий 4 уровня, затем нужно задать второй фактор (*День),* содержащий 2 уровня. В этот момент программа еще не знает, что в плане имеется сгруппированный фактор. Число переменных делится нацело на общее уровней. Если было бы задано 4 уровня первого фактора и 2 уровня второго фактора, то программа ожидала бы не менее 8 переменных (общее число уровней 2 x 4 = 8) в списке зависимых переменных. Теперь зададим группировку. Первый фактор повторных измерений группируется внутри фактора 2. По 2 уровня первого фактора располагаются внутри каждого уровня второго фактора. Поэтому нужно ввести 2 в соответствующее поле.

Исследование главных эффектов и взаимодействий. Исследование главных эффектов и взаимодействий в гнездовом (вложенном) плане с повторными измерениями производится тем же способом, что и в полных факторных внутригрупповых планах. Однако взаимодействия между сгруппированными факторами и факторами, в которые они группируются, не могут быть оценены.

В этом примере нельзя знать, как субъекты справились бы с первым и вторым тестами (уровень 1 и 2 первого фактора повторных измерений) во второй день (уровень 2 второго фактора повторных измерений). В этом исследовании не содержится необходимой информации для оценки такого взаимодействия.

Меж-внутригрупповые планы. В планах, содержащих как межгрупповые факторы, так и внутригрупповые (факторы повторных измерений), можно проводить группировку и межгрупповой части плана и внутригрупповой. Отметим, что межгрупповые факторы могут группироваться только в межгрупповые, а внутригрупповые (факторы повторных измерений) только во внутригрупповые.

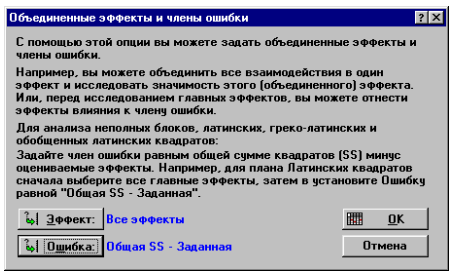
Латинские квадраты, греко-латинские квадраты, обобщенные латинские квадраты

Латинские квадраты являются обширным классом сбалансированных неполных планов (см. например, Hays, 1988; Lindman, 1974; Milliken and Johnson, 1984; Winer, 1962). Пример плана латинского квадрата обсуждался в разделе *Вводный обзор* (стр. 1720). В гнездовых планах используется такая комбинация уровней, которая позволяет оценить главные эффекты при

Copyright © StatSoft, 1995

StatSoft®

минимальном числе наблюдений. Для анализа планов такого типа используется опция *Объединенные эффекты/члены ошибки* в диалоговом окне *Результаты*.



Сначала выбирается нужный *Эффект* в диалоговом окне *Объединенные эффекты и члены ошибки* (или по умолчанию выбрать установку *Все эффекты*). Затем нужно нажать на кнопку *Ошибка* и отнести все главные эффекты к членам ошибки. Затем нужно вновь нажать на кнопку *Ошибки* и установить *Общая SS - Заданная*. Таким образом, может быть оценена значимость выбранных эффектов с соответствующим значением квадрата ошибок.

Другие специальные планы

Планы с одним наблюдением в ячейке

В некоторых случаях не так просто получить наблюдения для анализа, т.к. это связано с большими затратами. Тогда экономический аспект берет вверх над требованиями плана. Поэтому часто исследователь располагает лишь одним наблюдением в ячейках. Если план является полным факторным планом с одним наблюдением в каждой ячейке, то сумма квадратов эффектов “исчерпывает” число степеней свободы и *SS* ошибок не может быть подсчитана.

При анализе таких планов сначала вычисляется полная таблица всех эффектов. Так как *SS* ошибок не может быть вычислена, то *F* статистики тоже не вычисляются. Затем исследуются средние квадраты (*MS*; *SS* деленная на соответствующую степень свободы). Их величина оценивает “значимость” соответствующего эффекта. Используя эту информацию, можно решить, какие эффекты отнести к членам ошибки (маленькие, незначимые эффекты). Обычно в таких планах отбрасываются взаимодействия высокого порядка. Наконец, используя опцию *Объединенный эффект и член ошибки* в диалоговом окне *Результаты*, выбираются эффекты, относящиеся к членам ошибки, и исследуется статистическая значимость эффектов. Следует помнить, что в этом случае к статистической значимости критериев нужно относиться с осторожностью, так как вычисляется теоретически не совсем *F* статистика (маленькие эффекты относятся к членам ошибки).

Случайные блочные планы

Иногда эксперименты с одним наблюдением в ячейке специальным образом организуются для того, чтобы уменьшить *SS* ошибку и получить более чувствительный план. Например, наблюдения могут быть организованы в блоки, которые позволяют оценивать несмещенный главный эффект блочного фактора. Затем *SS* ошибки уменьшается за счет *SS* блочного фактора.

Пример. Предположим, исследуется урожайность различных сортов пшеницы при трех типах удобрений. Для исследований выбрано четыре поля, это можно использовать как дополнительный блочный фактор плана. План может быть представлен следующим образом:

StatSoft®
Copyright © StatSoft, 1995

		Поле (Блок)			
Удобрение	Сорт	I	II	III	IV
1	A	10.2	10.1	11.1	12.3
	B	11.1	9.8	8.6	9.4
	C	.	.	.	.
	D	.	.	.	.
2	A	.	.	.	.
	B	.	.	.	.
	C	.	.	.	.
	D	.	.	.	.
3	A	.	.	.	.
	B	.	.	.	.
	C	.	.	.	.
	D	.	.	.	.

В данном примере влияние блочного фактора не представляет интерес, иными словами, не представляет интерес статистическая значимость различия урожайности разных полей. Однако, оценивая *SS* для блочного фактора, можно уменьшить дисперсию ошибки, увеличивая тем самым чувствительность критерия значимости для эффектов факторов *Удобрение* и *Сорт* и их взаимодействий. В планах такого типа также принято игнорировать любые взаимодействия блочного фактора с исследуемыми факторами.

Задание файла данных. Файл данных для этого примера задается так же, как и для полного факторного межгруппового эксперимента. Файл будет содержать три группирующие переменные (*Удобрения, Сорт, Поле)* с кодами, однозначно определяющими, к какой ячейке плана принадлежит каждое наблюдения. Четвертой переменой в файле будет зависимая переменная *(Урожай).*

Задание и анализ плана. Этот план задается так, как если бы он был полным факторным планом. При исследовании статистической значимости эффектов, представляющих интерес, нужно пользоваться опцией *Объединенный эффект и член ошибки,* для отнесения соответствующих членов к ошибкам. Напомним, что в рассматриваемом примере представляют интерес эффекты *Удобрение* и *Сорт* и их взаимодействие. Кроме того, могут быть оценены главные эффекты блочного фактора (*Поле).* Для задания члена ошибки сначала нужно нажать на кнопку *Все,* а затем на кнопку *Ошибка* для установки переключателя на *Общая SS – Заданная*.

Случайные эффекты

В некоторых исследованиях требуется оценка значимости случайных эффектов. Например, можно исследовать гипотезу о различии в алгебраической образованности студентов в разных школах штата (т.е. школа представляет здесь случайный фактор). Если бы гипотеза была более специальной (например, знание алгебры в сельских школах лучше, чем в городских), то фактор школа имел бы постоянный эффект. Но основная гипотеза предполагает, что из совокупности школ в штате *случайным образом* выбираются школы, затем в каждой школе выбираются студенты для тестирования. Поэтому каждый уровень фактора школа (т.е. каждая школа выбранная случайным образом из совокупности школ штата) представляет один из возможных уровней этого фактора, выбранный случайным образом из совокупности всех его уровней (гипотетически допускаются *любые различия* между школами штата).

На практике фактор будет, скорее всего, исследоваться как случайный, если при повторении исследования не выбирается тот же уровень этого фактора. В нашем примере, если нужно повторить исследование, не следует тестировать студентов в тех же школах, а нужно выбрать другие школы. Фактор школа в нашем примере является случайным фактором.

С другой стороны, предположим нужно повторить эксперимент с обучением, где испытуемые должны запомнить список, состоящий либо из 10 слов, либо из 20 (уровень 1 и уровень 2 межгруппового фактора числа слов). В этом случае при повторении эксперимента будут выбираться точно те же уровни

Copyright © StatSoft, 1995

StatSoft®

межгруппового фактора, а именно, либо список из 10 слов, либо из 20. Этот межгрупповой фактор является фактором с постоянным эффектом.

Если различие между случайными и постоянными эффектами осталось неясным, то следует обратиться к учебникам по дисперсионному анализу (например, Lindman, 1974; Winer, 1962, 1971).

Сравнение экспериментов

Другой пример, когда можно задать фактор как случайный, появляется при сравнении различных экспериментов. Иногда нужно сравнить результаты различных экспериментов, в которых используется один и тот же план. В этом случае можно рассматривать межгрупповой фактор *Эксперимент* как случайный фактор. Можно считать, что случайным образом выбрано несколько экспериментов из совокупности всех случайных экспериментов.

Задание плана

Чтобы задать случайный фактор, нужно сначала задать все (межгрупповые и внутигрупповые) факторы. Затем в стартовой панели нужно нажать на кнопку *Случайные факторы* и выбрать факторы плана, которые будут исследоваться как случайные.

Исследование главных эффектов и взаимодействий

Если есть эффекты, которые полностью пересекаются со случайными эффектами, то взаимодействия соответствующих факторов со случайными фактором относятся к членам ошибки.

Технически, спланированные сравнения не могут оцениваться в плане со случайными эффектами (см. работу Kirk, 1982). Если попытаться сделать это, то модуль *Дисперсионный анализ* выдаст сообщение об ошибке, а затем поступит так, как будто план не содержит случайных факторов.

Фиксированные и переменные ковариаты в межгрупповых и внутригрупповых планах

Ковариаты представляют в планах непрерывные факторы. Они обычно включаются для того, чтобы уменьшить *SS* ошибки, для получения более чувствительных критериев.

Фиксированные ковариаты

Самое обычное применение ковариат в межгрупповых планах касается случая, когда имеется непрерывная переменная, которая, скорее всего, коррелирована с изучаемой зависимой переменной. Например, перед началом обучения специальному математическому курсу у всех студентов был измерен IQ (коэффициент умственного развития). Студенты случайным образом разбивались на две группы, в которых читались разные курсы, а затем измерялись их успехи в освоении этого курса. Эти измерения были исследуемой зависимой переменной.

В этом исследовании можно предполагать, что различие в умственных возможностях студентов вносит большой вклад в разброс зависимой переменной. В действительности этот вклад в разброс может быть настолько значительным, что он покроет различие в эффективности двух разных курсов. В этом случае можно определить IQ как ковариату. Введение этой ковариаты может существенно уменьшить *SS* ошибки.

Скорректированные средние. Если ковариаты не только коррелированы с

StatSoft®

Copyright © StatSoft, 1995

зависимыми переменными внутри каждой группы плана, но также и с межгрупповыми факторами, тогда необходимо корректировать средние перед интерпретацией любого эффекта. Ковариата коррелирована с межгрупповыми факторами, если эти факторы как-то влияют на нее.

Например, если в рассматриваемом примере, IQ измерялось *после* прохождения математических курсов, то, вполне возможно, что на (IQ) было оказано влияние различных курсов. В этих случаях, можно видеть, что включение ковариаты в план не только влияет на значение *SS* ошибок, но и на значение *SS* эффектов для межгрупповых факторов. Кроме того, скорректированные средние (вычисленные модулем *Дисперсионный анализ* с помощью *Опции вывода* в диалоговом окне результаты, см. стр. 1763) отличаются от значений наблюденных средних. В этих случаях включение ковариаты в план иногда даже уменьшает статистическую значимость эффектов.

Анализ контрастов. Во многих учебниках по дисперсионному анализу содержится до некоторой степени уклончивое отношение к анализу контрастов в планах с ковариатами. В модуле *Дисперсионный анализ* для таких планов контрасты задаются как обычно, используется общая вычислительная процедура, и результаты анализа контрастов автоматически выводятся с учетом ковариат. Отметим, что на самом деле анализ контрастов выполняется для скорректированных средних.

Переменные ковариаты

Эти типы ковариат встречаются только в планах с повторными измерениями (см. стр. 1720 и 1790). Основная идея переменных ковариат может быть проиллюстрирована следующим образом. Представим себе, что в плане с повторными измерениями, ковариата измеряется при каждом измерении зависимой переменной. Можно задать следующие два вопроса: (1) есть ли связь между зависимой переменной и ковариатой; (2) связаны ли между собой изменения ковариат при разных измерениях с изменениями зависимой переменной? Первый вопрос исследуется так же, как и для фиксированных ковариат. Второй вопрос требует, чтобы *SS* ошибки для плана с повторными измерениями была скорректирована соответствующими ковариатами.

Вычислительный подход. Модуль *Дисперсионный анализ* использует общую линейную модель дисперсионного анализа. Корректировка внутригрупповых эффектов ковариатами происходит практически автоматически. После вычисления соответствующих внутригрупповых гипотез и матриц ошибок, перед выводом окончательных результатов, вычисляются гипотезы и матрицы ошибок для средних, скорректированных ковариатами. Эта процедура многомерна по своей природе и должна быть определенная уверенность в правомерности применения многомерного критерия для повторных измерений (и ковариат) с более чем двумя уровнями.

Заметим, что результаты регрессии для переменных ковариат выводятся только, если эти результаты запрошены через опцию *Опции вывода* в диалоговом окне *Результаты* (стр. 1763). Затем исследуется *специфический эффект,* включающий фактор повторных измерений. Если выбрана опция *Регрессия внутри ячеек*, то будет выполнен регрессионный анализ внутри группы для каждого уровня фактора повторных измерений в качестве зависимой переменной.

Межгрупповой и внутригрупповой дисперсионный анализ. Если план содержит межгрупповые факторы и факторы повторных

Copyright © StatSoft, 1995

измерений, то результаты для ковариат рассматриваемой модели будут такими же, как и для модели с одной зависимой переменной равной сумме значений переменных по всем уровням фактора повторных измерений.

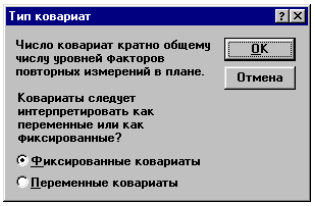
Задание переменных ковариат.

Переменные ковариаты задаются точно так же, как задаются факторы повторных измерений (см. стр. 1720 и 1790). Программа будет циклически проходить список ковариат (введенных через опцию *Ковариаты* в стартовой панели) и назначать переменным последовательно уровни факторов повторных измерений. Как и при задании факторов повторных измерений, быстрее всего будет меняться уровень фактора повторных измерений, который был задан последним, следующим по скорости изменения уровня будет фактор, заданный предпоследним и т. д. Если имеется несколько переменных ковариат, тогда самое быстрое изменение уровня фактора равно числу ковариат. “Соседние” переменные в списке ковариат с одним значением уровня фактора, введенного последним, "принадлежат" разным ковариатам.

Например, предположим, что имеется два теста замеренных три раза (1 месяц, 2 месяца, 3 месяца). Нужно задать эти два теста как переменные ковариаты. Предположим, что данные для ковариат были введены в файл следующим образом:

Тестовый контроль через					
1 месяц		2 месяца		3 месяца	
Тест		Тест		Тест	
1	2	1	2	1	2
24	30	26	34	28	28
30	32	30	28	32	34
30	27	29	28	28	33
30	28	29	26	34	35
35	25	36	27	32	36
..	..	..	..	..	..

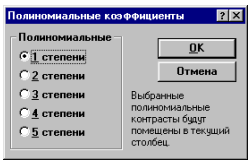
Зададим список ковариат в этом порядке (конечно, нужно задать внутригрупповую модель с одним фактором повторных измерений с тремя уровнями). При выходе из стартовой панели нужно задать тип ковариат (фиксированные или переменные).



Если выбраны *Переменные* ковариаты, то в этом примере два теста интерпретируются как две переменные ковариаты. Заметим, что вопрос, касающийся природы ковариат, возникает только, если число заданных ковариат делится нацело на общее число уровней факторов повторных измерений, в противном случае ковариаты интерпретируются как фиксированные.

Анализ полиномиальных контрастов

Так как пользователь может задавать контрастные коэффициенты для всех факторов плана, то легко исследовать полиномиальные (линейные, нелинейные) тренды (для главных эффектов, взаимодействий) по всем уровням факторов. *STATISTICA* позволяет легко и быстро задавать полиномиальные контрасты в соответствующем диалоговом окне. После выбора опции *Спланированные сравнения* из диалогового окна *Результаты* нужно в открывшемся окне *Задайте контрасты для этого фактора* нажать на кнопку *Полиномиальные.* После этого в новом окне выбрать степень полиномиального контраста.



StatSoft®
Copyright © StatSoft, 1995

Представленная ниже таблица контрастов может использоваться для факторов, содержащих до 10 уровней.

Уров-ни	Поли-ном	Коэффициенты
3	1 степени	-1 0 1
	2 степени	1 -2 1
4	1 степени	-3 -1 1 3
	2 степени	1 -1 -1 1
	3 степени	-1 3 -3 1
5	1 степени	-2 -1 0 1 2
	2 степени	2 -1 -2 -1 2
	3 степени	-1 2 0 -2 1
	4 степени	1 -4 6 -4 1
6	1 степени	-5 -3 -1 1 3 5
	2 степени	5 -1 -4 -4 -1 5
	3 степени	-5 7 4 -4 -7 5
	4 степени	1 -3 2 2 -3 1
	5 степени	-1 5-10 10 -5 1
7	1 степени	-3 -2 -1 0 1 2 3
	2 степени	5 0 -3 -4 -3 0 5
	3 степени	-1 1 1 0 -1 -1 1
	4 степени	3 -7 1 6 1 -7 3
	5 степени	-1 4 -5 0 5 -4 1
8	1 степени	-7 -5 -3 -1 1 3 5 7
	2 степени	7 1 -3 -5 -5 -3 1 7
	3 степени	-7 5 7 3 -3 -7 -5 7
	4 степени	7-13 -3 9 9 -3-13 7
	5 степени	-7 23-17-15 15 17-23 7
9	1 степени	-4 -3 -2 -1 0 1 2 3 4
	2 степени	28 7 -8-17-20-17 -8 7 28
	3 степени	-14 7 13 9 0 -9-13 -7 14
	4 степени	14-21-11 9 18 9-11-21 14
	5 степени	-4 11 -4 -9 0 9 4-11 4
10	1 степени	-9 -7 -5 -3 -1 1 3 5 7 9
	2 степени	6 2 -1 -3 -4 -4 -3 -1 2 6
	3 степени	-42 14 35 31 12-12-31-35-14 42
	4 степени	18-22-17 3 18 18 3-17-22 18
	5 степени	-6 14 -1-11 -6 6 11 1-14 6

Если нужно, можно ввести эти полиномиальные контрасты в любой комбинации с другими множествами контрастов в анализе контрастов. Например, критерии для полиномиальных трендов могут вычисляться для выбранного множества плана, или они могут быть включены, как часть сравнений полиномиальных взаимодействий.

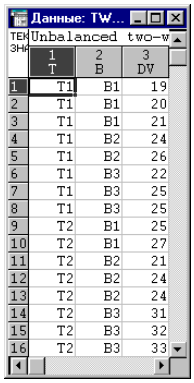
Copyright © StatSoft, 1995
StatSoft®

ПРИМЕЧАНИЯ, ЧАСТЬ II: ПРИМЕРЫ ДИСПЕРСИОННОГО АНАЛИЗА НЕСБАЛАНСИРОВАННЫХ И НЕПОЛНЫХ ПЛАНОВ

В следующих примерах демонстрируется, как надо проводить анализ различных несбалансированных и неполных планов. В очень хороших книгах Dodge (1985), Lindman (1974), Milliken и Johnson (1984), и Searle (1987) обсуждают вопросы, возникающие при анализе таких планов. Рекомендуется внимательно изучить эти работы перед исследованием сложных экспериментов.

Пример1: Несбалансированные планы и сумма квадратов типа I и II

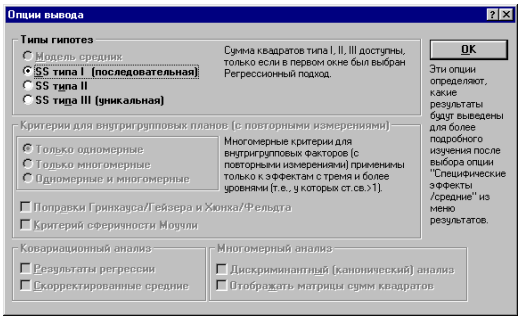
Milliken и Johnson (1984, стр. 129) обсуждают некоторые детали несбалансированного межгруппового плана 2 x 3. Файл данных *Twoway.sta* показан ниже в электронной таблице.



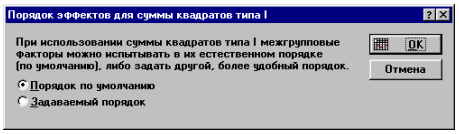
Этот план несбалансирован, так как ячейки *T1,B2* и *T2,B1* содержат по два наблюдения, так как оставшиеся ячейки содержат по три наблюдения. В главе 10 своей книги Милликен и Джонсон обсуждают вычисление и интерпретацию сумм квадратов *типа I,II и III* для факторов и их взаимодействий.

Для анализа, сначала нужно открыть файл данных *Twoway.sta* и задать переменные *T* и *B* как независимые переменные, а переменную *DV* — как зависимую. Также в стартовой панели нужно установить опцию *Регрессионный подход*.

Теперь, нужно нажать *OK* для начала анализа (*STATISTICA* автоматически выберет коды для независимых переменных). Для вычисления таблицы дисперсионного анализа *типа I* щелкните на кнопке *Опции вывода* и в получившемся диалоге выберите опцию *SS типа I (последовательная)*.

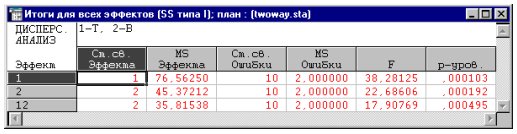


Нажмите *OK* в этом диалоговом окне, а затем в диалоговом окне *Результаты дисперсионного анализа* щелкните на кнопку *Все эффекты*. Откроется диалоговое окно *Порядок эффектов для суммы квадратов типа I*. В этом диалоговом окне можно выбрать порядок межгрупповых факторов в плане. Для этого примера выберите *Порядок по умолчанию*.



Нажмите *OK* для вывода таблицы результатов дисперсионного анализа *типа I* для всех эффектов.

StatSoft®
Copyright © StatSoft, 1995



Сумма квадратов типа I и опция *Специфический эффект/средние/графики*. Когда используется опция *Специфический эффект/средние/графики* в диалоговом окне *Результаты дисперсионного анализа,* то после запроса анализа *типа I* изучаемый эффект будет введен в модель после всех эффектов, которые ниже его по порядку, но перед всеми эффектами, которые выше его по порядку. Следовательно, изучаемый эффект исследуется после проверки всех низших по порядку эффектов. Эффекты такого же или высшего порядка игнорируются.

Пример 2: Простой гнездовой (вложенный) план

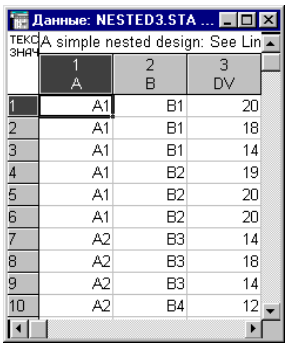
Теперь обсудим простой гнездовой (вложенный) план. Линдман (Lindman 1974, стр. 167) обсуждает двухфакторный гнездовой план, где фактор *A* имеет три уровня и фактор *B* — шесть. В каждом уровне фактора A расположены 2 уровня фактора B.

Структуру этого плана можно представить следующим образом (через X обозначаются наблюдаемые комбинации факторов):

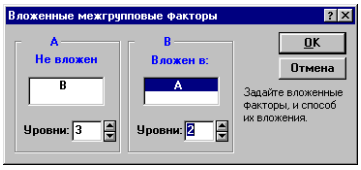
```
Фактор B        Фактор A
                1   2   3
-----------------------
1               x
2               x
3                   x
4                   x
5                       x
6                       x
```

Легко заметить, что в этом плане фактически 6 экспериментальных групп, тогда как для полного факторного плана нужно 18 групп. В этом плане можно оценить главные эффекты для двух факторов, которые будут несмешиваемыми, если между факторами нет никакого взаимодействия (это предположение существенно, так как оценить все взаимодействия можно только в полном плане).

Данные, напечатанные Линдманом, были введены в файл данных *Nested3.sta*, и представлены ниже.

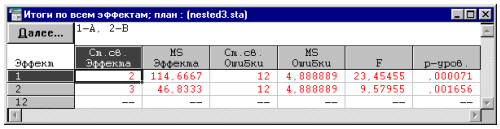


Для анализа этого плана, в стартовой панели выберите переменные *A* и *B* как независимые и переменную *DV* как зависимую. Нажмите на кнопку *Коды* и выберите коды для независимых переменных (*1-3* и *1-6* соответственно). Для задания группировки сначала нужно щелкнуть на кнопке *Гнездовой план,* а затем выбрать фактор B (сгруппированный), фактор A (куда группируется фактор B) и 2 (число уровней фактора B на каждом уровне фактора A).



Затем нужно щелкнуть *OK* в этом диалоговом окне и в стартовой панели модуля *Дисперсионный анализ* для начала анализа. В диалоговом окне *Результаты дисперсионного анализа*, щелкните на кнопке *Все эффекты* для просмотра таблицы полного дисперсионного анализа.

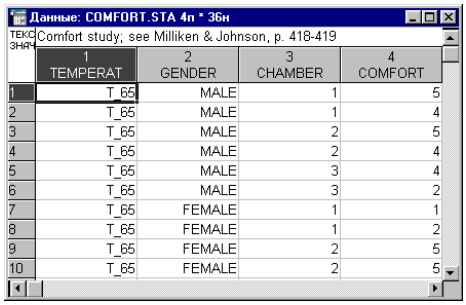
Copyright © StatSoft, 1995



Эти результаты обсуждаются на стр.172 в книге Lindman (1974).

Пример 3: Трехфакторный гнездовой (вложенный) план.

Теперь обсудим трехфакторный гнездовой план. Милликен и Джонсон (Milliken и Johnson 1984, стр. 418) представляют пример такого плана (соответствующий файл данных *Comfort.sta* представлен ниже).

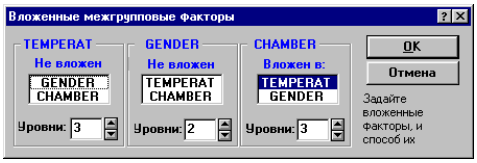


В этом эксперименте, субъекты (мужчины и женщины), случайно расположены в одной из 9 экологических камер (комнат). Эти камеры, в свою очередь, были распределены по трем уровням температурного фактора. Следовательно, в этом плане, фактор *Комната–Chamber* группируется в фактор *Температ — Temperature*, а субъекты группируются в фактор *Комната–Chamber*. Файл данных показан выше. Структура этого плана может быть представлена следующим образом (через X обозначается наблюдаемую комбинацию факторов):

```
Комната
Темпе-   Пол      -------------------------
ратура              1  2  3  4  5  6  7  8  9
-----------------------------------------
  65     мужчина   x  x  x
         женщина   x  x  x
  70     мужчина            x  x  x
         женщина            x  x  x
  75     мужчина                     x  x  x
         женщина                     x  x  x
```

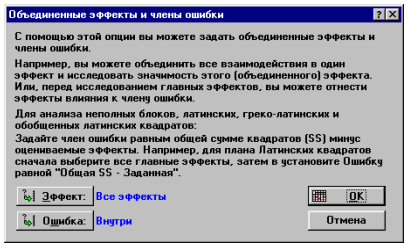
Для анализа плана откройте файл данных *Comfort.sta*. Затем щелкните на кнопке *Переменные* и выберите переменные *Температ - Temperat*, *Пол - Gender*, и *Комната - Chamber,* как независимые переменные, и переменную *Комфорт - Comfort,* как зависимую переменную. Затем щелкните на кнопке *Коды* и выберите коды для независимых переменных, а именно, *1-3* для *Температ - Temperat*, *1 2* для *Пола - Gender*, и *1-9* для *Комната - Chamber*.

Теперь задайте группировку, щелкнув на кнопке *Гнездовой план*, и определите фактор *Chamber*, как сгруппированный в факторе *Temperat* с 3 уровнями на каждом уровне этого фактора.

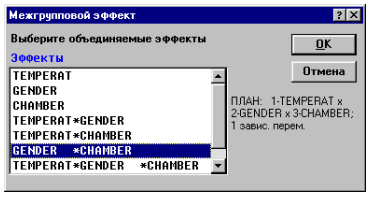


Нажмите *OK* здесь и в стартовой панели для начала анализа. Для получения результата, как показано в Milliken и Johnson (1984, стр. 419), используйте опцию *Объединенный эффект и член ошибки* в диалоговом окне *Результаты дисперсионного анализа.* Появится диалоговое окно *Объединенные эффекты и члены ошибки*.

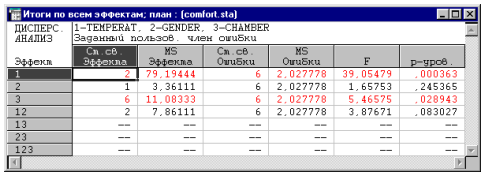
StatSoft®
Copyright © StatSoft, 1995



В этом диалоге нужно щелкнуть на кнопке *Ошибка* и отнести взаимодействие факторов *Пол - Gender* и *Комната - Chamber* в член ошибки (т.е. нужно добавить взаимодействие *23* во *Внутреннюю* ошибку).



Если щелкнуть *OK* в этом диалоговом окне, то появится текст (следом за кнопкой *Ошибка* в диалоговом окне *Объединенные эффекты и члены ошибки*). Этот текст будет таким: "*Заданная*". Для оценки суммы квадратов для всех эффектов нужно теперь щелкнуть *OK*.

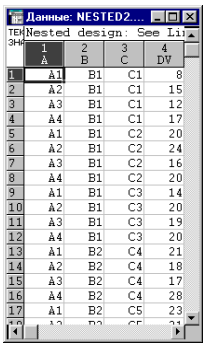


При анализе гнездовых (вложенных) планов важно понимать специфическую природу каждого индивидуального плана. Согласно анализу Milliken и Johnson, в этом частном случае имеются два члена ошибки. Во-первых, есть член, соответствующий распределению *Субъектов — Subjects* по уровням фактора *Комнаты – Chambers*. В таблице, приведенной выше, средний квадрат для этой ошибки содержится в столбце *MS ошибки.* Этот член ошибки используется для исследования эффекта фактора *Комнаты– Chamber*. Однако для того чтобы исследовать эффект фактора *Температ - Temperat* (температура) нужно использовать член ошибки, возникающей при распределении *Комнат - Chamber* по уровням фактора *Температура - Temperat*. Следовательно, нужно использовать опцию *Объединенный эффект и член ошибки* еще раз.

Пример 4:
Анализ плана с повторными измерениями как гнездового (вложенного) плана со случайными эффектами

Линдман (Lindman 1974, стр. 173) рассмотрел интересный пример гнездового плана, в котором группирующий фактор – случаен. Каждому из 9 субъектов было предложено изучить по четыре списка бессмысленных символов по одной из трех инструкций. Здесь имеется три фактора. Фактор A с четырьмя уровнями, которые соответствуют четырем спискам бессмысленных символов. Фактор *B* с тремя уровнями, соответствующими трем разным типам инструкций и фактор C с 9 уровнями, соответствующим 9 субъектам.

Заметим, что фактор C (субъекты) сгруппирован внутри фактора B (каждые из трех субъектов, обучаются по одной из трех инструкций). Ниже показан файл данных *Nested2.sta* из Lindman (1974).



Copyright © StatSoft, 1995

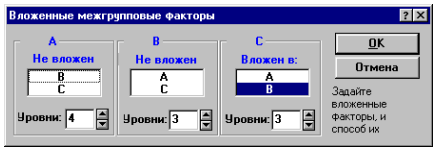
StatSoft®

Структура эксперимента может быть представлена следующим образом (через X обозначены комбинации, участвующие в эксперименте):

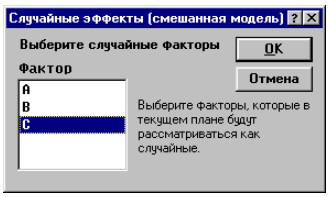
Субъект Фактор A	Фактор B	1	2	3	4	5	6	7	8	9
A1	B1	x	x	x						
	B2				x	x	x			
	B3							x	x	x
A2	B1	x	x	x						
	B2				x	x	x			
	B3							x	x	x
A3	B1	x	x	x						
	B2				x	x	x			
	B3							x	x	x
A4	B1	x	x	x						
	B2				x	x	x			
	B3							x	x	x

Для анализа этого плана, сначала откроем файл *Nested2.sta*. Зададим *A*, *B*, и *C*, как независимые переменные и *DV* как зависимую переменную. Затем зададим коды для независимых переменных (*1-4*, *1-3*, и *1-9,* соответственно).

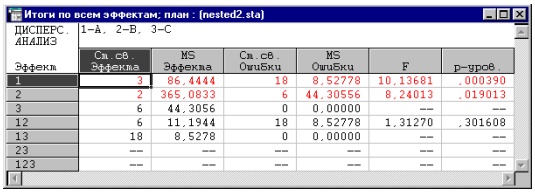
Как видно из представленной выше схемы плана, фактор C (субъекты) группируется внутри уровней фактора B. Для задания этой группировки, нужно нажать на кнопку *Гнездовой план* и определить фактор C, как сгруппированный в фактор B. В каждый уровень фактора B группируются 3 уровня фактора C.



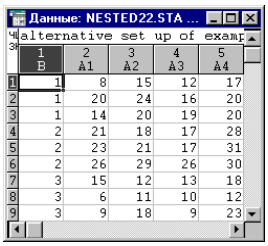
Фиксированные и случайные эксперименты обсуждаются на стр. 1801 в этом разделе. Короче говоря, когда уровни фактора выбираются случайным образом из совокупности уровней и нужно получить сведения обо всей этой совокупности, тогда соответствующий фактор будет рассматриваться как случайный. В рассматриваемом примере фактор C (сгруппированный в фактор B) должен рассматриваться как случайный, так как обычно нужно получить сведения относительно всей популяции, из которой случайным образом выбираются субъекты. Поэтому фактор *C* задается как случайный с помощью кнопки *Случайные факторы*.



Теперь нужно нажать на *OK* в стартовой панели и начать анализ. В диалоговом окне *Результаты дисперсионного анализа* нужно нажать на кнопку *Все эффекты*. Тогда появится таблица, такая же, как и в книге Lindman (1974, стр. 178).

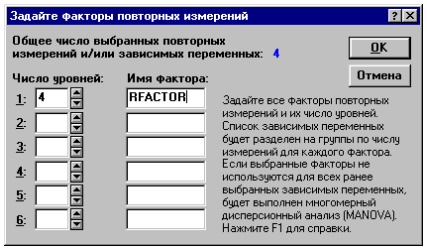


Анализ плана с повторными измерениями. Как видно план, описанный выше, похож на план с повторными измерениями. Можно рассматривать фактор A (четыре списка бессмысленных символов) как фактор повторных измерений, так как каждый субъект изучал каждый из четырех списков. Фактор B является межгрупповым фактором, как так различные субъекты слушали различные инструкции. Поэтому данные могли быть введены в файл следующим образом:



StatSoft®
Copyright © StatSoft, 1995

Чтобы анализировать эти данные с помощью плана с повторными измерениями, нужно нажать на кнопку *Переменные* и выбрать переменную B как независимую переменную, а переменные с A1 по A4 как зависимые переменные. Затем нужно задать фактор повторных измерений (для этого нужно нажать на кнопку *План с повторными измерениями* с четырьмя уровнями.



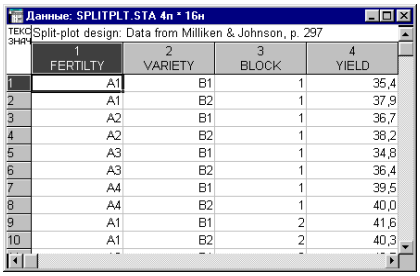
После этого нужно нажать *OK* в этой панели и в стартовой панели *Дисперсионный анализ* и перейти в диалоговое окно *Результаты дисперсионного анализа*. В этом окне нужно нажать на кнопку *Все эффекты* и появится таблица всех эффектов, представленная ниже.

Эффекта	Сч. св. Эффекта	MS Эффекта	Сч. св. Ошибки	MS Ошибки	F	p-уров.
1	2	365,0833	6	44,30556	8,24013	,019013
2	3	86,4444	18	8,52778	10,13681	,000390
12	6	11,1944	18	8,52778	1,31270	,301608

Видно, что результаты совпадают с теми, что были получены ранее.

Пример 5: Планы с расщепленными делянками

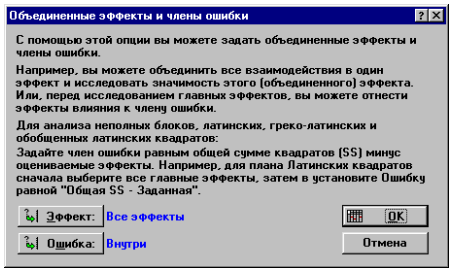
План, рассмотренный выше, иногда относят к так называемым планам с расщепленными делянками, так как такие планы обычно используются в сельскохозяйственных исследованиях. Допустим, что нужно изучить эффективность четырех типов удобрений (фактор A) на двух сортах пшеницы (фактор B). Исследователь делит поле на два блока, каждый из которых делится затем на четыре отдельных участка (делянки). Затем каждое из четырех удобрений наносится на один из участков внутри каждого блока. Каждый участок затем делится пополам (делянка расщепляется), и на разные половины этого участка высаживаются два разных сорта пшеницы. Пример такого плана представлен в Milliken and Johnson (1984, стр. 297). Данные, введенные в файл *Splitplt.sta*, показаны ниже.



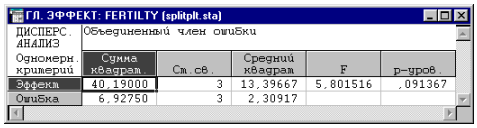
	1 FERTILTY	2 VARIETY	3 BLOCK	4 YIELD
1	A1	B1	1	35,4
2	A1	B2	1	37,9
3	A2	B1	1	36,7
4	A2	B2	1	38,2
5	A3	B1	1	34,8
6	A3	B2	1	36,4
7	A4	B1	1	39,5
8	A4	B2	1	40,0
9	A1	B1	2	41,6
10	A1	B2	2	40,3

Для получения таблицы дисперсионного анализа, представленной в Milliken и Johnson (стр. 299), сначала нужно задать обычный межгрупповой план. Нужно щелкнуть на кнопке *Переменные* в стартовой панели *Дисперсионный анализ* и выбрать переменные *Удобрения, Сорт, Блок - Fertility*, *Variety* и *Block* как независимые переменные и переменную *Урожайность* как зависимую переменную. Затем нужно задать соответствующие коды для независимых переменных. Нажмите *OK* для ввода данных. Соответствующий член ошибки для фактора *Удобрения - Fertility* в этом плане – это взаимодействие между факторами *Удобрения - Fertility* и *Блок - Block* (так как экспериментальными единицами являются участки, см. Milliken и Johnson, 1984, стр. 297, для подробного описания этого плана). Следовательно, щелкните на кнопке *Объединенный эффект и член ошибки* в диалоговом окне *Результаты дисперсионного анализа*. Появится диалоговое окно *Объединенные эффекты и члены ошибки*.

Copyright © StatSoft, 1995

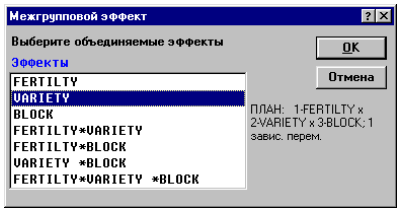


Теперь нужно щелкнуть на кнопке *Эффект* и выберите для исследования эффект *Удобрения – Fertility*. Затем щелкните на кнопке *Ошибки* и задайте взаимодействие фактора *Удобрения - Fertility* с фактором *Блок - Block* (*13*) как член ошибки. Ниже показана итоговая таблица дисперсионного анализа

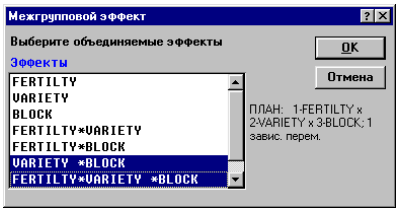


Фактор *Сорт - Variety* так же, как взаимодействие факторов *Сорт - Variety* с *Удобрения - Fertilizer* включает в себя сравнения между подделянками. Поэтому соответствующие эффекты должны сравниваться с разбросом по подделянками.

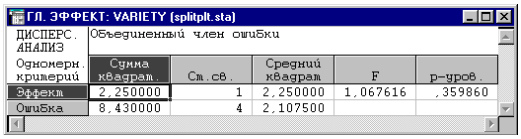
Теперь еще раз щелкните на кнопке *Объединенный эффект и член ошибки.* В получившемся диалоговом окне *Объединенные эффекты и члены ошибки* щелкните на кнопке *Эффект* и выберите для исследования фактор *Сорт – Variety.*



Нажмите на кнопку *Ошибка* и выберите попарное взаимодействие факторов *Сорт - Variety* с фактором *Блок - Block* и взаимодействие трех факторов *Удобрения – Fertility Сорт – Variety, Блок - Block* как объединенный член ошибки.

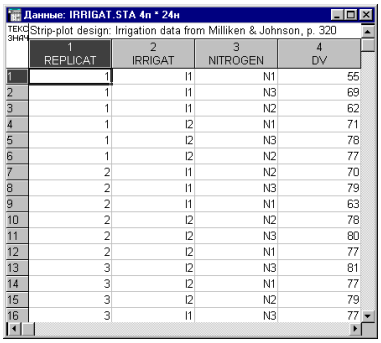


Ниже приведена таблица результатов дисперсионного анализа.



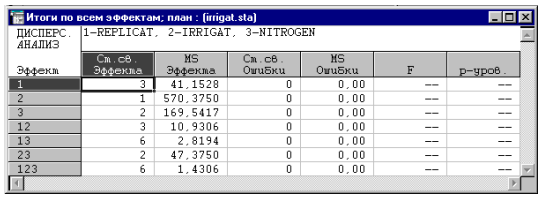
Пример 6:
Планы с полосными делянками

План с полосными делянками похож на план с расщепленными делянками. Детальное обсуждение таких планов можно найти в Milliken и Johnson (1984, стр. 321). В обсуждаемом здесь эксперименте, исследуется влияние взаимодействия двух способов ирригации и трех типов азотных удобрений на урожайность зерна. Данные примера введены в файл данных *Irrigat.sta*.



StatSoft®
Copyright © StatSoft, 1995

Можно легко получить таблицу дисперсионного анализа, представленную в книге Milliken and Johnson на стр. 320. Все компоненты дисперсии автоматически вычисляются в модуле *Дисперсионный анализ* при выборе опции *Все эффекты.* Заметим, что в таблице, представленной в Milliken and Johnson, содержится опечатка, а именно сумма квадратов для фактора *Ирригация - Irrigation* равна 570.4 (а не 507.4). С помощью опции *Объединенный эффект и член ошибки* получены результаты, представленные ниже.


Итоги по всем эффектам; план : (irrigat.sta)

ДИСПЕРС. АНАЛИЗ 1-REPLICAT, 2-IRRIGAT, 3-NITROGEN

Эффект	Сч. св. Эффекта	MS Эффекта	Сч. св. Ошибки	MS Ошибки	F	p-уров.
1	3	41,1528	0	0,00	--	--
2	1	570,3750	0	0,00	--	--
3	2	169,5417	0	0,00	--	--
12	3	10,9306	0	0,00	--	--
13	6	2,8194	0	0,00	--	--
23	2	47,3750	0	0,00	--	--
123	6	1,4306	0	0,00	--	--

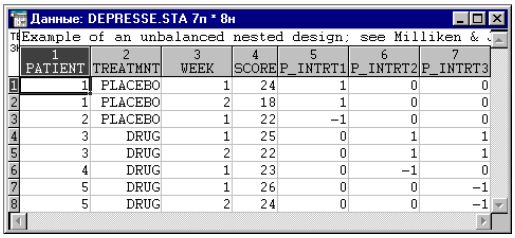
Пример 7: Планы с расщепленными делянками с неравным числом подделянок

Все примеры, рассматриваемые до сих пор, были “правильно” сбалансированы и можно было легко получать несмешанные оценки для эффектов. Так как эти оценки на практике обычно просто интерпретируются, то желательно, если возможно, иметь дело лишь со сбалансированными планами. Несбалансированные гнездовые планы редко обсуждаются в литературе, но в реальности такие планы могут появиться.

В работе Milliken and Johnson (1984, стр. 385) обсуждается пример таких планов. Пяти пациентам, страдающим депрессией, были назначены два разных вида лечения (*Лечение: плацебо и лекарство) – (Treatment*: *Placebo* vs. *Drug).* Пациенты обследовались после одной недели лечения и после пяти недель. Результаты обследований (замеренные в некоторой шкале степени депрессивности), образуют зависимую переменную. Здесь представлены данные обследования (два пациента не явились на повторное обследование):

Пациент	Лечение	Неделя 1	Неделя 2
1	Плацебо	24	18
2	Плацебо	22	--
3	Лекарство	25	22
4	Лекарство	23	--
5	Лекарство	26	24

Неполные планы с более или менее случайным расположением пустых ячеек можно анализировать с помощью анализа контрастов (см. ниже), но здесь будет продемонстрирована и другая техника. В данном примере, при вводе файла данных, были введены три фиктивные переменные (с кодами, которые однозначно определяют сравнения между субъектами внутри каждого вида лечения). Файл данных *Depresse.sta* представлен ниже.


Данные: DEPRESSE.STA 7п * 8н

Example of an unbalanced nested design; see Milliken &

	1 PATIENT	2 TREATMNT	3 WEEK	4 SCORE	5 P_INTRT1	6 P_INTRT2	7 P_INTRT3
1	1	PLACEBO	1	24	1	0	0
2	1	PLACEBO	2	18	1	0	0
3	2	PLACEBO	1	22	-1	0	0
4	3	DRUG	1	25	0	1	1
5	3	DRUG	2	22	0	1	1
6	4	DRUG	1	23	0	-1	0
7	5	DRUG	1	26	0	0	-1
8	5	DRUG	2	24	0	0	-1

Видно, что переменная *5* (*P_IntrT1*) задает сравнение между первым и вторым пациентом (проходящими лечение *Плацебо – Placebo),* переменные *6* и *7* (*P_IntrT2* и *P_IntrT3*) задают два возможных сравнения между пациентами 3, 4, и 5 (получающими *Лекарство – Drug).*

Для получения таблицы дисперсионного анализа со всеми компонентами дисперсии, напечатанной на стр. 394 книге Milliken and Johnson, нужно использовать эти фиктивные переменные как ковариаты и задать план как простой 2 (*Лечение - Treatment*) x 2 (*Неделя - Week*) межгрупповой план с ковариатами (см. стр.1751 для примера плана с ковариатами).

Copyright © StatSoft, 1995

StatSoft®

Заметим, что конструкция критериев значимости для этого плана не так проста, и нужно обратиться к работе Milliken and Johnson для подробного их описания.

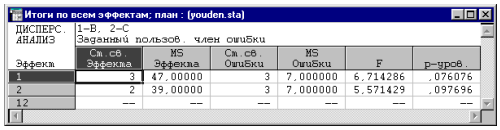
Пример 8: Планы типа квадрата Юдена

Стратегия, примененная в примере 7 (см. выше), будет использована еще в одном примере. План типа квадрата Юдена иногда используется для уравновешивания влияния расположения комбинации условий таким образом, чтобы при этом не смешивались никакие пары исследуемых факторов. Здесь представлен пример 4 x 4 плана типа квадрата Юдена с тремя факторами *A*, *B*, и *C*. Числа в скобках – это данные из примера Lindman (1974, стр. 209):

	B1	B2	B3	B4
C1	A1(15)	A2(7)	A3(14)	A4(8)
C2	A2(17)	A3(6)	A4(8)	A1(7)
C3	A3(22)	A4(14)	A1(11)	A2(15)

Видно, что уровни фактора *A* “чередуются” с уровнями фактора B. Для анализа этих данных были введены три фиктивных переменных с кодами, которые задавали три сравнения с 4 уровнями фактора A (см. файл данных *Youden.sta*). Эти фиктивные переменные затем использовались как ковариаты. План рассматривался как полный 4 (*B*) x 3 (*C*) план ковариационного анализа.

Ниже представлена таблица всех эффектов. При этом взаимодействия *B* x *C* отнесены к членам ошибки (см. также таблицу результатов в книге Lindman, стр. 209):


Итоги по всем эффектам; план : (youden.sta)

ДИСПЕРС. АНАЛИЗ 1-B, 2-C Заданный пользоб. член ошибки

Эффекм	См.сб. Эффекма	MS Эффекма	См.сб. Ошибки	MS Ошибки	F	p-уров.
1	3	47,00000	3	7,000000	6,714286	,076076
2	2	39,00000	3	7,000000	5,571429	,097696
12	--	--	--	--	--	--

Заметим, что для исследования значимости фактора A была использована опция *Регрессия внутри ячеек* из диалогового окна *Результаты дисперсионного анализа*.

Планы с пропущенными ячейками

В следующих примерах экспериментальные планы обычно “предполагаются” полными, однако по тем или иным причинам содержат некоторое количество пропущенных ячеек. В отличие от планов, рассмотренных ранее, расположение пропущенных ячеек в таких планах, не планируется заранее (для того, чтобы можно было легче исследовать главные эффекты), а является случайным. Часто трудно определить, как в таких планах оценивать главные эффекты и взаимодействия.

Как обсуждалось ранее (см. стр.1729) в таких случаях нужно аккуратно изучить расположение наблюдаемых и пропущенных ячеек. Затем сформулировать частные гипотезы (в терминах сравнений средних по ячейкам), которые аппроксимируют гипотезы настолько близко, насколько это возможно. Подробное описание того, как подбирать такие сравнения выходит за рамки нашего изложения. Детальное описание этого вопроса можно найти в Dodge (1985), Milliken и Johnson (1984), и Searle (1987).

Пример 9: План 3 x 3 с пропущенными ячейками

Милликен и Джонсон (Milliken и Johnson 1984, стр. 173) представляют следующий план 3 x 3 с пропущенными ячейками и не равным количеством наблюдений в ячейке.

StatSoft®

Copyright © StatSoft, 1995

Фактор T	Фактор B: B1	B2	B3
T1	2,6	пропущена	8,6
T2	3	14	12,9
T3	6	9	пропущена

Для дальнейшего обсуждения, ячейки пронумерованы по строкам (пропущенные ячейки заключены в круглые скобки).

Фактор T	Фактор B: B1	B2	B3
T1	Ячейка 1	(Ячейка 2)	Ячейка 3
T2	Ячейка 4	Ячейка 5	Ячейка 6
T3	Ячейка 7	Ячейка 8	(Ячейка 9)

Легко заметить, что ячейки 2 и 9 пропущены в этом плане. Как могут быть оценены главные эффекты и взаимодействия? Начнем с фактора *T*. Для того чтобы не смешивать оценку главного эффекта фактора *T* с взаимодействием, нужно быть уверенным в том, что сравниваются средние по уровням *T* внутри уровней фактора *B*. Можно добиться этого, разбивая план на два полных факторных эксперимента:

План 1:

T	B: B1	B3
T1	Ячейка 1	Ячейка 3
T2	Ячейка 4	Ячейка 6

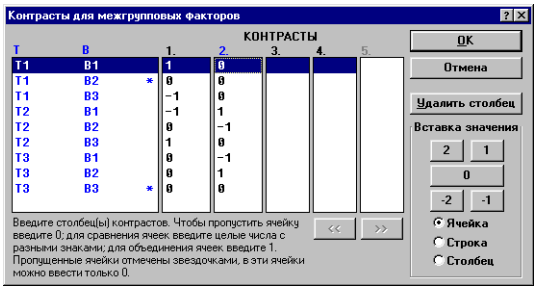
План 2:

T	B: B1	B2
T1	Ячейка 4	Ячейка 5
T2	Ячейка 7	Ячейка 8

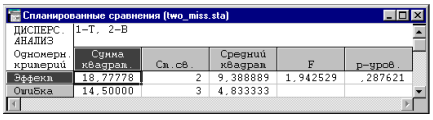
Внутри каждого из этих планов можно составлять сравнения для исследования главных эффектов и взаимодействий обычным способом. Для фактора *T* это может быть проведено с помощью сравнений: (a) Ячейка 1 + Ячейка 3 с Ячейка 4 + Ячейка 6; и (b) Ячейка 4 + Ячейка 5 с Ячейка 7 + Ячейка 8.

Аналогично, для исследования главного эффекта фактора *B* можно сравнить: (a) Ячейка 1 + Ячейка 4 с Ячейка 3 + Ячейка 6; и (b) Ячейка 4 + Ячейка 7 с Ячейка 5 + Ячейка 8. Наконец, взаимодействие может быть оценено сравнением: (a) Ячейка 1 + Ячейка 6 с Ячейка 4 + Ячейка 3; и (b) Ячейка 4 + Ячейка 8 с Ячейка 7 + Ячейка 5.

В примере 2 (стр. 1747) используется контрастный анализ (см. обсуждение на стр. 1723 и 1785). Заметим, что в рассматриваемом примере контрастные коэффициенты должны быть введены как один длинный вектор (для этого нужно щелкнуть на кнопке *Опции* в диалоговом окне *Результаты дисперсионного анализа* и установить опцию *Вводить контрасты для всех факторов вместе*). Данные примера 9 были введены в файл *Two_miss.sta*. Внизу представлены контрастные коэффициенты для взаимодействий.



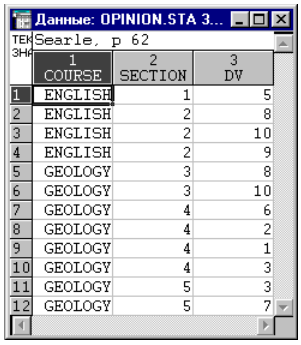
Если в приведенном выше диалоговом окне щелкнуть *OK*, то появится следующая таблица.



Пример 10: Взвешенные средние Анализ гнездового плана с неравным числом наблюдений в ячейках

Этот пример взят из книги Searle (1987, стр. 62). Данные были введены в файл *Opinion.sta* (показан ниже).

Copyright © StatSoft, 1995



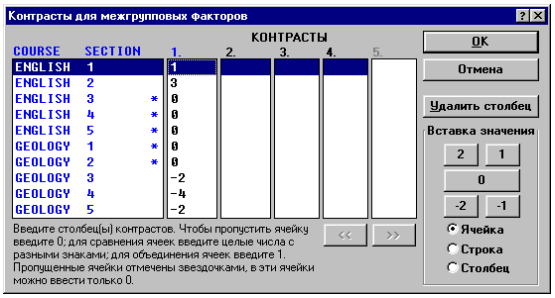
	1 COURSE	2 SECTION	3 DV
1	ENGLISH	1	5
2	ENGLISH	2	8
3	ENGLISH	2	10
4	ENGLISH	2	9
5	GEOLOGY	3	8
6	GEOLOGY	3	10
7	GEOLOGY	4	6
8	GEOLOGY	4	2
9	GEOLOGY	4	1
10	GEOLOGY	4	3
11	GEOLOGY	5	3
12	GEOLOGY	5	7

Даже если это и гнездовой план, вложение факторов не было спланировано экспериментатором, а получилось из данных “само собой”. Представленные данные являются результатом изучения отношения студентов к компьютерам. В эксперименте участвовали студенты двух разных курсов. Фактор *Курс – Course* имеет два уровня: *Геология - Geology* и *Английский - English.* На геологическом курсе имеются две разные группы, а на курсе английского языка – три. Во всех группах преподают разные преподаватели. Видно, что представленные данные образуют двухфакторный гнездовой (вложенный) план. Структура эксперимента и данные могут быть представлены следующим образом:

```
Секция
Курс     1      2      3        4       5
-------------------------------------------
Англ.яз.5  8,10,9
Геология              8,10   6,2,1,3   3,7
```

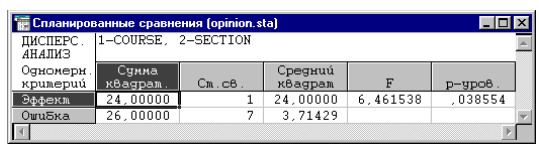
Для исследования главного эффекта фактора *Курс — Course*, в работе Searle использовано сравнение взвешенных средних. Заметим, что в этом примере неравномерное распределение числа наблюдений в ячейках (1, 3, 2, 4, 2, если идти по таблице слева направо, см. выше), отвечало не выбору экспериментатора, а скорее отражало действительное распределение студентов по курсам и группам. Так как результаты эксперимента желательно обобщить на всех студентов, то в этом случае разумно рассматривать взвешенные средние для фактора *Курс—Course*, а не просто средние.

Для вычисления таблицы сумм квадратов, напечатанной в Searle (1987, стр. 71) сначала нужно задать план так, если бы он был полным. Затем нужно ввести следующие контрастные коэффициенты для фактора *Курс—Course*, используя число наблюдений в ячейках в качестве весов:



Заметим, что в этом случае сумма контрастных коэффициентов не равна нулю. Модуль *Дисперсионный анализ* автоматически подправляет контрастные коэффициенты так, что они будут в сумме давать нуль (за исключением случая, когда контрастные коэффициенты состоят только из положительных целых чисел, тогда рассматриваемая гипотеза является гипотезой об общем среднем).

Ниже показан результат этого сравнения.



Заметим, что (как сказано в Searle, стр. 71), общее среднее и главный эффект фактора *Секция—Section* могут быть исследованы аналогично.

StatSoft®
Copyright © StatSoft, 1995

Пример 11:
Гнездовой план 4 x 11
с неравным количеством уровней

Следующий пример - пример гнездового плана, когда разное число уровней фактора вкладывается в разные уровни другого фактора. И вновь, такие планы лучше всего исследовать с помощью анализа контрастов..

Милликен и Джонсон (Milliken и Johnson 1984, стр. 415) представляют результаты сравнения 11 различных средств от насекомых, произведенных различными компаниями.

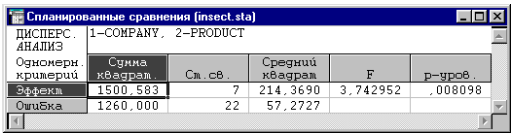
```
                Продукт
Компания  1 2 3 4 5 6 7 8 9 10 11
--------------------------------
   A      x x x
   B            x x
   C                x x
   D                    x x  x  x
```

Одна компания производит три различных средства от насекомых, другая—четыре, оставшиеся производят по два типа инсектицида. План эксперимента представлен выше (через X обозначены наблюдаемые комбинации вариантов факторов).

Данные из книги Милликена и Джонсона содержатся в файле *Insect.sta*. Видно, что главный эффект для фактора *Компания—Company* может быть оценен обычным способом. Однако для исследования фактора *Продукт—Product* нужно использовать спланированные сравнения. Точнее, можно задать два контраста для сравнения трех продуктов, произведенных компанией *A*, один контраст для продуктов, произведенных компанией *B*, один — для *C* и три контраста для четырех продуктов, произведенных компанией *D* - всего 7 контрастов.

Для того чтобы построить эти контрасты, щелкните на кнопке *Опции* в диалоговом окне *Результаты дисперсионного анализа* и установите опцию *Вводить контрасты для всех факторов вместе*. Затем щелкните на кнопке *Спланированные сравнения* и в появившемся диалоговом окне введите подходящие коэффициенты.

Заметим, что пропущенные ячейки в плане отмечены звездочками, для этих ячеек можно вводить только *0.* Ниже показана таблица с результатами одновременного исследования 7 контрастов.



Заметим, что эти результаты слегка отличаются от тех, которые напечатаны в книге Милликена и Джонсона (стр. 422). Содержащиеся там результаты противоречивы и, возможно, имеют опечатку (например, сравните среднее, для последней группы напечатанное на стр.417 с данными на стр. 415).

Пример 12:
Сложные планы с большим количеством пропущенных ячеек (Исследование гипотез типа IV)

Милликен и Джонсон (Milliken and Johnson 1984, стр. 202) обсуждают пример сложного плана с большим количеством пропущенных ячеек. Проводится исследование питания. План исследования содержит 3 фактора: *Группа - Group* (2 уровня: получал или нет субъект *талоны на бесплатное питание - food stamps*); *Возраст - Age* (субъекты разбиваются на три возрастные группы); *Раса-Race* (*черные - black*, *испанцы - hispanic*, *белые - white*). Зависимая переменная – это изменение некоторой характеристики до и после исследования.

Copyright © StatSoft, 1995

StatSoft®

Распределение пропущенных ячеек и число наблюдений в ячейках выглядят следующим образом:

Возрастная Группа	(Талоны)	Раса Черные	Испанцы	Белые
1	Нет			3
	Да	2		2
2	Нет	7		8
	Да			8
3	Нет	6	2	20
	Да	4	1	31
4	Нет	1		
	Да	1		11

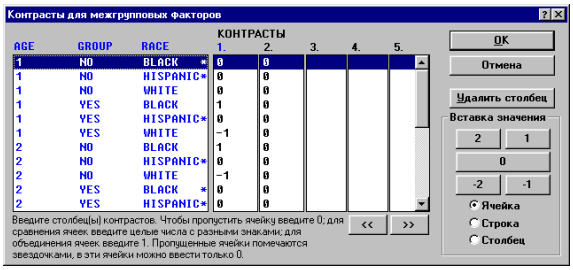
Теперь рассмотрим главный эффект фактора *Раса - Race*. В примере 9 кратко обсуждается стратегия построения множества сравнений для отдельного эффекта, который сбалансирован по всем факторам не включенным в эффект. Для исследования главного эффекта нужно было бы исследовать распределение наблюденных и пропущенных ячеек и попытаться выделить "подплан", который уже будет полон относительно рассматриваемого фактора.

Фактор *Раса -Race* имеет 3 уровня. Для изучения этого фактора можно выделить два подплана (как и раньше, в таблице содержатся числа наблюдений в соответствующих ячейках).

Подплан 1: Возраст. Группа	(Талоны)	Раса Черные	Белые
1	Да	2	2
2	Нет	7	8
3	Нет	6	20
	Да	4	31
4	Да	1	11

Подплан 2: Возраст. Группы	(Талоны)	Раса Испанцы	Белые
3	Нет	2	20
	Да	1	31

По этим двум подпланам можно строить контрасты главных эффектов фактора *Раса - Race* (см. также Таблицу 17.3 в Milliken and Johnson, 1984, стр. 204). Ниже показаны результаты для данных, содержащихся в файле *Food.sta.* Для того чтобы ввести контрасты, показанные ниже, нужно сначала в диалоговом окне *Результаты дисперсионного анализа* щелкнуть на кнопку *Спланированные сравнения* (сначала в диалоговом окне *Результаты дисперсионного анализа* нужно нажать на кнопку *Опции,* и затем выбрать опцию *Вводить для всех факторов вместе*, а потом ввести контрастные коэффициенты, как показано ниже).



Полное множество контрастных коэффициентов представлено ниже.

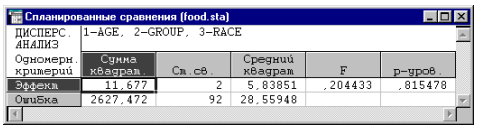
Контраст 1:

```
0 0 0 1 0 -1 1 0 -1 0 0 0
1 0 -1 1 0 -1 0 0 0 1 0 -1
```

Контраст 2:

```
0 0 0 0 0 0 0 0 0 0 0 0
0 1 -1 0 1 -1 0 0 0 0 0 0
```

Если щелкнуть на *OK* в этом диалоговом окне, появится таблица результатов дисперсионного анализа.



Рассмотрим теперь взаимодействие факторов *Раса - Race* и *Группа - Group*. Рассмотрим опять распределение ячеек с наблюдениями, представленное выше. Видно, что полные подпланы для исследования взаимодействия указанных факторов возможны лишь, когда фактор *Возраст - Age = 3.* Два подплана выглядят следующим образом (в скобках добавлены значения соответствующих контрастов для эффекта взаимодействия):

StatSoft®

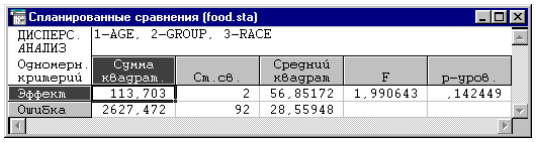
Copyright © StatSoft, 1995

```
Подплан 1:                  Раса
Возраст.              ---------------
Группа   (Талоны)   Черные    Белые
  3        Нет        6(+1)   20(-1)
           Да         4(-1)   31(+1)
Подплан 2:                  Раса
Возраст.              ----------------
Группа   (Талоны)   Испанцы    Белые
  3        Нет        2(+1)    20(-1)
           Да         1(-1)    31(+1)
```

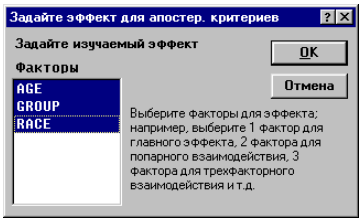
Заметим, что в таблице, представленной в Milliken and Johnson (1984, стр. 204), имеется опечатка в первой гипотезе для взаимодействия факторов *Группа - Group* и *Раса - Race* (положительный вес должен быть у последних средних). Нужно теперь щелкнуть на кнопке *Спланированные сравнения* и ввести соответствующие контрасты (см. ниже) в диалоговом окне *Контрасты для межгрупповых факторов*.

```
Контраст 1:

0 0 0 0 0 0 0 0 0 0 0 0
1 0 -1 -1 0 1 0 0 0 0 0 0

Контраст 2:

0 0 0 0 0 0 0 0 0 0 0 0
0 1 -1 0 -1 1 0 0 0 0 0 0
```

Если теперь в диалоговом окне *Контрасты для межгрупповых факторов* щелкнуть на *OK*, появится следующая таблица результатов дисперсионного анализа.

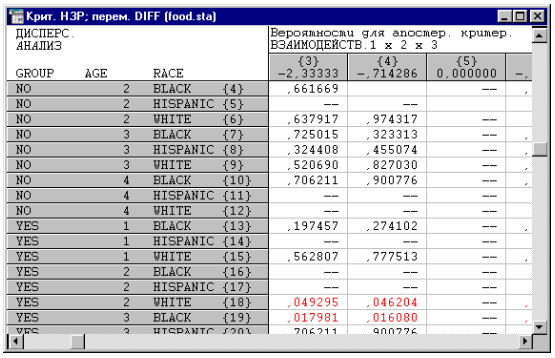


Эти результаты совпадают с результатами анализа суммы квадратов *типа IV,* проведенным в Milliken and Johnson (Таблица 17.2, стр. 203). До известной степени, анализ, представленный здесь, точно такой же, как если бы план был однофакторным планом дисперсионного анализа, в котором исследуются отдельные гипотезы о разности между средними. Следуя этому пути, было бы полезно рассмотреть сравнение сразу всех средних. Это можно выполнить, щелкнув на кнопку *Апостериорные сравнения* в диалоговом окне *Результаты дисперсионного анализа* и затем задать взаимодействие *123*.



После этого можно щелкнуть OK и в диалоговом окне результатов *Апостериорные сравнения средних* выбрать НЗР критерий или опцию спланированные сравнения.

Ниже показана таблица результатов (см. также Milliken and Johnson, таблица 17.5, стр. 206; заметим, что факторы были переупорядочены так, чтобы порядок средних совпадал с порядком представленным в книге Милликена и Джонсона):



Copyright © StatSoft, 1995

StatSoft®

ТЕХНИЧЕСКИЕ ЗАМЕЧАНИЯ

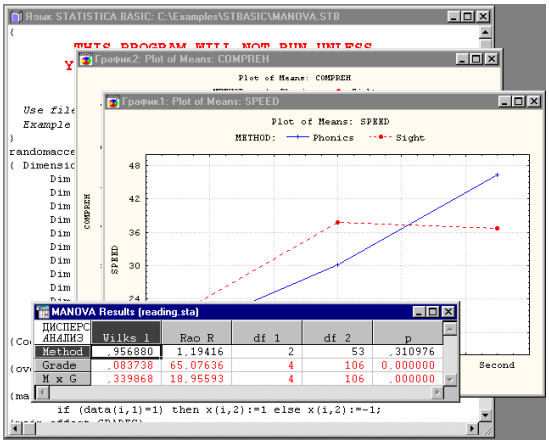
Вычислительные процедуры

Дисперсионный анализ позволяет исследовать общую линейную модель (см. Bock, 1975; Finn, 1974, 1977; Hays, 1981; Lindeman, Merenda, and Gold, 1980). Программа сначала вычисляет ковариационную матрицу внутри ячеек для зависимых переменных (и ковариат, см. ниже). Матрица плана для главных эффектов (или матрица для контрастных коэффициентов) сначала ортонормируется (см. Bjorck, 1967), а затем используется для вычисления суммы квадратов гипотез (со средними по ячейкам) и сумм квадратов ошибок (с ковариационной матрицей внутри ячеек). Если план содержит ковариаты, то они добавляются в ковариационную матрицу зависимых переменных. Затем, перед вычислением статистических критериев, средние и ковариационная матрица ошибок корректируются (Dempster, 1969). Эти процедуры подробно описаны в Finn (1974, 1977).

Подробное изложение вычислительных аспектов рассматриваемых моделей можно найти, например, в Bock (1975); Finn (1974); Hocking and Speed (1975); Morrison (1967); Timm (1975), and Timm and Carlson (1973, 1975).

Использование библиотеки матричных функций *STATISTICA BASIC* в многомерном дисперсионном анализе

Используя язык программирования *STATISTICA BASIC,* можно написать собственную процедуру для анализа данных. Этот язык содержит полную библиотеку матричных функций (включая функции для вычисления обратной и псевдообратной матрицы, ортонормирование, разложение по сингулярным значениям и т.д.). В качестве примера такой программы приведена программа *Manova.stb* (в поддиректории *STBASIC*), в которой вычисления многомерного дисперсионного анализа реализованы с помощью нескольких строк (с использованием возможностей матричных функций).



Существует еще несколько примеров программ на языке *STATISTICA BASIC,* включенных в поддиректорию *STBASIC.* Например, программы для проведения преобразований Бокса-Кокса и Бокса-Тидвелла зависимых переменных (*Boxcox.stb* и *Boxtid.stb*, соответственно), программа, осуществляющая оценку методом взвешенных наименьших квадратов (*Wls.stb*), программаоценивания двухступенчатым методом наименьших квадратов (*2stls.sta*) и т. д. Эти программы содержат подробные комментарии.

Многомерный дисперсионный анализ и моделирование структурными уравнениями

В последнее время моделирование структурными уравнениями широко используется в качестве

StatSoft®

Copyright © StatSoft, 1995

альтернативы многомерного дисперсионного анализа (см., например, Bagozzi and Yi, 1989; Bagozzi, Yi, and Singh, 1991; Cole, Maxwell, Arvey, and Salas, 1993). Этот подход позволяет не только подробно исследовать гипотезы о средних в разных группах, но и гипотезы о различии структуры корреляционных матриц зависимых переменных. Например, можно ослабить предположение об однородности ковариационных матриц и явно включить в модель различные ковариационные матрицы ошибок для каждой группы. Модуль *Моделирование структурными уравнениями* (см. том III) позволяет выполнить такой анализ, см. соответствующую главу описания. Там же обсуждаются примеры анализа так называемых структурных моделей для нескольких групп данных.

Copyright © StatSoft, 1995

StatSoft®

ПРЕДМЕТНЫЙ УКАЗАТЕЛЬ

А

Анализ взвешенных средних в гнездовом плане с неравным числом наблюдений в ячейках, 1815
Анализ контрастов, 1724, 1781, 1784, 1785, 1786, 1803, 1817
Апостериорные сравнения, 1724, 1764, 1769

В

Вектор контрастов, 1764, 1781, 1783
Взаимодействия высших порядков, 1719
Влияние ковариат на F критерий, 1721
Внутригрупповые планы (с повторными измерениями), 1720, 1723, 1732, 1758, 1766, 1790, 1798

Г

Гипер-греко-латинские, 1779
Гипотезы модели средних, 1765
Гистограммы категоризованные, 1772
Главные эффекты, 1718
Гнездовые (вложенные) планы, 1720, 1759, 1761, 1797, 1807, 1808, 1809, 1817
Гнездовые (вложенные) планы с неравным числом наблюдений в ячейках, 1815
График ствол и листья, 1775, 1777
Графики взаимодействия
  график взаимодействия для средних, 1780
  назначение факторов осям графика, 1769
Греко-латинские квадраты, 1779, 1799

Д

Двухфакторные взаимодействия, 1718
Диаграммы размаха
  категоризованные диаграммы размаха, 1772
Диаграммы рассеяния
  для зависимых переменных и ковариат, 1775
  для средних и стандартных отклонений, 1774
  категоризованные диаграммы рассеяния, 1773
Дискриминантная функция, 1755
Дискриминантный анализ, 1753, 1755, 1767
Дисперсионный анализ, 1715
  анализ контрастов, 1724, 1781
  апостериорные сравнения, 1764, 1769
  взаимодействия высших порядков, 1719
  внутригрупповые планы, 1720, 1723, 1732, 1758, 1760, 1766
  гипотеза параллельности, 1755, 1773
  гипотезы модели средних, 1765
  главные эффекты, 1718
  гнездовые (вложенные) планы, 1720, 1759, 1761
  зависимые и независимые переменные, 1717
  изолированная контрольная группа, 1759, 1761
  исследование предположений, 1730, 1732, 1741, 1755, 1771, 1773, 1776
  ковариаты, 1758, 1762
  коррелированность средних и дисперсий, 1731
  латинские квадраты, 1720, 1778, 1779
  матричные функции *STATISTICA BASIC*, 1820
  меж- и внутригрупповые планы, 1720, 1795
  межгрупповые планы, 1720, 1784
  многомерная регрессия и дисперсионный анализ, 1724
  многомерные критерии, 1722, 1732, 1753
  многомерные планы с повторными измерениями, 1760
  многомерный дисперсионный анализ, 1722, 1732, 1767, 1784
  многомерный ковариационный анализ (MANCOVA), 1751
  многофакторный дисперсионный анализ, 1717
  модели структурных средних, 1820
  моделирование структурными уравнениями и многомерный дисперсионный анализ, 1820
  модель дисперсионного анализа для средних по ячейкам, 1727, 1729
  независимые гипотезы при повторных измерениях, 1733
  необходимость независимости гипотез, 1732
  неполные планы, 1720, 1728, 1778, 1806
  несбалансированные и сбалансированные планы, 1725
  объединенные эффекты и члены ошибки, 1764
  одномерные критерии, 1753
  однородность дисперсии, 1730, 1773
  однородность дисперсии и ковариаций, 1731, 1773
  основная логика дисперсионного анализа, 1717
  отклонения от предположений, 1730
  оценивание двухступенчатым методом наименьших квадратов, 1820
  оценивание методом взвешенных наименьших квадратов, 1820
  переменные ковариаты, 1722, 1758, 1762, 1803

StatSoft®

Copyright © StatSoft, 1995

планы с несколькими случайными эффектами, 1762
планы с повторными измерениями, 1720, 1732, 1758, 1760, 1766
планы с пропущенными ячейками, 1728
преобразования Бокса-Кокса и Бокса-Тидвелла, 1820
проверка предположений, 1735
проверка специфического эффекта, 1767
пропущенные ячейки и объединенные эффекты/члены ошибки, 1730
пропущенные ячейки и проверка специфического эффекта, 1729
регрессионный подход в дисперсионном анализе, 1728, 1759
регрессия внутри ячеек, 1764
сложные планы, 1719
случайные эффекты, 1759, 1762, 1801, 1809
смешанные модели, 1759, 1762
спланированные сравнения, 1724, 1728, 1763, 1770, 1781
сумма квадратов, 1726
сумма квадратов типа I, 1724, 1726, 1765
сумма квадратов типа II, 1726, 1765
сумма квадратов типа III, 1726, 1765
сумма квадратов типа IV, 1727, 1765
сферичность и сложная симметрия, 1732
фиксированные ковариаты, 1722, 1758, 1762, 1803
эффекты взаимодействия, 1718
Дисперсионный анализ с многомерными повторными измерениями (*MANOVA*), 1792

З

Зависимые переменные, 1717
Задание контрастов, 1781

И

Изолированная контрольная группа, 1759, 1761, 1788
Исследование главных эффектов и взаимодействий, 1786, 1791

К

Канонический анализ, 1753
Категоризованные гистограммы, 1772
Категоризованные графики
диаграммы рассеяния, 1773
нормальные вероятностные графики, 1772
Категоризованные диаграммы размаха, 1772
Ковариаты
взаимодействие между ковариатами и факторами, 1722
вычисление скорректированных средних, 1721
множество ковариат, 1721
переменные ковариаты, 1722, 1758, 1762
фиксированные ковариаты, 1721, 1722, 1758, 1762
Ковариационный анализ (ANCOVA), 1721, 1751
вводный обзор, 1721
взаимодействие между ковариатами, 1722
влияние ковариат на F критерий, 1721
множество ковариат, 1721
опции вывода, 1766
переменные ковариаты, 1722, 1758, 1762
результаты, 1767
скорректированные средние, 1721, 1753, 1766, 1767, 1802
фиксированные ковариаты, 1721, 1762
Контрольная группа, 1759, 1761
Критерий ДЗР Тьюки, 1771
Критерий ДЗР Тьюки для выборок неравных объемов, 1771
Критерий Левена, 1735, 1773, 1777
Критерий множественных сравнений Дункана, 1770
Критерий НЗР, 1770
Критерий Ньюмена-Кеулса, 1770
Критерий Спйотвола и Столайна, 1771
Критерий сферичности Моучли, 1766
Критерий Хартли, 1735, 1773, 1777
Критерий Шеффе, 1770
Критические размахи для критерия множественных сравнений Дункана, 1770
Критические размахи для критерия Ньюмена-Кеулса, 1770

Л

Латентные корни, 1753
Латинские квадраты, 1720, 1779

М

Матрица диаграмм рассеяния, 1776
Матрицы сумм квадратов, 1735, 1750, 1767
Матричные функции *STATISTICA BASIC*, 1820
Меж- и внутригрупповые многомерные планы, 1795
Меж- и внутригрупповые планы, 1720, 1795
Межгрупповые гнездовые планы, 1797
Межгрупповые планы, 1720, 1784
М-критерий Бокса, 1777
Многомерные критерии
М-критерий Бокса, 1777
непараметрический критерий Сена и Пури, 1777
Многомерные межгрупповые планы, 1722, 1784
Многомерные многофакторные внутригрупповые планы (с повторными измерениями), 1795
Многомерные планы, 1722
Многомерные планы с повторными измерениями, 1760
Многомерный дисперсионный анализ, 1785
межгрупповые планы, 1722
планы с повторными измерениями, 1723
суммирование значений переменных, 1723

Copyright © StatSoft, 1995

Многомерный дисперсионный анализ зависимых измерений, 1797
Многомерный ковариационный анализ (MANCOVA), 1751
Многофакторные планы с повторными измерениями, 1792
Многофакторный дисперсионный анализ, 1717
Многофакторный межгрупповой дисперсионный анализ, 1786
Модели структурных средних, 1820
Моделирование структурными уравнениями и многомерный дисперсионный анализ, 1820
Модель дисперсионного анализа для средних по ячейкам, 1727, 1729, 1765

Н

Назначение факторов осям графика, 1769
Независимые переменные, 1717
Непараметрический критерий Сена и Пури, 1777
Неполные планы, 1720, 1728, 1778, 1806
Несбалансированные планы, 1725, 1806
Нормальные вероятностные графики
  категоризованные нормальные вероятностные графики, 1772

О

Обобщенные латинские квадраты, 1799
Объединенные эффекты и члены ошибки, 1730, 1764, 1778, 1779, 1800, 1808
Одномерные межгрупповые планы, 1784
Однородность дисперсии, 1776
Однородность дисперсии и ковариаций, 1773
Однофакторный дисперсионный анализ
  с многомерными повторными измерениями, 1792
  с повторными измерениями, 1790
Однофакторный межгрупповой дисперсионный анализ, 1784
Опции вывода, 1765
Ортогональные эффекты, определения, 1725
Оценивание двухступенчатым методом наименьших квадратов, 1820
Оценивание методом взвешенных наименьших квадратов, 1820

П

Параллельность регрессионных линий/плоскостей, 1755, 1773
Переменные ковариаты, 1722, 1803
Переупорядочивание эффектов для сумм квадратов типа I, 1780
Планы латинские квадраты, 1799
Планы с несколькими случайными эффектами, 1762
Планы с одним наблюдением в ячейке, 1800
Планы с повторными измерениями, 1720, 1732, 1758, 1760, 1766, 1790, 1792, 1795, 1809
Планы с полосными делянками, 1812
Планы с пропущенными ячейками, 1728
Планы с расщепленными делянками, 1811
  с неравным числом подделянок, 1813
Планы типа квадрата Юдена, 1814
Полиномиальные контрасты, 1782, 1804
Поправка Гринхауса-Гейзера, 1750, 1766
Поправка Хюнха-Фельдта, 1750, 1766
Последовательные суммы квадратов, 1779
Постоянные ковариаты (см. также фиксированные ковариаты), 1802
Правила формирования контрастов, 1785
Предопределенные контрасты, 1782
Предположения и эффекты
  нарушение предположения о сферичности и сложной симметрии, 1732
  нарушение предположения об однородности дисперсии, 1730
  нарушение предположения об однородности дисперсии и ковариаций, 1731
  отклонение от нормальности распределения, 1730
Преобразования Бокса-Кокса и Бокса-Тидвелла, 1820
Проверка значимости, 1716
Процедура множественных сравнений, 1724

Р

Разбиение суммы квадратов, 1715
Регрессионный подход в дисперсионном анализе, 1727, 1728, 1759, 1765
Регрессия внутри ячеек, 1764
Результаты для ковариат, 1767

С

Сбалансированные неполные планы, 1729, 1799
Сбалансированные планы, 1725
Скорректированные средние, 1721, 1753, 1766, 1767, 1802
Сложная симметрия, 1732
Сложные планы, 1719
Случайные блочные планы, 1800
Случайные эффекты, 1759, 1762, 1801, 1809
Смешанные модели, 1759, 1762
Спланированные сравнения, 1724, 1728, 1763, 1770, 1781, 1784, 1785, 1786, 1789, 1791, 1794, 1796, 1803, 1817
Сумма квадратов ошибок, 1715, 1716, 1724
Сумма квадратов типа I, 1726, 1765, 1779, 1806
Сумма квадратов типа II, 1726, 1765, 1806
Сумма квадратов типа III, 1726, 1765
Сумма квадратов типа IV, 1726, 1727, 1729, 1765, 1817
Сумма квадратов эффекта, 1716
Сферичность и сложная симметрия, 1732, 1766

StatSoft®

Copyright © StatSoft, 1995

У

Уровень значимости для таблицы
всех эффектов, 1780

Ф

Фиксированные ковариаты, 1721,
1802

Ц

Цель дисперсионного анализа, 1715

Э

Эффекты взаимодействия, 1718
высших порядков, 1719
двухфакторные, 1718

Copyright © StatSoft, 1995

StatSoft®

StatSoft®

Copyright © StatSoft, 1995

Приложения:

Оглавление

Приложение I: функции распределения ... 1831
Генерирование случайных чисел, имеющих заданное распределение ... 1831
Бета распределение ... 1831
Биномиальное распределение ... 1833
Распределение Коши ... 1833
Распределение *хи-квадрат* ... 1834
Экспоненциальное распределение ... 1834
Распределение экстремального значения (Гумбеля) ... 1835
F распределение ... 1837
Гамма распределение ... 1838
Геометрическое распределение ... 1839
Распределение Лапласа ... 1839
Логистическое распределение ... 1840
Логнормальное распределение ... 1841
Нормальное распределение ... 1842
Распределение Парето ... 1843
Распределение Пуассона ... 1844
Распределение Релея ... 1844
t распределение Стьюдента ... 1846
Распределение Вейбулла ... 1846
Приложение II: итоговые статистики ... 1848
Сумма ... 1848
Минимум ... 1848
Максимум ... 1848
Среднее ... 1848
Доверительные границы для среднего ... 1849
Размах ... 1849
Процентили ... 1849
Дисперсия ... 1851
Стандартная ошибка среднего ... 1851
Эксцесс ... 1851
Стандартная ошибка эксцесса ... 1852
Асимметрия ... 1852
Стандартная ошибка асимметрии ... 1852
Предметный указатель ... 1853

Подробное оглавление см. на следующей странице.

Copyright © StatSoft, 1995

StatSoft®

Подробное оглавление

ПРИЛОЖЕНИЕ I: ФУНКЦИИ РАСПРЕДЕЛЕНИЯ....1831
Генерирование случайных чисел, имеющих заданное распределение 1831
Бета распределение 1831
Бета распределение для формул таблиц результатов, *STATISTICA BASIC* и определенных пользователем функций 1831
Бета распределение в *Вероятностном калькуляторе* 1832
Бета распределение в *Квантиль-Квантиль* графиках 1832
Бета распределение в графиках *Вероятность- Вероятность* 1832
Биномиальное распределение 1833
Биномиальное распределение в формулах таблиц результатов, *STATISTICA BASIC* и в определенных пользователем функциях 1833
Распределение Коши 1833
Распределение Коши в формулах таблиц результатов, в *STATISTICA BASIC* и определенных пользователем функциях 1833
Распределение Коши в *Вероятностном калькуляторе* 1833
Распределение *хи-квадрат* 1834
Распределение *хи-квадрат* в формулах таблиц результатов, в *STATISTICA BASIC* и определенных пользователем функциях 1834
Распределение *хи-квадрат* в Вероятностном калькуляторе 1834
Экспоненциальное распределение 1834
Экспоненциальное распределение в формулах таблиц результатов, в *STATISTICA BASIC* и определенных пользователем функциях 1834
Экспоненциальное распределение в *Вероятностном калькуляторе* 1835
Экспоненциальное распределение в графиках *Квантиль-Квантиль* 1835
Экспоненциальное распределение в графиках *Вероятность- Вероятность* 1835
Распределение экстремального значения (Гумбеля) 1835
Распределение экстремального значения в формулах таблиц результатов, в *STATISTICA BASIC* и определенных пользователем функциях 1836
Распределение экстремального значения в *Вероятностном калькуляторе* 1836
Распределение экстремального значения в графиках *Квантиль-Квантиль* 1836
Распределение экстремального значения в графиках *Вероятность- Вероятность* 1836
F распределение 1837
F распределение в формулах таблиц результатов, в *STATISTICA BASIC* и определенных пользователем функциях 1837
F распределение в *Вероятностном калькуляторе* 1837
Гамма распределение 1838
Гамма распределение в формулах таблиц результатов, в *STATISTICA BASIC* и определенных пользователем функциях 1838
Гамма распределение в *Вероятностном калькуляторе* 1838
Гамма распределение в графиках *Квантиль-Квантиль* 1838

StatSoft®
Copyright © StatSoft, 1995

Гамма распределение в графиках *Вероятность- Вероятность* 1839
Геометрическое распределение 1839
Геометрическое распределение в формулах таблиц результатов, в *STATISTICA BASIC* и определенных пользователем функциях 1839
Распределение Лапласа 1839
Распределение Лапласа в формулах таблиц результатов, *STATISTICA BASIC* и определенных пользователем функциях 1840
Распределение Лапласа в *Вероятностном калькуляторе* 1840
Логистическое распределение 1840
Логистическое распределение в формулах таблиц результатов, в *STATISTICA BASIC* и определенных пользователем функциях 1840
Логистическое распределение в *Вероятностном калькуляторе* 1840
Логнормальное распределение 1841
Логнормальное распределение в формулах таблиц результатов, в *STATISTICA BASIC* и в определенных пользователем функциях 1841
Логнормальное распределение в *Вероятностном калькуляторе* 1841
Логнормальное распределение в графиках *Квантиль-Квантиль* 1841
Логнормальное распределение в графиках *Вероятность- Вероятность* 1842
Нормальное распределение 1842
Нормальное распределение в формулах таблиц результатов, в *STATISTICA BASIC* и определенных пользователем функциях 1842
Нормальное распределение в *Вероятностном калькуляторе* 1842
Нормальное распределение в графиках *Квантиль-Квантиль* 1843
Нормальное распределение в графиках *Вероятность- Вероятность* 1843
Распределение Парето 1843
Распределение Парето в формулах таблиц результатов, в *STATISTICA BASIC* и определенных пользователем функциях 1843
Распределение Парето в *Вероятностном калькуляторе* 1844
Распределение Пуассона 1844
Распределение Пуассона в формулах таблиц результатов, в *STATISTICA BASIC* и определенных пользователем функциях 1844
Распределение Релея 1844
Распределение Релея в формулах таблиц результатов, в *STATISTICA BASIC* и определенных пользователем функциях 1845
Распределение Релея в *Вероятностном калькуляторе* 1845
Распределение Релея в графиках *Квантиль-Квантиль* 1845
Распределение Релея в графиках *Вероятность- Вероятность* 1845
t распределение Стьюдента 1846
t распределение Стьюдента в формулах таблиц результатов, в *STATISTICA BASIC* и определенных пользователем функциях 1846
t распределение Стьюдента в *Вероятностном калькуляторе* 1846
Распределение Вейбулла 1846
Распределение Вейбулла в формулах таблиц результатов, в *STATISTICA BASIC* и определенных пользователем функциях 1847

Copyright © StatSoft, 1995

Распределение Вейбулла в *Вероятностном калькуляторе* ... 1847
Распределение Вейбулла в графиках *Квантиль-Квантиль* ... 1847
Распределение Вейбулла в графиках *Вероятность- Вероятность* ... 1847

ПРИЛОЖЕНИЕ II: ИТОГОВЫЕ СТАТИСТИКИ ... 1848

Число наблюдений *n* ... 1848
Построчное удаление пропущенных данных ... 1848
Попарное удаление пропущенных данных ... 1848
Подстановка среднего вместо пропусков ... 1848
Сумма ... 1848
Минимум ... 1848
Максимум ... 1848
Среднее ... 1848
Доверительные границы для среднего ... 1849
Размах ... 1849
Процентили ... 1849
Метод взвешенного среднего в X_{np} ... 1849
Метод взвешенного среднего в $X_{(n+1)p}$ ... 1850
Метод эмпирической функции распределения ... 1850
Эмпирическая функция распределения с методом усреднения ... 1850
Метод эмпирической функции распределения и интерполяции (MS Excel) ... 1851
Метод ближайшего наблюдения ... 1851
Дисперсия ... 1851
Стандартная ошибка среднего ... 1851
Эксцесс ... 1851
Стандартная ошибка эксцесса ... 1852
Асимметрия ... 1852
Стандартная ошибка асимметрии ... 1852

ПРЕДМЕТНЫЙ УКАЗАТЕЛЬ ... 1853

StatSoft®

Copyright © StatSoft, 1995

Приложения:

ПРИЛОЖЕНИЕ I: ФУНКЦИИ РАСПРЕДЕЛЕНИЯ

STATISTICA предоставляет широкий набор функций плотности распределения, интегральных функций распределения и обратных функций распределения, которые можно использовать в формулах электронных таблиц (см. Глава 3), в *Вероятностном калькуляторе* (см. Глава 10), в *графиках* (см. Том II), в программах *STATISTICA BASIC* (см. *Электронное руководство*), других определяемых пользователем функциях. Ниже описаны функции распределения, доступные в *STATISTICA*.

Генерирование случайных чисел, имеющих заданное распределение

Используя датчик равномерно распределенных чисел и обратные функции распределения, можно построить случайные числа с разнообразными распределениями. Например, формула в электронной таблице:

```
= vWeibull(rnd(1), 1, 2, 0)
```

задает случайные числа, имеющие распределение Вейбулла с параметром масштаба *1*, параметром формы *2* и параметром положения 0. Детали см. в Evans и Peacock (1993, стр. 23-24).

Бета распределение

Бета распределение имеет функцию плотности вида:

$$f(x) = \Gamma(\nu+\omega)/[\Gamma(\nu)\Gamma(\omega)]*[x^{\nu-1} * (1-x)^{\omega-1}]$$

$$0 \le x \le 1,\ \nu > 0,\ \omega > 0$$

где

Γ *гамма* функция

ν, ω параметры *формы*

θ параметр *положения*

Положение (порог) θ и сигма (σ). Стандартное *бета* распределение сосредоточено на отрезке [0,1]. Однако в некоторых случаях (например, в модуле *Анализ процессов* и в графиках *Квантиль-Квантиль*) пользователь может задать меньший параметр положения и параметр масштаба (*сигма*). Программа вычислит стандартизованную *бета* величину:

```
(x-положение)/сигма
```

Заметим, что параметр *положения* (порог) должен быть меньше минимального наблюдаемого значения, а *положение* + *сигма* больше максимального наблюдаемого значения.

Бета распределение для формул таблиц результатов, *STATISTICA BASIC* и определенных пользователем функций

В формулах электронной таблицы, в *STATISTICA BASIC*, в определенных пользователем функциях для бета распределения параметр *положения* θ равен *0,* а параметр *масштаба* σ равен *1*. Поэтому необходимо задать только параметры *формы*. Для плотности *бета* распределения, используйте обозначение *beta(x,ν,ω)*; для функции *бета* распределения, используйте обозначение *ibeta(x,ν,ω)*; для обратного *бета*

Copyright © StatSoft, 1995

распределения - обозначение *vbeta(x,ν,ω)*. Например, следующая формула в электронной таблице

```
= beta(v5,5,9)
```

вычисляет значения плотности вероятности для значений переменной *v5* и записывает их в текущую переменную.

Бета распределение в *Вероятностном калькуляторе*

В *Калькуляторе вероятностных распределений,* кратко, *Вероятностном калькуляторе* параметр *положения* θ равен *0*, а параметр *масштаба* σ равен *1*. Поэтому нужно лишь определить параметры *формы* (ν и ω, соответственно). Заметьте, можно изменить значения этих параметров или редактируя их в соответствующих полях ввода (необходимо нажать кнопку *Вычислить* для выполнения вычислений), или используя "стрелки" для того, чтобы изменить значения полей (в данном способе *STATISTICA* автоматически пересчитает остальные значения). Если один или оба этих параметра изменены, значение *p* будет пересчитано.

Бета распределение в *Квантиль-Квантиль* графиках

В графиках *Квантиль-Квантиль* (*К-К*) (см. Том II), используется стандартизованное *бета* распределение с параметрами *формы* ν и ω для того, чтобы найти распределение наиболее точно соответствующее данным. Параметры *формы* можно задать одним из двух способов в диалоге *Графики Квантиль-Квантиль*:

(1) Ввести определенные значения в поля *Параметр1* и *Параметр2* (ν и ω, соответственно) и отменить опцию *Вычислить параметры*.

(2) Выбрать опцию *Вычислить параметры* ввести параметры θ и σ и оценить параметры формы ν и ω, соответственно.

Дальнейшую информацию о графиках *К-К* см. в Томе II или в *Электронном руководстве*. Если точки на графике *К-К* имеют форму прямой линии, тогда соответствующее семейство распределений (в данном случае *бета* распределения с параметрами формы ν и ω) обеспечивает хорошую подгонку к данным. Свободный член и наклон в данном случае можно рассматривать как оценки параметров положения (порога) θ и масштаба σ, соответственно.

Бета распределение в графиках *Вероятность-Вероятность*

Можно использовать также графики *Вероятность-Вероятность* (*В-В*) для того, чтобы определить насколько хорошо *Бета* распределение соответствует наблюдаемым данным. Для того чтобы создать этот график, теоретическая функция распределения должна быть полностью определена. Поэтому параметры распределения нужно либо задать, либо оценить из данных. Задать параметры распределения можно следующим образом:

Отменить выбор опции *Вычислить по данным*. Когда вы отмените опцию в диалоговом окне, нужно определить два параметра: *формы* ν и ω, и параметры *положения* и *масштаба* θ и σ, соответственно.

Выбрать *Вычислить по данным*. Если выбрана эта опция и определены параметры (θ и σ, соответственно), *STATISTICA* оценит оба параметра ν и ω из данных.

Если точки на графике *В-В* имеют форму прямой линии, тогда соответствующее семейство распределений (в данном случае *бета* распределения с соответствующими

StatSoft®

Copyright © StatSoft, 1995

параметрами ν и ω) обеспечивает хорошую подгонку к данным. Свободный член и наклон в данном случае можно рассматривать как оценки θ и σ параметров, соответственно. Дальнейшую информацию относительно *K-K* и *B-B* графиков см. в Томе II или в *Электронном руководстве*.

Биномиальное распределение

Биномиальные вероятности задаются формулой:

$$f(x) = \{n!/[x!*(n-x)!]\} * p^{x} * q^{n-x}$$

$$0 < p < 1,\ x = 0, 1, \ldots, n$$

где

p вероятность успеха в отдельном испытании

q равна *1-p* (вероятность неуспеха)

n число испытаний

Биномиальное распределение в формулах таблиц результатов, *STATISTICA BASIC* и в определенных пользователем функциях

Для биномиальных вероятностей используйте обозначение *binom(x,p,n)*. Для биномиального распределения используйте обозначение *ibinom(x,p,n)*. Например, следующая формула:

```
= binom(v2,.5,200)
```

вычисляет биномиальные вероятности с вероятностью успеха 0.5, числом испытаний 200 и записывает их в ячейки переменной v2.

Распределение Коши

Распределение Коши имеет плотность вида:

$$f(x) = 1/\theta*\pi*\{1+[(x-\eta)/\theta]^{2}\}$$

$$0 < \theta$$

где

η параметр *положения* (медиана)

θ параметр *масштаба*

π константа *Pi* (3.14...)

Интегральная функция распределения имеет вид:

$$F(x) = 1/2 + 1/\pi*\arctan[(x-\eta)/\theta]$$

Распределение Коши в формулах таблиц результатов, в *STATISTICA BASIC* и определенных пользователем функциях

Для плотности распределения Коши используйте обозначение *Cauchy(x,η,θ)*. Для функции распределения Коши используйте обозначение *iCauchy(x,η,θ)*; для обратного распределения Коши, используйте обозначение *vCauchy(x,η,θ)*. Например, оператор *STATISTICA BASIC*

```
v1 := Cauchy(v2,25,10);
```

вычисляет плотность распределения Коши для значений переменной v2 с параметрами положения и масштаба 25 и 10, соответственно, и присваивает их переменной v1.

Распределение Коши в *Вероятностном калькуляторе*

Вы можете определить параметры *положения* и *масштаба* (η и θ, соответственно) распределения Коши в *Вероятностном калькуляторе* (доступном из опции *Анализ - Основные Быстрые Стат - Вероятностный Калькулятор*). Заметьте, можно изменить значения параметров: или в соответствующих полях ввода (необходимо щелкнуть кнопку *Вычислить* для выполнения вычислений), или используя «стрелки» для того, чтобы инструментально изменить значения полей (в данном способе *STATISTICA* автоматически пересчитает другие значения.) Если один или

Copyright © StatSoft, 1995

StatSoft®

оба этих параметра изменены, значение *p* будет пересчитано.

Распределение *хи-квадрат*

Плотность распределения *хи-квадрат* в точке x равна:

$$f(x)=\{1/[2^{\nu/2}*\Gamma(\nu/2)]\}*[x^{(\nu/2)-1}*e^{-x/2}]$$

$$\nu = 1, 2, ..., \quad 0 < x$$

где

ν степень свободы

e константа Эйлера (2.72...)

Γ *гамма* функция

Распределение *хи-квадрат* в формулах таблиц результатов, в *STATISTICA BASIC* и определенных пользователем функциях

Для обозначения плотности распределения *хи-квадрат* используйте *chi2(x,ν).* Для функции распределения *хи-квадрат* используйте обозначение *ichi2(x,ν)*; для обратной функции распределения *хи-квадрат* используйте обозначение *vchi2(x,ν)*. Например, формула электронной таблицы

```
= chi2(v1,20)
```

вычисляет плотность для значений переменной 1 и записывает их в текущую переменную.

Распределение *хи-квадрат* в Вероятностном калькуляторе

Вы можете определить параметр, ν для распределения *хи-квадрат* в *Вероятностном калькуляторе* (доступном из опции *Анализ - Основные Быстрые Стат - Вероятностный Калькулятор*). Заметьте, можно изменить значения параметров: либо вручную, редактируя их в соответствующих полях ввода (необходимо щелкнуть кнопку *Вычислить* для выполнения вычислений), или используя «стрелки» (в последнем способе *STATISTICA* автоматически пересчитает значение *p*.) При изменении параметров автоматически пересчитывается значение *p*.

Экспоненциальное распределение

Плотность экспоненциального распределения в точке x имеет вид:

$$f(x) = \lambda*e^{[-\lambda(x-\theta)]}$$

$$0 < \lambda, \ \theta < x$$

где

θ параметр *положения*

λ (*лямбда*) параметр *масштаба* (альтернативная параметризация λ = 1/b)

e константа Эйлера (2.72...)

Кумулятивная функция распределения (для θ=0):

$$F(x) = 0 \qquad x \le 0$$
$$= 1 - e^{-\lambda x} \qquad x > 0$$

Обратная функция распределения (вероятности α) равна (для θ=0):

$$-1/\lambda*\log(1-\alpha)$$

Экспоненциальное распределение в формулах таблиц результатов, в *STATISTICA BASIC* и определенных пользователем функциях

В формулах электронной таблицы, *STATISTICA BASIC*, определенных пользователем функциях, параметр θ равен *0*. Поэтому нужно только задать параметр λ распределения. Для плотности экспоненциального распределения используйте обозначение, *expon(x,λ)*; для

StatSoft®

Copyright © StatSoft, 1995

интегральной экспоненциальной функции используйте *iexpon(x,λ);* для обратной функции распределения используйте *vexpon(x,λ)*. Например, следующая определенная пользователем функция (в *Двумерный график функции пользователя*)

```
expon(x,2)
```

построит график плотности экспоненциального распределения с параметром лямбда = 2 в заранее определенном диапазоне значений *x*.

Экспоненциальное распределение в *Вероятностном калькуляторе*

В *Вероятностном калькуляторе* параметр θ равен *0*. Поэтому нужно только определить параметр λ распределения. Заметьте, можно изменить значения параметра: либо вручную, в соответствующем поле ввода (необходимо щелкнуть кнопку *Вычислить* для выполнения вычислений), или используя «стрелки» для того, чтобы инструментально изменить значения (в данном случае *STATISTICA* автоматически пересчитает остальные значения.) Если параметр изменен, значение *p* будет пересчитано на основе измененного значения.

Экспоненциальное распределение в графиках *Квантиль-Квантиль*

В графиках *Квантиль-Квантиль* (*К-К*) (см. Том II, или *Электронное руководство*) используется стандартизованное экспоненциальное распределение, чтобы найти распределение, наилучшим образом аппроксимирующее данные. Необязательно определять все параметры распределения.

Если точки на графике *К-К* имеют форму прямой линии, то соответствующее семейство распределений (экспоненциальное распределение) обеспечивает хорошую подгонку к данным. Свободный член и наклон в данном случае можно рассматривать как графическую интерпретацию оценок параметров θ и λ соответственно.

Экспоненциальное распределение в графиках *Вероятность-Вероятность*

Вы можете использовать графики *Вероятность-Вероятность* (или *В-В*) для того, чтобы определить насколько хорошо экспоненциальное распределение соответствует наблюдаемым данным. Для того чтобы создать график, теоретическая функция распределения должна быть полностью определена. Поэтому параметры распределения должны быть либо заданы, либо оценены. Можно следующим образом задать параметры распределения:

Отменить выбор *Вычислить по данным*. Когда вы отмените эту опцию, нужно ввести параметры λ и θ.

Выбрать *Вычислить по данным*. Когда вы выберете эту опцию и определите параметр θ, *STATISTICA* оценит λ.

Если точки на графике *В-В* имеют форму прямой линии, тогда соответствующее семейство распределений (в данном случае экспоненциальное распределение с соответствующими параметрами) обеспечивает хорошую подгонку к данным. По поводу дальнейшей информации относительно *К-К* и *В-В* графиков, см. Том II или *Электронное руководство*.

Распределение экстремального значения (Гумбеля)

Распределение экстремального значения (Тип I, Гумбеля) имеет функцию плотности вида:

Copyright © StatSoft, 1995

StatSoft®

$$f(x) = 1/b * e^{-(x-a)/b} * e^{-e^{-(x-a)/b}}$$

$$-\infty < x < \infty, \; 0 < b$$

где

a параметр *положения (порог)*

b параметр *масштаба*

e константа Эйлера (2.72...)

Интегральная функция распределения имеет вид:

$$F(x) = e^{-e^{-(x-a)/b}}$$

Обратная функция распределения (вероятности α) имеет вид:

$$a-[b*\text{loglog}(1/\alpha)]$$

Это распределение иногда также называется распределением наибольшего экстремума.

Распределение экстремального значения в формулах таблиц результатов, в *STATISTICA BASIC* и определенных пользователем функциях

Для плотности распределения экстремального значения используйте обозначение *extreme(x,a,b)*; для интегральной функции распределения - обозначение *iextreme(x,a,b)*; для обратной функции распределения используйте обозначение *vextreme(x,a,b)*. Например, следующая формула в электронной таблице

```
= extreme(v1,-1,1)
```

вычислит функцию плотности с параметрами *a=-1*, *b=1* для значений переменной *v1* и занесет их в текущую переменную.

Распределение экстремального значения в *Вероятностном калькуляторе*

Вы можете определить параметры *положения* и масштаба (*a* и *b*, соответственно) для распределения экстремального значения в *Вероятностном Калькуляторе* (доступном из опции *Анализ - Основные Быстрые Стат - Вероятностный Калькулятор*). Заметьте, можно изменить значения параметров: либо задавая их в соответствующих полях ввода (необходимо щелкнуть кнопку *Вычислить* для выполнения вычислений), либо используя «стрелки» (в последнем способе *STATISTICA* автоматически пересчитает другие значения.) Если один или оба этих параметра изменены, значение *p* будет пересчитано.

Распределение экстремального значения в графиках *Квантиль-Квантиль*

В графиках *Квантиль-Квантиль* (*К-К*) (см. Том II, или *Электронное руководство*) используется стандартизованное распределение экстремального значения, чтобы найти распределение, наилучшим образом согласующееся с данными. При этом вовсе не обязательно определять все параметры этого распределения.

Если точки на графике *К-К* имеют форму прямой линии, соответствующее семейство распределений (в данном случае, распределение Гумбеля) обеспечивает хорошую подгонку к данным; свободный член и наклон в данном случае можно рассматривать как графическую интерпретацию оценки параметров *положения (порог)* (*a*) и *масштаба,* соответственно.

Распределение экстремального значения в графиках *Вероятность-Вероятность*

Вы можете использовать графики *Вероятность-Вероятность* (или *В-В*) для того, чтобы определить насколько хорошо распределение Гумбеля соответствует наблюдаемым данным. Для того чтобы создать график, теоретическая функция распределения должна быть полностью определена. Поэтому параметры распределения должны быть либо заданы, либо оценены.

StatSoft®

Copyright © StatSoft, 1995

Можно следующим образом задать параметры распределения:

Отменить выбор *Вычислить по данным*. Когда вы отмените опцию, нужно определить параметры *a* и *b*.

Выбрать *Вычислить по данным*. Когда вы выберете эту опцию, *STATISTICA* оценит *a* и *b* из данных.

Если точки на графике *B-B* имеют форму прямой линии, тогда соответствующее семейство распределений (в данном случае распределения Гумбеля) обеспечивает хорошую подгонку к данным. По поводу дальнейшей информации относительно *K-K* и *B-B* графиков, см. Том II или *Электронное руководство*.

F распределение

F распределение имеет плотность вида:

$$f(x) = \{\Gamma[(\nu+\omega)/2]*(\nu/\omega)^{\nu/2}*x^{(\nu-2)/2}\}/$$
$$\{\Gamma(\nu/2)*\Gamma(\omega/2)[1+(\nu/\omega)*x]^{(\nu+\omega)/2}\}$$

$$0 \le x < \infty$$

$$\nu = 1,2,..., \omega = 1,2,...$$

где

ν, **ω** степени свободы

Γ *гамма* функция

F распределение в формулах таблиц результатов, в *STATISTICA BASIC* и определенных пользователем функциях

Для плотности *F* распределения используйте обозначение *F(x,ν,ω)*; для интегральной функции распределения *F* используйте обозначение *iF(x,ν,ω)*; для обратной функции *F* распределение, используйте обозначение *vF(x,ν,ω)*.

Например, следующая формула в электронной таблице

```
= F(v3,5,10)
```

вычислит функцию *F* плотности со степенями свободы 5 и 10 для значений переменной *v3* и поместит их в текущую переменную электронной таблицы.

В языке *STATISTICA BASIC* для плотности *F* распределения, используется обозначение *FDistr(x,ν,ω)*. Потому что функция *if* в *STATISTICA BASIC* уже есть, для интегральной функции *F* распределения используйте *iFDistr(x,ν,ω)*. Для обратной функции *F* распределения используйте обозначение *vFDistr(x,ν,ω)*.

Например, следующий оператор *STATISTICA BASIC*

```
v10 := FDistr(v3,5,10);
```

присваивает значения плотности вероятности *F* распределения со степенями свободы 5 и 10 в точках *v* 3.

F распределение в *Вероятностном калькуляторе*

Вы можете определить параметры *формы* (*ν*, и *ω*, или *df1* и *df2*, соответственно) для *F* распределения в *Вероятностном Калькуляторе* (доступном из опции *Анализ - Основные Быстрые Стат - Вероятностный Калькулятор*). Заметьте, можно изменить значения параметров: или задавая их в соответствующих полях ввода (необходимо щелкнуть кнопку *Вычислить* для выполнения вычислений), или используя «стрелки» для того, чтобы инструментально изменить значения (в данном способе *STATISTICA* автоматически пересчитает другие значения.) Если один или оба этих параметра изменены, значение *p* будет пересчитано.

Copyright © StatSoft, 1995

StatSoft®

Гамма распределение

Гамма распределение имеет функцию плотности вида:

$$f(x) = \{1/[\Gamma(c)b]\}*[(x-\theta)/b]^{c-1}*e^{[-(x-\theta)/b]}$$

$$0 \leq x$$

где

Γ *гамма* функция

θ параметр *положения*

c параметр *формы*

b параметр *масштаба*

e константа Эйлера (2.72...)

Гамма распределение в формулах таблиц результатов, в *STATISTICA BASIC* и определенных пользователем функциях

В формулах электронной таблицы, в *STATISTICA BASIC* и в определенных пользователем функциях, параметр θ равен *0,* параметр *b* равен *1*. Поэтому нужно задать лишь параметр формы *c*.

Для плотности *гамма* распределения используйте обозначение *gamma(x,c)*; для интегральной функции распределения используйте *igamma(x,c)*; для обратной функции *гамма* распределения используйте *vgamma(x,c)*. Например, следующая формула в электронной таблице

```
= gamma(v10,5)
```

вычисляет плотность вероятности для значений переменной *v10* и помещает в текущую переменную.

Гамма распределение в *Вероятностном калькуляторе*

В *Вероятностном Калькуляторе*, параметр θ равен *0,* параметр *b* равен *1*. Нужно задать лишь параметр *c*. Заметьте, можно изменить значения параметров: либо задавая их в соответствующих полях ввода (необходимо щелкнуть кнопку *Вычислить* для выполнения вычислений), или используя «стрелки» для того, чтобы инструментально изменить значения (в данном способе *STATISTICA* автоматически пересчитает другие значения.) Если один или оба этих параметра изменены, значение *p* будет пересчитано.

Гамма распределение в графиках *Квантиль-Квантиль*

Стандартизованное экспоненциальное распределение используется, чтобы найти распределение, наилучшим образом согласующееся с данными. Параметр *формы* может быть определен одним из следующих способов:

(1) Введите определенное значение для параметра *форма* и отмените опцию *Вычислить параметры*.

(2) Оцените параметр *формы,* выбрав опцию *Вычислить параметры* и введя значение параметра *положения (порога)*.

Если точки на графике *К-К* имеют форму прямой линии, то соответствующее семейство распределений (*гамма* распределение с параметром формы c) обеспечивает хорошую подгонку к данным; свободный член и наклон прямой можно рассматривать как оценки параметров *положения* (θ) и *масштаба* (*b*), соответственно.

StatSoft®

Copyright © StatSoft, 1995

Гамма распределение в графиках *Вероятность-Вероятность*

Вы можете использовать графики *Вероятность-Вероятность* (или *B-B*) для того, чтобы определить насколько хорошо *гамма* распределение соответствует наблюдаемым данным. Для того чтобы создать график, теоретическая функция распределения должна быть определена. Поэтому параметры распределения должны быть либо заданы, либо оценены. Можно следующим образом задать параметры распределения:

Отменить выбор *Вычислить по данным*. Когда вы отмените опцию, нужно определить параметры *формы* и *масштаба* (*c* и *b*, соответственно) и параметр θ..

Выбрать *Вычислить по данным*. Выберете эту опцию и определите параметр θ. *STATISTICA* оценит оставшиеся параметры, параметры *формы* и *масштаба* (*c* и *b*, соответственно), из данных.

Если точки на графике *B-B* имеют форму прямой линии, тогда соответствующее семейство распределений (в данном случае *гамма* распределения с соответствующими параметрами) обеспечивает хорошую подгонку к данным. По поводу дальнейшей информации относительно *K-K* и *B-B* графиков, см. Том II, или *Электронное руководство*.

Геометрическое распределение

Геометрическое распределение имеет функцию плотности вида:

$$f(x) = p*(1-p)^{x-1}$$

$$0 < p < 1,\ x = 1, 2, \ldots$$

где

p вероятность успеха.

Геометрическое распределение в формулах таблиц результатов, в *STATISTICA BASIC* и определенных пользователем функциях

Для плотности геометрического распределения используйте обозначение *geom(x,p)*; для интегральной функции распределения используйте *igeom(x,p)* (обратная функция распределения недоступна здесь). Например, формула в *STATISTICA BASIC*

```
v7 := geom(v1,.5);
```

преобразует значения *v1* в значения плотности геометрического распределения и присвоит их переменной *v7*.

Распределение Лапласа

Плотность распределения Лапласа имеет вид:

$$f(x) = 1/(2b) * e^{-|x-a|/b}$$

$$-\infty < x < \infty,\ 0 < b$$

где

a среднее

b параметр *масштаба*

e константа Эйлера (2.72...)

Интегральная функция распределения Лапласа имеет вид:

$$F(x) = 1/2 * e^{-(a-x)/b} \qquad x < a$$
$$= 1 - \{1/2 * e^{-(x-a)/b} \qquad x \geq a$$

Copyright © StatSoft, 1995

Распределение Лапласа в формулах таблиц результатов, *STATISTICA BASIC* и определенных пользователем функциях

Для плотности распределения Лапласа, используйте обозначение *Laplace(x,a,b)*; для интегральной функции распределения используйте *iLaplace(x,a,b)*; для обратной *vLaplace(x,a,b)*. Например, следующая функция

```
Laplace(x,-2,5)
```

может быть введена в *Двумерный график* для построения графика плотности распределения Лапласа с параметрами *a=-2*, *b=5*.

Распределение Лапласа в *Вероятностном калькуляторе*

Вы можете определить параметры *положения* и *масштаба* (*a* и *b*, соответственно) для распределения Лапласа в *Вероятностном Калькуляторе* (доступном из опции *Анализ - Основные Быстрые Стат - Вероятностный Калькулятор*). Заметьте, можно изменить значения параметров: или задавая их в соответствующих полях ввода (необходимо щелкнуть кнопку *Вычислить* для выполнения вычислений), или используя «стрелки» для того, чтобы инструментально изменить значения полей (в данном способе *STATISTICA* автоматически пересчитает другие значения.) Если один или оба этих параметра изменены, значение *p* будет пересчитано.

Логистическое распределение

Функция плотности логистического распределения имеет вид:

$$f(x) = (1/b)*e^{-(x-a)/b} * [1+e^{-(x-a)/b}]^{-2}$$

$$-\infty < x < \infty,\ 0 < b$$

где

a *среднее* распределения

b параметр *масштаба*

e константа Эйлера (2.72...)

Интегральная функция логистического распределения равна:

$$F(x) = \{1 + e^{[-(x-a)/b]}\}^{-1}$$

Обратная функция логистического распределения имеет вид:

```
a-b * log(1/α - 1)
```

Логистическое распределение в формулах таблиц результатов, в *STATISTICA BASIC* и определенных пользователем функциях

Для плотности логистического распределения, используйте обозначения *logis(x,a,b)*; для интегральной функции распределения используйте *ilogis(x,a,b)*; для обратной функции распределения используйте *vlogis(x,a,b)*. Например, следующая формула электронной таблицы

```
= logis(v0,-4,15)
```

присваивает имеющейся текущей переменной значения плотности для значения *v0* (номер наблюдения).

Логистическое распределение в *Вероятностном калькуляторе*

Вы можете определить параметры *положения* и *масштаба* (*a* и *b*, соответственно) для логистического распределения в *Вероятностном Калькуляторе* (доступном из опции *Анализ - Основные Быстрые Стат - Вероятностный Калькулятор*). Заметьте, можно изменить значения параметров: либо задавая их в соответствующих полях ввода (необходимо щелкнуть кнопку *Вычислить* для выполнения вычислений), или используя «стрелки» для того,

StatSoft

Copyright © StatSoft, 1995

чтобы инструментально изменить значения полей (в данном способе *STATISTICA* автоматически пересчитает другие значения.) Если один или оба этих параметра изменены, значение *p* будет пересчитано.

Логнормальное распределение

Функция плотности логнормального распределения имеет вид:

$$f(x)=1/[(x-\theta)\sigma(2\pi)^{1/2}]*e^{-[\log(x-\theta)-\mu]^2/2\sigma^2}$$

$$\theta < x < \infty,\ 0 < \mu,\ 0 < \sigma$$

где

μ параметр *масштаба*

σ параметр *форма*

θ параметр *положения*

e константа Эйлера (2.72...)

Логнормальное распределение в формулах таблиц результатов, в *STATISTICA BASIC* и в определенных пользователем функциях

В формулах электронной таблицы, *STATISTICA BASIC*, и определенных пользователем функциях параметр положения θ равен *0*. Поэтому нужно только определить параметры *масштаба* и *формы* (μ и σ, соответственно).

Для плотности логнормального распределения используйте обозначение *lognorm(x,μ,σ)*; для интегральной функции распределения используйте *ilognorm(x,μ,σ)*; для обратной функции распределения используйте *vlognorm(x,μ,σ)*.

Например, формула

```
= lognorm(v12,0,1)
```

вычисляет плотность вероятности для значений в переменной *v12* и записывает в текущую переменную.

Логнормальное распределение в *Вероятностном калькуляторе*

В *Вероятностном Калькуляторе* параметр *положения* θ равен *0*. Поэтому нужно только определить параметры *масштаба* и *формы* (μ и σ, соответственно). Заметьте, можно изменить значения параметров: или задавая их в соответствующих полях ввода (необходимо щелкнуть кнопку *Вычислить* для выполнения вычислений), или используя «стрелки» для того, чтобы инструментально изменить значения полей (в данном способе *STATISTICA* автоматически пересчитает другие значения.) Если один или оба этих параметра изменены, значение *p* будет пересчитано.

Логнормальное распределение в графиках *Квантиль-Квантиль*

Стандартизованное логнормальное распределение с параметром σ используется, чтобы найти распределение, наилучшим образом согласующееся с данными. Параметр σ можно определить одним из следующих способов:

(1) Отмените опцию *Вычислить параметры* и введите значение параметра σ

(2) Выберите опцию *Вычислить параметры,* введите значение параметра θ и оцените σ.

Если точки на графике *К-К* имеют форму прямой линии, то соответствующее семейство распределений (логнормальное распределение с параметром σ) обеспечивает хорошую подгонку к данным; свободный член и наклон прямой можно рассматривать как оценки параметров *положения* (θ) и *масштаба* (μ).

Copyright © StatSoft, 1995

StatSoft®

Логнормальное распределение в графиках *Вероятность-Вероятность*

Вы можете использовать графики *Вероятность-Вероятность* (или *B-B*) для того, чтобы определить, насколько хорошо логнормальное распределение подходит к наблюдаемым данным. Для того чтобы создать график, теоретическая функция распределения должна быть определена. Поэтому параметры распределения должны быть либо заданы, либо оценены из данных. Можно следующим образом задать параметры распределения:

Отменить выбор *Вычислить по данным*. Когда вы отмените опцию, нужно ввести параметры *формы* и *масштаба* (σ и μ, соответственно) и параметр θ.

Выбрать *Вычислить по данным*. Когда вы выберете эту опцию и определите параметр θ, *STATISTICA* оценит параметры *формы* и *масштаба* (σ и μ, соответственно).

Если точки на графике *B-B* имеют форму прямой линии, тогда соответствующее семейство распределений (в данном случае логнормальное распределение с соответствующими параметрами) обеспечивает хорошую подгонку к данным. По поводу дальнейшей информации относительно *K-K* и *B-B* графиков см. Том II или *Электронное руководство*.

Нормальное распределение

Функция плотности нормального распределения имеет вид:

$$f(x) = 1/[\sigma(2\pi)^{1/2}] * e^{\{-1/2[(x-\mu)/\sigma]^2\}}$$

$$-\infty < x < \infty,\ 0 < \sigma$$

где

μ мю - среднее

σ сигма - стандартное отклонение

e константа Эйлера (2.72...)

π константа *Pi* (3.14...)

Нормальное распределение в формулах таблиц результатов, в *STATISTICA BASIC* и определенных пользователем функциях

Для плотности нормального распределения используйте обозначение *Normal(x,μ,σ)*; для интегральной функции распределения используйте обозначение *iNormal(x,μ,σ)*; для обратной функции распределения используйте обозначение *vNormal(x,μ,σ)*.

Например, формула

```
= Normal(v2,0,1)
```

вычисляет плотность нормального распределения для значений переменной *v2* и записывает в текущую переменную.

Так как обозначение *normal* уже использовано в *STATISTICA BASIC*, то для плотности нормального распределения используйте *dNormal(x,μ,σ)*, для интегральной функции распределения используйте *iNormal(x,μ,σ)*; для обратной функции распределения используйте *vNormal(x,μ,σ)*.

Например, оператор в *STATISTICA BASIC*

```
v25 := dNormal(v15,10,5);
```

вычислит значения плотности нормального распределения (с параметрами 10 и 5) в точках *v* 15 и запишет их в переменную *v* 25.

Нормальное распределение в *Вероятностном калькуляторе*

Вы можете определить среднее и стандартное отклонение параметров (μ и σ, соответственно) для нормального распределения в *Вероятностном Калькуляторе*, (доступном из

StatSoft®

Copyright © StatSoft, 1995

опции *Анализ - Основные Быстрые Стат - Вероятностный калькулятор*). Заметьте, можно изменить значения параметров: либо задавая их в соответствующих полях ввода (необходимо щелкнуть кнопку *Вычислить* для выполнения вычислений), или используя «стрелки» (в данном способе *STATISTICA* автоматически пересчитает другие значения.) Если один или оба этих параметра изменены, значение *p* будет пересчитано на основе изменений значений.

Нормальное распределение в графиках *Квантиль-Квантиль*

Стандартизованное нормальное распределение используется, чтобы найти распределение, наилучшим образом согласующееся с данными. В данном случае нет необходимости определять какие-либо параметры.

Если точки на графике *К-К* имеют форму прямой линии, то соответствующее семейство распределений (нормальное распределение) обеспечивает хорошую подгонку к данным; свободный член и наклон можно рассматривать как оценки среднего (μ) и стандартного отклонения (σ).

Нормальное распределение в графиках *Вероятность-Вероятность*

Вы можете использовать графики *Вероятность-Вероятность* (или *В-В*) для того, чтобы определить, насколько хорошо нормальное распределение подходит к наблюдаемым данным. Для того чтобы создать график, теоретическая функция распределения должна быть определена. Поэтому параметры распределения должны быть либо заданы, либо оценены. Можно следующим образом задать параметры распределения:

Отменить выбор *Вычислить по данным*. Когда вы отмените опцию, нужно ввести параметры *Мю* (μ) и *Сигма* (σ).

Выбрать *Вычислить по данным*. Когда вы выберете опцию *Вычислить по данным* и определите параметр θ, *STATISTICA* оценит параметры *Мю* (μ) и *Сигма* (σ) из данных.

Если точки на графике *В-В* имеют форму прямой линии, то соответствующее семейство распределений (в данном случае нормальное распределение с соответствующими параметрами) обеспечивает хорошую подгонку к данным. По поводу дальнейшей информации относительно *К-К* и *В-В* графиков, см. Том II, или *Электронное руководство*.

Распределение Парето

Функция плотности распределения Парето имеет вид:

$$f(x) = c/x^{c+1}$$

$$1 \le x,\ c < 0$$

где

c параметр *формы* распределения.

Интегральная функция распределения Парето имеет вид:

$$F(x) = 1 - x^{-c}$$

Обратная функция распределения Парето (от вероятности α) имеет вид:

$$(1-\alpha)^{c}$$

Распределение Парето в формулах таблиц результатов, в *STATISTICA BASIC* и определенных пользователем функциях

Для плотности распределения Парето используйте обозначение *Pareto(x,c)*; для

Copyright © StatSoft, 1995

интегральной функции распределения используйте *iPareto(x,c)*; для обратной функции распределения используйте *vPareto(x,c)*. Например, формула в электронной таблице:

```
= Pareto(v1,1)
```

вычисляет функцию плотности для значений переменной v1 записывает их в текущую переменную.

Распределение Парето в *Вероятностном калькуляторе*

Вы можете определить параметр *c* распределения Парето в *Вероятностном калькуляторе* (доступном из опции *Анализ - Основные Быстрые Стат - Вероятностный калькулятор*). Заметьте, можно изменить значение параметра: либо вручную редактируя его в соответствующем поле ввода (необходимо щелкнуть кнопку *Вычислить* для выполнения вычислений), или используя «стрелки» для того, чтобы инструментально изменить значение поле (в данном способе *STATISTICA* автоматически пересчитает другие значения.) Если параметр изменен, значение *p* будет пересчитано.

Распределение Пуассона

Функция плотности распределения Пуассона имеет вид:

$$f(x) = (\lambda^x * e^{-\lambda})/x!$$

$$0 < \lambda,\ x = 0, 1, \ldots$$

где

λ (*лямбда)* среднее

e константа Эйлера (2.72...)

Распределение Пуассона в формулах таблиц результатов, в *STATISTICA BASIC* и определенных пользователем функциях

Для плотности распределения Пуассона используйте обозначение *Poisson(x,λ)*; для интегральной функции распределения используйте *iPoisson(x,λ)*. Обратная функция недоступна для этого распределения.

Например, функция

```
Poisson(x,25)
```

может быть введена как определенная пользователем функция в диалог *Двумерный график функции пользователя* для того, чтобы построить график плотности распределения Пуассона для значений *x* из определенного диапазона.

Так как термин *poisson* уже используется в *STATISTICA BASIC*, то для обозначения плотности распределения Пуассона применяется *dpoisson(x,λ),* для интегральной функции распределения применяется *ipoisson(x,λ)*.

Например, следующий *STATISTICA BASIC* оператор

```
v4 := dPoisson(v1,2);
```

присваивает переменной v4 значения плотности распределения Пуассона с параметром 2 в точках v1.

Распределение Релея

Функция плотности распределения Релея имеет вид:

$$f(x) = (x-\theta)/b^2 * e^{-[(x-\theta)^2/(2b^2)]}$$

$$\theta \le x < \infty,\ b > 0$$

StatSoft®

Copyright © StatSoft, 1995

где

b параметр *масштаба*

θ параметр *положения*

e константа Эйлера (2.72...)

Интегральная функция распределения имеет вид:

$$F(x) = 1 - e^{-(x^2/(2b^2))}$$

Обратная функция распределения (вероятностный уровень α) имеет вид (для θ=0):

$$\{-2b^2[\log(1-\alpha)]\}^{1/2}$$

Распределение Релея в формулах таблиц результатов, в *STATISTICA BASIC* и определенных пользователем функциях

В формулах электронной таблицы, в *STATISTICA BASIC* и определенных пользователем функциях параметр θ равен *0*. Поэтому вам нужно определить лишь параметр *масштаба b*. Для плотности Релея используйте обозначение *Rayleigh(x,b)*; для интегральной функции распределения используйте *iRayleigh(x,b)*; для обратной функции распределения используйте *vRayleigh(x,b)*.

Например, формула в электронной таблице

```
= Rayleigh(v2,4)
```

вычисляет плотность вероятности для значений переменной *v2* и записывает в текущую переменную.

Распределение Релея в *Вероятностном калькуляторе*

В *Вероятностном калькуляторе* параметр θ равен *0*. Поэтому нужно определить только параметр *масштаба* (*b*). Заметьте, можно изменить значения этого параметра: либо вводя его в поле ввода (необходимо щелкнуть кнопку *Вычислить* для выполнения вычислений), или используя «стрелки» для того, чтобы инструментально изменить значение поля (в данной случае *STATISTICA* автоматически пересчитает другие значения распределения в соответствии с изменением параметра.) Если этот параметр изменен, значение *p* будет пересчитано.

Распределение Релея в графиках *Квантиль-Квантиль*

Стандартизованное распределение Релея используется, чтобы найти распределение, наилучшим образом согласующееся с данными. Нет необходимости здесь определять какие-либо параметры распределения.

Если точки на графике *К-К* имеют форму прямой линии, тогда соответствующее семейство распределений (в данном случае, распределение Релея) обеспечивает хорошую подгонку к данным; свободный член и наклон можно рассматривать как графические оценки параметров θ и *b*, соответственно.

Распределение Релея в графиках *Вероятность-Вероятность*

Вы можете использовать графики *Вероятность-Вероятность* (или *В-В*) для того, чтобы определить, насколько хорошо распределение Релея подходит к данным. Для того чтобы создать график, теоретическая функция распределения должна быть определена. Поэтому параметры распределения должны быть либо заданы, либо оценены из данных. Можно следующим образом задать параметры распределения:

Отменить выбор *Вычислить по данным*. Когда вы отмените эту опцию, нужно задать как параметр *масштаба,* так и параметр *положения* θ.

Copyright © StatSoft, 1995

Выбрать *Вычислить по данным*. Когда вы выберете эту опцию и зададите порог θ, *STATISTICA* оценит параметр *масштаба b* из данных.

Если точки на графике *B-B* имеют форму прямой линии, тогда соответствующее семейство распределений (в данном случае распределение Релея с соответствующими параметрами) обеспечивает хорошую подгонку к данным. По поводу дальнейшей информации относительно *K-K* и *B-B* графиков, см. Том II, или *Электронное руководство*.

t распределение Стьюдента

Функция плотности t распределения Стьюдента имеет вид:

$$f(x) = \Gamma[(\nu+1)/2] / \Gamma(\nu/2) * (\nu*\pi)^{-1/2} * [1+(x^2/\nu)]^{-(\nu+1)/2}$$

$$\nu = 1,2,...$$

где

ν число степеней свободы

Γ *гамма* функция

π константа *Pi* (3.14...)

Интегральная функция распределения Стьюдента в зависимости от четности *ν* всесторонне описана в книге Evans, Hastings, и Peacock, 1993.

t распределение Стьюдента в формулах таблиц результатов, в *STATISTICA BASIC* и определенных пользователем функциях

Для плотности распределения Стьюдента, используйте обозначение *student(x,df)*; для интегральной функции распределения используйте *istudent(x,df)*; для обратной функции распределения используйте *vstudent(x,df)*. Например, следующая формула

```
= student(v4,25)
```

вычисляет плотность распределения Стьюдента с числом степеней свободы 25 для значений переменной *v4* и помещает в ячейки текущей переменной.

t распределение Стьюдента в *Вероятностном калькуляторе*

Вы можете определить число степеней свободы распределения Стьюдента в *Вероятностном калькуляторе* (доступном из опции *Анализ - Основные Быстрые Стат - Вероятностный калькулятор*). Заметьте, можно изменить значение этого параметра: либо задавая его в соответствующем поле ввода (необходимо щелкнуть кнопку *Вычислить* для выполнения вычислений), или используя «стрелки» для того, чтобы инструментально изменить значения (в данном способе *STATISTICA* автоматически пересчитает другие значения.) Если параметр изменен, значение *p* будет пересчитано.

Распределение Вейбулла

Функция плотности распределения Вейбулла имеет вид:

$$f(x) = c/b*[(x-\theta)/b]^{c-1} * e^{-[(x-\theta)/b]^c}$$

$$x \geq \theta,\ 0 < b,\ 0 < c\theta$$

где

b параметр *масштаба*

c параметр *форма*

θ параметр *положения*

e константа Эйлера (2.72...)

Обратная функция распределения имеет вид:

$$b\{\log[1/(1-\alpha)]\}^{1/c}$$

StatSoft®

Copyright © StatSoft, 1995

Распределение Вейбулла в формулах таблиц результатов, в *STATISTICA BASIC* и определенных пользователем функциях

Для функции плотности распределения Вейбулла используйте обозначение *Weibull(x,b,c,θ)*; для интегральной функции распределения используйте *iWeibull(x,b,c,θ)*; для обратной функции распределения используйте *vWeibull(x,b,c,θ)*. Например, функция

```
Weibull(x,1,3,1)
```

может быть введена как определенная пользователем функция в диалог *Двумерный график* для того, чтобы построить график плотности Вейбулла с параметрами b=1,c=3, θ=1.

Распределение Вейбулла в *Вероятностном калькуляторе*

В *Вероятностном калькуляторе* параметр θ равен *0*. Поэтому нужно лишь задать параметры *масштаба* и *формы* (*b* и *c*, соответственно). Заметьте, можно изменить эти значения: либо вводя их в соответствующих полях ввода (необходимо щелкнуть кнопку *Вычислить* для выполнения вычислений), или используя «стрелки» для того, чтобы инструментально изменить значения полей (в данном способе *STATISTICA* автоматически пересчитает другие значения.) Если один или оба этих параметра изменены, значение *p* будет пересчитано.

Распределение Вейбулла в графиках *Квантиль-Квантиль*

Стандартизованное распределение Вейбулла с параметром *c* используется, чтобы найти распределение, наилучшим образом согласующееся с данными. Параметр *c* может быть определен одним из следующих способов:

(1) Отмените выбор опции *Вычислить параметры* и введите значение параметра *c*.

(2) Оцените параметр *c* из данных, выбрав опцию *Вычислить параметры*.

Если точки на графике *K-K* имеют форму прямой линии, тогда соответствующее семейство распределений (в данном случае семейство Вейбулла с параметром *формы c*) обеспечивает хорошую подгонку к данным; свободный член и наклон прямой можно рассматривать оценки параметров *положения* (θ) и *масштаба* (*b*).

Распределение Вейбулла в графиках *Вероятность-Вероятность*

Вы можете использовать графики *Вероятность-Вероятность* (или *B-B*) для того, чтобы определить насколько хорошо распределение Вейбулла подходит к наблюдаемым данным. Для того чтобы создать график, теоретическая функция распределения должна быть определена. Поэтому параметры распределения должны быть либо заданы, либо оценены. Можно следующим образом задать параметры распределения:

Отменить выбор *Вычислить по данным*. Когда вы отмените опцию, нужно задать параметры *формы* и *масштаба* (*b* и *c*, соответственно) а также параметр θ.

Выбрать *Вычислить по данным*. Когда вы выберете эту опцию и определите параметр θ, *STATISTICA* оценит параметры *формы* и *масштаба* (*b* и *c*, соответственно) из данных.

Если точки на графике *B-B* имеют форму прямой линии, тогда соответствующее семейство распределений (в данном случае распределение Вейбулла с соответствующими параметрами) обеспечивает хорошую подгонку к данным. По поводу дальнейшей информации относительно *K-K* и *B-B* графиков, см. Том II, или *Электронное руководство*.

Copyright © StatSoft, 1995

ПРИЛОЖЕНИЕ II: ИТОГОВЫЕ СТАТИСТИКИ

Следующие итоговые статистики доступны, когда вы выбрали опцию *Основные Быстрые Стат - Описательные* (см. Глава 10).

Число наблюдений *n*

Это число всех (непропущенных) значений соответствующей переменной. Пропущенные данные обрабатываются в *STATISTICA* одним из следующих способов:

Построчное удаление пропущенных данных

Если выбрано построчное удаление пропущенных данных (например, *Построчное удаление ПД* в диалоговом окне *Основные Быстрые Стат - Расширенные опции*), то в анализ будут включены только случаи, которые не содержат пропущенных данных во всех выбранных переменных. При вычислении коэффициентов корреляций исключаются случаи, в которых есть пропуски хотя бы в одной выбранной переменной (таким образом, все корреляции оцениваются по одному и тому же множеству данных).

Попарное удаление пропущенных данных

Если выбрано попарное удаление пропущенных данных (например, если не поставлена галочка, отмечающая *Построчное удаление ПД* в диалоговом окне *Основные Быстрые Стат – Расширенные опции*), то все наблюдаемые значения для пары переменных будут включены в анализ соответствующих переменных. Возможно, при этом будет иметься неравное *число наблюдений n* для каждой пары переменных.

Подстановка среднего вместо пропусков

Если выбран способ *Подстановка среднего*, пропущенные данные будут заменены средним значением. (Эта опция доступна только в ряде модулей; заметьте, что можно также заменить пропущенные данные средними с помощью опции *Заменить пропущенные данные* из меню *Анализ* в модуле *Управление данными*.)

Сумма

Значение суммы для данных без пропусков равно:

$$\text{сумма} = \Sigma x_i \quad \text{для } i = 1 - n$$

Минимум

Минимальное значение соответствующей переменной.

Максимум

Максимальное значение соответствующей переменной.

Среднее

Среднее показывает “центральное положение” или центр переменной и рассматривается вместе с доверительным интервалом (см. ниже). Обычно мы интересуемся статистиками (например, среднее), построенными по выборке, для того, чтобы получить информацию о популяции в целом. Чем больше размер выборки, тем более надежна оценка среднего. Чем больше изменчивость данных, тем менее

StatSoft®

Copyright © StatSoft, 1995

надежна оценка среднего (см. *Элементарные понятия*, Глава 8).

$$\text{Среднее} = (\Sigma x_i)/n$$

где

n размер выборки.

Доверительные границы для среднего

Доверительный интервал для среднего дает диапазон значений вокруг оценки среднего, где, как мы ожидаем, находится "истинное" среднее популяции (с данным уровнем доверия, см. *Элементарные понятия*, Глава 8).

Например, если среднее в вашем примере равно 23, и нижняя и верхняя границы доверительного интервала с p=.95 равны 19 и 27 соответственно, то вы можете заключить, что с вероятностью 95% интервал с границами 19 и 27 накрывает среднее популяции. Если вы возьмете больший p-уровень, то интервал станет шире, и в связи с этим возрастет вероятность, с какой он накрывает неизвестное среднее популяции, и наоборот; например, хорошо известно, что чем "неопределенней" прогноз погоды (т.е., шире доверительный интервал), тем вероятнее он оправдается.

Заметим, что ширина доверительного интервала зависит от размера выборки и разброса данных. Доверительные интервалы вычисляются в предположении, что переменная в популяции распределена нормально. Эти оценки могут быть неточными при небольшом размере выборки, скажем меньше, чем n = 100.

Размах

Размах равен: максимальное значение переменной минус минимальное значение

Процентили

Процентиль - это значение, слева от которого находится заданная доля (процент) наблюдений количественной переменной. Можно выбрать шесть способов вычисления процентилей или процентных точек, включая медиану (50% процентиль) и квартили (25% и 75% процентили). Выбранный метод применим в любом модуле *STATISTICA,* где вычисляются медиана, квартили или процентили например:

- процентили в модуле *Непараметрическая статистика и распределения*;
- медиана и квартили в модуле *Основные статистики и таблицы*;
- медиана и квартили в *диаграммах размаха*;
- медиана и квартили в графиках ящики с усами;
- медиана и квартили в графиках ящики с усами в *Дискриминантном анализе*, *Надежности и позиционном анализе*, *Основных статистиках, Множественной регрессии*.

В описанных далее методах n число наблюдений и p процентиль, деленный на 100 (например, *50/100 = .5* для медианы).

Метод взвешенного среднего в X_{np}

Метод Это взвешенное с центром в точке X_{np}.среднее. Представим np (n, умноженное на p) как $np=j+g$, где j целая часть np, и g дробная часть np; затем вычислить:

$$\text{ЗначениеПроцентиля} = (1-g)x_j + gx_{j+1}$$

Заметим, X_0 в этом вычислении заменяется X_1 (например, если $j = 0$, то формула имеет вид $(1-g)x_1 + gx_{(0+1)}$).

Copyright © StatSoft, 1995

StatSoft®

Пример. Рассмотрим следующие упорядоченные данные (вариационный ряд):

1, 2, 4, 7, 8, 9, 10, 12, 13

Здесь $n = 9$ и пусть $p = .25$ (25% процентиль). Представим np (n, умноженное на p) как:

np = 9*.25 = 2.25 = j + g

поэтому $j = 2$ и $g = .25$. Теперь 25% процентиль вычисляется по формуле:

$$\text{25\% Процентиль} = (1-.25)x_2 + .25x_3$$
$$= (.75)2 + (.25)4$$
$$= 2.5$$

Метод взвешенного среднего в $X_{(n+1)p}$

Это взвешенное среднее с центром в $X_{(n+1)p}$. Выразим $(n+1)p$ как $(n+1)p=j+g$ где j целая часть $(n+1)p$, и g дробная часть $(n+1)p$; тогда:

$$\text{ЗначениеПроцентиля} = (1-g)x_j + gx_{j+1}$$

Заметим, что X_{n+1} в этих вычислениях заменяется X_n (например, если $j = n$, тогда предыдущая формула представляется в виде $(1-g)x_n + gx_n$).

Пример. Рассмотрим следующие упорядоченные данные (вариационный ряд):

1, 2, 4, 7, 8, 9, 10, 12, 13

Здесь, $n = 9$ и пусть $p = .25$ (25% процентиль). Представим $(n+1)p$ как:

(n+1)p = 10*.25 = 2.5 = j + g

поэтому $j = 2$ и $g = .5$. Теперь вычислите 25% процентиль:

$$\text{25\% Процентиль} = (1-.5)x_2 + .5x_3$$
$$= (.5)2 + (.5)4$$
$$= 3.0$$

Метод эмпирической функции распределения

Представим np (n, умноженное на p,где n число наблюдений, p-выбранный уровень) как $np=j+g$ где j есть целая часть np, и g дробная часть np; тогда выберем значение процентной точки как:

$$\text{ЗначениеПроцентиля} = x_j \quad \text{если } g=0$$
$$\text{ЗначениеПроцентиля} = x_{j+1} \quad \text{если } g>0$$

Пример. Рассмотрим следующие упорядоченные данные (вариационный ряд):

1, 2, 4, 7, 8, 9, 10, 12, 13

Здесь $n = 9$ и пусть $p = .25$ (25% процентиль). Представим np (n, умноженное на p) как:

np = 9*.25 = 2.25 = j + g

поэтому, $j = 2$ и $g = .25$. Так как. $g>0$, вычислим 25% процентиль как:

$$\text{25\% Процентиль} = x_3$$
$$= 4.0$$

Эмпирическая функция распределения с методом усреднения

Выразим np (n, умноженное на p) как $np=j+g$, где j целая часть np, и g дробная часть np; тогда значение процентиля вычисляется по формуле:

$$\text{ЗначениеПроцентиля} = (x_j + x_{j+1})/2 \quad \text{если } g=0$$
$$\text{ЗначениеПроцентиля} = x_{j+1} \quad \text{если } g>0$$

Пример. Рассмотрим вариационный ряд:

1, 2, 4, 7, 8, 9, 10, 12, 13

Здесь $n = 9$ и пусть $p = .25$ (25% процентиль). Представим np (n, умноженное на p) как:

np = 9*.25 = 2.25 = j + g

поэтому $j = 2$ и $g = .25$. Так как $g>0$, вычислим 25% процентиль по формуле:

$$\text{25\% Процентиль} = x_3$$
$$= 4.0$$

StatSoft®

Copyright © StatSoft, 1995

Метод эмпирической функции распределения и интерполяции (MS Excel)

Представим *(n-1)p* (*(n-1),* умноженное на *p*) как *(n-1)p=j+g*, где *j* целая часть *(n-1)p*, и дробная часть *(n-1)p*; тогда значение процентиля вычисляется по формуле:

$\text{ЗначениеПроцентиля}=x_{j+1}$ если g=0

$\text{ЗначеиеПроцентиля}=x_{j+1}+g(x_{j+2}-x_{j+1})$ если g>0

Пример. Рассмотрим вариационный ряд:

1, 2, 4, 7, 8, 9, 10, 12, 13

Здесь *n* = 9 и пусть *p* = .25 (т.е. вычисляет 25% процентиль). Представим *(n-1)p* как:

(n-1)p = 8*.25 = 2.0 = j + g

отсюда, *j = 2* и *g = 0*. Так как *g*=0, вычислим 25% процентиль как:

$25\%\ \text{Процентиль} = x_3$

$= 4.0$

Метод ближайшего наблюдения

Упорядочим наблюдения по возрастанию (построим вариационный ряд). Вычислим *j* как целую часть *np+1/2*, затем определим:

$\text{ЗначениеПроцентиля} = x_j$

Где x_j есть i-й член вариационного ряда.

Пример. Рассмотрим следующий вариационный ряд:

1, 2, 4, 7, 8, 9, 10, 12, 13

Здесь *n* = 9 и пусть *p* = .25 (25% процентиль). Представим *np* в виде:

np + 1/2 = (9*.25) + .5 = 2.75 = j + g

отсюда *j = 2*. Теперь вычислим 25% процентиль по формуле:

$25\%\ \text{Процентиль} = x_2$

$= 2.0$

Дисперсия

Выборочная оценка дисперсии вычисляется по формуле:

$s^2 = \Sigma(x_i - \text{xbar})^2 / (n-1)$

где

xbar выборочное среднее

n размер выборки.

Стандартная ошибка среднего

Стандартная ошибка средняя – это стандартное отклонение среднего, вычисляется по формуле:

$\text{std.err.} = \sqrt{(s^2/n)}$

где

s^2 стандартное отклонение

n размер выборки.

Эксцесс

Эксцесс (или точнее, коэффициент эксцесса) измеряет “пикообразность” распределения. Если эксцесс значимо отличен от 0, то функция плотности либо более пологая, чем функция плотности нормального распределения либо имеет более острый пик, чем пик плотности нормального распределения. Функция плотности нормального распределения имеет эксцесс равный 0. Оценка эксцесса (выборочный эксцесс) вычисляется по формуле:

$$\text{Эксцесс} = [n*(n+1)*M_4 - 3*M_2*M_2*(n-1)] / [(n-1)*(n-2)*(n-3)*\sigma^4]$$

где

Copyright © StatSoft, 1995

M_j равен: $\sum(x_i\text{-Среднее})^j$

σ^4 стандартное отклонение (*сигма*), возведенное в четвертую степень

n число наблюдений.

Стандартная ошибка эксцесса

Стандартная ошибка эксцесса вычисляется по формуле:

$$\text{Std.Err.} = \{[4*(n^2-1)*SE(S)^2]/[(n-3)*(n+5)]\}^{1/2}$$

где

SE(S) равен стандартной ошибке асимметрии (см. ниже)

n число наблюдений.

Асимметрия

Асимметрия является мерой несимметричности распределения. Если асимметрия отчетливо отличается от 0, распределение асимметричное, плотность нормального распределения – симметрична относительно среднего.

$$\text{Ассиметрия} = n*M_3/(n-1)*(n-2)*\sigma^3$$

где

M^j равен: $\sum(x_i\text{-Среднее})^j$

σ^3 стандартное отклонение (*сигма*) в степени 4

n число наблюдений.

Стандартная ошибка асимметрии

Стандартная ошибка асимметрии вычисляется по формуле:

$$\text{Std.Err.} = \{[6n*(n-1)]/[(n-2)*(n+1)*(n+3)]\}^{1/2}$$

где

n число наблюдений.

StatSoft®

Copyright © StatSoft, 1995

ПРЕДМЕТНЫЙ УКАЗАТЕЛЬ

F

F распределение, 1837
  вероятностный калькулятор, 1837

А

Асимметрия, 1852

Б

Бета распределение, 1831
  вероятностный калькулятор, 1832
  Вероятность-Вероятность графики, 1832
  Квантиль-Квантиль графики, 1832
Биномиальное распределение, 1833

В

Вейбулла распределение, 1846
  вероятностный калькулятор, 1847
  Вероятность-Вероятность графики, 1847
  Квантиль-Квантиль графики, 1847
Вероятностный калькулятор
  F распределение, 1837
  t распределение Стьюдента, 1846
  бета распределение, 1832
  гамма распределение, 1838
  логистическое распределение, 1840
  логнормальное распределение, 1841
  нормальное распределение, 1842
  распределение Вейбулла, 1847
  распределение Коши, 1833
  распределение Лапласа, 1840
  распределение Парето, 1844
  распределение Релея, 1845
  распределение хи-квадрат, 1834
  распределение экстремального значения, 1836
  экспоненциальное распределение, 1835
Вероятность-Вероятность графики
  бета распределение, 1832
  гамма распределение, 1839
  логнормальное распределение, 1842
  нормальное распределение, 1843
  распределение Вейбулла, 1847
  распределение Релея, 1845
  распределение экстремального значения (Гумбеля), 1836
  экспоненциальное распределение, 1835

Г

Гамма распределение, 1838
  вероятностный калькулятор, 1838
  Вероятность-Вероятность графики, 1839
  Квантиль-Квантиль графики, 1838
Геометрическое распределение, 1839
Гумбеля распределение, 1835

Д

Дисперсия, 1851
Доверительные границы для среднего, 1849

И

Итоговые статистики, 1848
  асимметрия, 1852
  дисперсия, 1851
  доверительные границы для среднего, 1849
  квантили, 1849
  квартили, 1849
  максимум, 1848
  минимум, 1848
  процентили, 1849
  размах, 1849
  среднее, 1848
  стандартная ошибка асимметрии, 1852
  стандартная ошибка среднего, 1851
  стандартная ошибка эксцесса, 1852
  сумма, 1848
  число наблюдений n, 1848
  эксцесс, 1851

К

Квантили, 1849
Квантиль-Квантиль графики
  бета распределение, 1832
  гамма распределение, 1838
  логнормальное распределение, 1841
  нормальное распределение, 1843
  распределение Вейбулла, 1847
  распределение Релея, 1845
  распределение экстремального значения (Гумбеля), 1836
  экспоненциальное распределение, 1835
Квартили, 1849
Коши распределение, 1833
  вероятностный калькулятор, 1833

Л

Лапласа распределение, 1839
  вероятностный калькулятор, 1840
Логистическое распределение, 1840
  вероятностный калькулятор, 1840
Логнормальное распределение, 1841
  вероятностный калькулятор, 1841
  Вероятность-Вероятность графики, 1842
  Квантиль-Квантиль графики, 1841

М

Максимальное значение, 1848
Метод взвешенного в X_{np} среднего вычисления процентилей, 1849
Минимальное значение, 1848

Н

Нормальное распределение, 1842

Copyright © StatSoft, 1995

StatSoft®

вероятностный калькулятор, 1842
Вероятность-Вероятность графики, 1843
Квантиль-Квантиль графики, 1843

О

Общие установки по умолчанию
процентили, 1849
Определенные пользователем функции
F распределение, 1837
t распределение Стьюдента, 1846
бета распределение, 1831
биномиальное распределение, 1833
гамма распределение, 1838
геометрическое распределение, 1839
логистическое распределение, 1840
логнормальное распределение, 1841
нормальное распределение, 1842
распределение Вейбулла, 1847
распределение Коши, 1833
распределение Лапласа, 1840
распределение Парето, 1843
распределение Пуассона, 1844
распределение Релея, 1845
распределение экстремального значения, 1836
хи-квадрат распределение, 1834
экспоненциальное распределение, 1834

П

Парето распределение, 1843
вероятностный калькулятор, 1844
Подстановка среднего вместо пропущенных данных, 1848
Попарное удаление пропущенных данных, 1848
Построчное удаление пропущенных данных, 1848
Процентили, 1849
Пуассона распределение, 1844

Р

Размах, 1849
Распределения
F, 1837
t Стьюдента, 1846
бета, 1831
биномиальное, 1833
Вейбулла, 1846
гамма, 1838
геометрическое, 1839
Гумбеля, 1835
Коши, 1833
Лапласа, 1839
логистическое, 1840
логнормальное, 1841
нормальное, 1842
Парето, 1843
Пуассона, 1844
Релея, 1844
Тип I (Гумбеля), 1835
хи-квадрат, 1834
экспоненциальное, 1834
экстремального значения, 1835
Релея распределение, 1844
вероятностный калькулятор, 1845
Вероятность-Вероятность графики, 1845
Квантиль-Квантиль графики, 1845

С

Случайные числа, 1831
Среднее, 1848
Стандартная ошибка
асимметрии, 1852
среднего, 1851
эксцесса, 1852
Стьюдента t распределение, 1846
вероятностный калькулятор, 1846
Сумма значений, 1848

Т

Типа I (Гумбеля) распределение, 1835

У

Удаление пропущенных данных
подстановка среднего, 1848
попарное, 1848
построчное, 1848
Установки по умолчанию
STATISTICA
процентили, 1849

Ф

Формулы
F распределение, 1837
t распределение Стьюдента, 1846
бета распределение, 1831
биномиальное распределение, 1833
гамма распределение, 1838
геометрическое распределение, 1839
логистическое распределение, 1840
логнормальное распределение, 1841
нормальное распределение, 1842
распределение Вейбулла, 1847
распределение Коши, 1833
распределение Лапласа, 1840
распределение Парето, 1843
распределение Пуассона, 1844
распределение Релея, 1845
распределение экстремального значения, 1836
хи-квадрат распределение, 1834
экспоненциальное распределение, 1834

Х

Хи-квадрат распределение, 1834
вероятностный калькулятор, 1834

Ч

Число наблюдений n, 1848

Э

Экспоненциальное распределение, 1834
вероятностный калькулятор, 1835
Вероятность-Вероятность графики, 1835

StatSoft®

Copyright © StatSoft, 1995

Квантиль-Квантиль графики, 1835
Экстремального значения (Гумбеля) распределение
Вероятность-Вероятность графики, 1836
Квантиль-Квантиль графики, 1836
Экстремального значения распределение, 1835
вероятностный калькулятор, 1836
Эксцесс, 1851

Я

Язык *STATISTICA BASIC*
F распределение, 1837
t распределение Стьюдента, 1846
бета распределение, 1831
биномиальное распределение, 1833
гамма распределение, 1838
геометрическое распределение, 1839
логистическое распределение, 1840
логнормальное распределение, 1841
нормальное распределение, 1842
распределение Вейбулла, 1847
распределение Коши, 1833
распределение Лапласа, 1840
распределение Парето, 1843
распределение Пуассона, 1844
распределение Релея, 1845
распределение экстремального значения, 1836
хи-квадрат распределение, 1834
экспоненциальное распределение, 1834

Copyright © StatSoft, 1995

StatSoft®

Copyright © StatSoft, 1995

СПИСОК ЛИТЕРАТУРЫ

Abraham, B., & Ledolter, J. (1983). *Statistical methods for forecasting*. New York: Wiley.

Adorno, T. W., Frenkel-Brunswik, E., Levinson, D. J., & Sanford, R. N. (1950). *The authoritarian personality.* New York: Harper.

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In B. N. Petrov and F. Csaki (Eds.), *Second International Symposium on Information Theory.* Budapest: Akademiai Kiado.

Akaike, H. (1983). Information measures and model selection. *Bulletin of the International Statistical Institute: Proceedings of the 44th Session, Volume 1.* Pages 277-290.

Aldrich, J. H., & Nelson, F. D. (1984). *Linear probability, logit, and probit models*. Beverly Hills, CA: Sage Publications.

Almon, S. (1965). The distributed lag between capital appropriations and expenditures. *Econometrica*, *33*, 178-196.

American Supplier Institute (1984-1988). *Proceedings of Supplier Symposia on Taguchi Methods.* (April, 1984; November, 1984; October, 1985; October, 1986; October, 1987; October, 1988), Dearborn, MI: American Supplier Institute.

Anderson, O. D. (1976). *Time series analysis and forecasting*. London: Butterworths.

Anderson, S. B., & Maier, M. H. (1963). 34,000 pupils and how they grew. *Journal of Teacher Education*, *14*, 212-216.

Anderson, T. W. (1958). *An introduction to multivariate statistical analysis.* New York: Wiley.

Anderson, T. W., & Rubin, H. (1956). Statistical inference in factor analysis. *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability.* Berkeley: The University of California Press.

Andrews, D. F. (1972). Plots of high-dimensional data. *Biometrics*, *28*, 125-136.

AT&T (1956). *Statistical quality control handbook, Select code 700-444.* Indianapolis, AT&T Technologies.

Auble, D. (1953). Extended tables for the Mann-Whitney statistic. *Bulletin of the Institute of Educational Research, Indiana University*, *1*, No. 2.

Bagozzi, R. P., & Yi, Y. (1989). On the use of structural equation models in experimental design. *Journal of Marketing Research*, *26*, 271-284.

Bagozzi, R. P., Yi, Y., & Singh, S. (1991). On the use of structural equation models in experimental designs: Two extensions. *International Journal of Research in Marketing*, *8*, 125-140.

Bails, D. G., & Peppers, L. C. (1982). *Business fluctuations: Forecasting techniques and applications*. Englewood Cliffs, NJ: Prentice-Hall.

Bain, L. J. (1978). *Statistical analysis of reliability and life-testing models*. New York: Decker.

Baird, J. C. (1970). *Psychophysical analysis of visual space.* New York: Pergamon Press.

Baird, J. C., & Noma, E. (1978). *Fundamentals of scaling and psychophysics.* New York: Wiley.

Barcikowski, R., & Stevens, J. P. (1975). A Monte Carlo study of the stability of canonical correlations, canonical weights, and canonical variate-variable correlations. *Multivariate Behavioral Research*, *10*, 353-364.

Barker, T. B. (1986). Quality engineering by design: Taguchi's philosophy. *Quality Progress*, *19*, 32-42.

Copyright © StatSoft, 1995

Barlow, R. E., & Proschan, F. (1975). *Statistical theory of reliability and life testing*. New York: Holt, Rinehart, & Winston.

Barnard, G. A. (1959). Control charts and stochastic processes. *Journal of the Royal Statistical Society*, Ser. B, *21*, 239.

Bartholomew, D. J. (1984). The foundations of factor analysis. *Biometrika*, *71*, 221-232.

Bates, D. M., & Watts, D. G. (1988). *Nonlinear regression analysis and its applications.* New York: Wiley.

Bayne, C. K., & Rubin, I. B. (1986). *Practical experimental designs and optimization methods for chemists.* Deerfield Beach, FL: VCH Publishers.

Becker, R. A., Denby, L., McGill, R., & Wilks, A. R. (1986). Datacryptanalysis: A case study. *Proceedings of the Section on Statistical Graphics, American Statistical Association*, 92-97.

Bendat, J. S. (1990). *Nonlinear system analysis and identification from random data.* New York: Wiley.

Bentler, P. M, & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin*, *88*, 588-606.

Bentler, P. M. (1986). Structural modeling and Psychometrika: A historical perspective on growth and achievements. *Psychometrika*, *51*, 35-51.

Bentler, P. M. (1989). *EQS Structural equations program manual.* Los Angeles, CA: BMDP Statistical Software.

Bentler, P. M., & Weeks, D. G. (1979). Interrelations among models for the analysis of moment structures. *Multivariate Behavioral Research*, *14*, 169-185.

Bentler, P. M., & Weeks, D. G. (1980). Linear structural equations with latent variables. *Psychometrika*, *45*, 289-308.

Berkson, J., & Gage, R. R. (1950). The calculation of survival rates for cancer. *Proceedings of Staff Meetings, Mayo Clinic*, *25*, 250.

Bhote, K. R. (1988). *World class quality.* New York: AMA Membership Publications.

Binns, B., & Clark, N. (1986). The graphic designer's use of visual syntax. *Proceedings of the Section on Statistical Graphics, American Statistical Association*, 36-41.

Birnbaum, Z. W. (1952). Numerical tabulation of the distribution of Kolmogorov's statistic for finite sample values. *Journal of the American Statistical Association*, *47*, 425-441.

Birnbaum, Z. W. (1953). Distribution-free tests of fit for continuous distribution functions. *Annals of Mathematical Statistics*, *24*, 1-8.

Bishop, Y. M. M., Fienberg, S. E., & Holland, P. W. (1975). *Discrete multivariate analysis*. Cambridge, MA: MIT Press.

Bjorck, A. (1967). Solving linear least squares problems by Gram-Schmidt orthonormalization. *Bit*, *7*, 1-21.

Blackman, R. B., & Tukey, J. (1958). *The measurement of power spectral from the point of view of communication engineering.* New York: Dover.

Blackwelder, R. A. (1966). *Taxonomy: A text and reference book.* New York: Wiley.

Blalock, H. M. (1972). *Social statistics* (2nd ed.). New York: McGraw-Hill.

Bloomfield, P. (1976). *Fourier analysis of time series: An introduction*. New York: Wiley.

Bock, R. D. (1963). Programming univariate and multivariate analysis of variance. *Technometrics*, *5*, 95-117.

StatSoft®

Copyright © StatSoft, 1995

Bock, R. D. (1975). *Multivariate statistical methods in behavioral research*. New York: McGraw-Hill.

Bolch, B.W., & Huang, C. J. (1974). *Multivariate statistical methods for business and economics*. Englewood Cliffs, NJ: Prentice-Hall.

Bollen, K. A. (1989). *Structural equations with latent variables.* New York: John Wiley & Sons.

Borg, I., & Lingoes, J. (1987). *Multidimensional similarity structure analysis*. New York: Springer.

Borg, I., & Shye, S. (in press). *Facet Theory*. Newbury Park: Sage.

Bowker, A. G. (1948). A test for symmetry in contingency tables. *Journal of the American Statistical Association*, *43*, 572-574.

Box, G. E. P. (1954a). Some theorems on quadratic forms applied in the study of analysis of variance problems: I. Effect of inequality of variances in the one-way classification. *Annals of Mathematical Statistics*, *25*, 290-302.

Box, G. E. P. (1954b). Some theorems on quadratic forms applied in the study of analysis of variance problems: II. Effect of inequality of variances and of correlation of errors in the two-way classification. *Annals of Mathematical Statistics*, *25*, 484-498.

Box, G. E. P., & Anderson, S. L. (1955). Permutation theory in the derivation of robust criteria and the study of departures from assumptions. *Journal of the Royal Statistical Society*, *17*, 1-34.

Box, G. E. P., & Behnken, D. W. (1960). Some new three level designs for the study of quantitative variables. *Technometrics*, *2*, 455-475.

Box, G. E. P., & Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society*, *26*, 211-253.

Box, G. E. P., & Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society, B26*, 211-234.

Box, G. E. P., & Draper, N. R. (1987). *Empirical model-building and response surfaces.* New York: Wiley.

Box, G. E. P., & Jenkins, G. M. (1970). *Time series analysis*. San Francisco: Holden Day.

Box, G. E. P., & Jenkins, G. M. (1976). *Time series analysis: Forecasting and control*. San Francisco: Holden-Day.

Box, G. E. P., & Tidwell, P. W. (1962). Transformation of the independent variables. *Technometrics*, *4*, 531-550.

Box, G. E. P., Hunter, W. G., & Hunter, S. J. (1978). *Statistics for experimenters: An introduction to design, data analysis, and model building.* New York: Wiley.

Brenner, J. L., et al. (1968). Difference equations in forecasting formulas. *Management Science*, *14*, 141-159.

Brent, R. F. (1973). *Algorithms for minimization without derivatives.* Englewood Cliffs, NJ: Prentice-Hall.

Breslow, N. E. (1970). A generalized Kruskal-Wallis test for comparing *K* samples subject to unequal pattern of censorship. *Biometrika*, *57*, 579-594.

Breslow, N. E. (1974). Covariance analysis of censored survival data. *Biometrics*, *30*, 89-99.

Brigham, E. O. (1974). *The fast Fourier transform.* Englewood Cliffs, NJ: Prentice-Hall.

Brillinger, D. R. (1975). *Time series: Data analysis and theory.* New York: Holt, Rinehart. & Winston.

Brown, D. T. (1959). A note on approximations to discrete probability distributions. *Information and Control*, *2*, 386-392.

Brown, R. G. (1959). *Statistical forecasting for inventory control*. New York: McGraw-Hill.

Browne, M. W. (1968). A comparison of factor analytic techniques. *Psychometrika*, *33*, 267-334.

Copyright © StatSoft, 1995

Browne, M. W. (1974). Generalized least squares estimators in the analysis of covariance structures. *South African Statistical Journal*, *8*, 1-24.

Browne, M. W. (1982). Covariance Structures. In D. M. Hawkins (Ed.) *Topics in Applied Multivariate Analysis.* Cambridge, MA: Cambridge University Press.

Browne, M. W. (1984). Asymptotically distribution free methods for the analysis of covariance structures. *British Journal of Mathematical and Statistical Psychology*, *37*, 62-83.

Browne, M. W., & Cudeck, R. (1990). Single sample cross-validation indices for covariance structures. *Multivariate Behavioral Research*, *24*, 445-455.

Browne, M. W., & Cudeck, R. (1992). Alternative ways of assessing model fit. In K. A. Bollen and J. S. Long (Eds.), *Testing structural equation models.* Beverly Hills, CA: Sage.

Browne, M. W., & DuToit, S. H. C. (1982). *AUFIT* (Version 1). A computer programme for the automated fitting of nonstandard models for means and covariances. Research Finding WS-27. Pretoria, South Africa: Human Sciences Research Council.

Browne, M. W., & DuToit, S. H. C. (1987). *Automated fitting of nonstandard models*. Report WS-39. Pretoria, South Africa: Human Sciences Research Council.

Browne, M. W., & DuToit, S. H. C. (1992). Automated fitting of nonstandard models. *Multivariate Behavioral Research*, *27*, 269-300.

Browne, M. W., & Mels, G. (1992). *RAMONA User's Guide.* The Ohio State University: Department of Psychology.

Browne, M. W., & Shapiro, A. (1989). *Invariance of covariance structures under groups of transformations*. Research Report 89/4. Pretoria, South Africa: University of South Africa Department of Statistics.

Browne, M. W., & Shapiro, A. (1991). Invariance of covariance structures under groups of transformations. *Metrika*, *38*, 335-345.

Buffa, E. S. (1972). *Operations management: Problems and models* (3rd. ed.). New York: Wiley.

Buja, A., & Tukey, P. A. (Eds.) (1991). *Computing and Graphics in Statistics.* New York: Springer-Verlag.

Buja, A., Fowlkes, E. B., Keramidas, E. M., Kettenring, J. R., Lee, J. C., Swayne, D. F., & Tukey, P. A. (1986). Discovering features of multivariate data through statistical graphics. *Proceedings of the Section on Statistical Graphics, American Statistical Association*, 98-103.

Burman, J. P. (1979). Seasonal adjustment - a survey. *Forecasting, Studies in Management Science*, *12*, 45-57.

Burns, L. S., & Harman, A. J. (1966). *The complex metropolis, Part V of profile of the Los Angeles metropolis: Its people and its homes.* Los Angeles: University of Chicago Press.

Campbell D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, *56*, 81-105

Carmines, E. G., & Zeller, R. A. (1980). *Reliability and validity assessment*. Beverly Hills, CA: Sage Publications.

Carroll, J. D., & Wish, M. (1974). Multidimensional perceptual models and measurement methods. In E. C. Carterette and M. P. Friedman (Eds.), *Handbook of perception*. (Vol. 2, pp. 391-447). New York: Academic Press.

StatSoft®

Copyright © StatSoft, 1995

Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research*, *1*, 245-276.

Cattell, R. B., & Jaspers, J. A. (1967). A general plasmode for factor analytic exercises and research. *Multivariate Behavioral Research Monographs.*

Chambers, J. M., Cleveland, W. S., Kleiner, B., & Tukey, P. A. (1983). *Graphical methods for data analysis*. Bellmont, CA: Wadsworth.

Chan, L. K., Cheng, S. W., & Spiring, F. (1988). A new measure of process capability: C_{pm}. *Journal of Quality Technology*, *20*, 162-175.

Chernoff, H. (1973). The use of faces to represent points in k-dimensional space graphically. *Journal of American Statistical Association*, *68*, 361-368.

Christ, C. (1966). *Econometric models and methods*. New York: Wiley.

Clarke, G. M., & Cooke, D. (1978). *A basic course in statistics*. London: Edward Arnold.

Cleveland, W. S. (1979). Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association*, *74*, 829-836.

Cleveland, W. S. (1984). Graphs in scientific publications. *The American Statistician*, *38*, 270-280.

Cleveland, W. S. (1985). *The elements of graphing data*. Monterey, CA: Wadsworth.

Cleveland, W. S., Harris, C. S., & McGill, R. (1982). Judgements of circle sizes on statistical maps. *Journal of the American Statistical Association*, *77*, 541-547.

Cliff, N. (1983). Some cautions concerning the application of causal modeling methods. *Multivariate Behavioral Research*, *18*, 115-126.

Cochran, W. G. (1950). The comparison of percentages in matched samples. *Biometrika*, *37*, 256-266.

Cole, D. A., Maxwell, S. E., Arvey, R., & Salas, E. (1993). Multivariate group comparisons of variable systems: MANOVA and structural equation modeling. *Psychological Bulletin*, *114*, 174-184.

Conover, W. J. (1974). Some reasons for not using the Yates continuity correction on 2 x 2 contingency tables. *Journal of the American Statistical Association*, *69*, 374-376.

Cook, R. D. (1977). Detection of influential observations in linear regression. *Technometrics*, *19*, 15-18.

Cook, R. D., & Nachtsheim, C. J. (1980). A comparison of algorithms for constructing exact D-optimal designs. *Technometrics*, *22*, 315-324.

Cooke, D., Craven, A. H., & Clarke, G. M. (1982). *Basic statistical computing*. London: Edward Arnold.

Cooley, J. W., & Tukey, J. W. (1965). An algorithm for the machine computation of complex Fourier series. *Mathematics of Computation*, *19*, 297-301.

Cooley, W. W., & Lohnes, P. R. (1971). *Multivariate data analysis.* New York: Wiley.

Cooley, W. W., & Lohnes, P. R. (1976). *Evaluation research in education*. New York: Wiley.

Coombs, C. H. (1950). Psychological scaling without a unit of measurement. *Psychological Review*, *57*, 145-158.

Coombs, C. H. (1964). *A theory of data.* New York: Wiley.

Corballis, M. C., & Traub, R. E. (1970). Longitudinal factor analysis. *Psychometrika*, *35*, 79-98.

Cormack, R. M. (1971). A review of classification. *Journal of the Royal Statistical Society*, *134*, 321-367.

Cornell, J. A. (1990a). *How to run mixture experiments for product quality.* Milwaukee, WI: ASQC.

Copyright © StatSoft, 1995

Cornell, J. A. (1990b). *Experiments with mixtures: designs, models, and the analysis of mixture data* (2nd ed.). New York: Wiley.

Cox, D. R. (1957). Note on grouping. *Journal of the American Statistical Association*, *52*, 543-547.

Cox, D. R. (1959). The analysis of exponentially distributed life-times with two types of failures. *Journal of the Royal Statistical Society*, *21*, 411-421.

Cox, D. R. (1964). Some applications of exponential ordered scores. *Journal of the Royal Statistical Society*, *26*, 103-110.

Cox, D. R. (1970). *The analysis of binary data*. New York: Halsted Press.

Cox, D. R. (1972). Regression models and life tables. *Journal of the Royal Statistical Society*, *34*, 187-220.

Cox, D. R., & Oakes, D. (1984). *Analysis of survival data*. New York: Chapman & Hall.

Cramer, H. (1946). *Mathematical methods in statistics*. Princeton, NJ: Princeton University Press.

Crowley, J., & Hu, M. (1977). Covariance analysis of heart transplant survival data. *Journal of the American Statistical Association*, *72*, 27-36.

Cudeck, R. (1989). Analysis of correlation matrices using covariance structure models. *Psychological Bulletin*, *105*, 317-327.

Cudeck, R., & Browne, M. W. (1983). Cross-validation of covariance structures. *Multivariate Behavioral Research*, *18*, 147-167.

Cutler, S. J., & Ederer, F. (1958). Maximum utilization of the life table method in analyzing survival. *Journal of Chronic Diseases*, *8*, 699-712.

Dahlquist, G., & Bjorck, A. (1974). *Numerical Methods.* Englewood Cliffs, NJ: Prentice-Hall.

Daniel, C. (1976). *Applications of statistics to industrial experimentation.* New York: Wiley.

Daniell, P. J. (1946). Discussion on symposium on autocorrelation in time series. *Journal of the Royal Statistical Society, Suppl*. *8*, 88-90.

Darlington, R. B. (1990). *Regression and linear models*. New York: McGraw-Hill.

Darlington, R. B., Weinberg, S., & Walberg, H. (1973). Canonical variate analysis and related techniques. *Review of Educational Research*, *43*, 433-454.

Davies, P. M., & Coxon, A. P. M. (1982). *Key texts in multidimensional scaling.* Exeter, NH: Heinemann Educational Books.

De Boor, C. (1978). *A practical guide to splines*. New York: Springer-Verlag.

De Gruitjer, P. N. M., & Van Der Kamp, L. J. T. (Eds.). (1976). *Advances in psychological and educational measurement*. New York: Wiley.

Deming, S. N., & Morgan, S. L. (1993). *Experimental design: A chemometric approach.* (2nd ed.). Amsterdam, The Netherlands: Elsevier Science Publishers B.V.

Deming, W. E., & Stephan, F. F. (1940). The sampling procedure of the 1940 population census. *Journal of the American Statistical Association*, *35*, 615-630.

Dempster, A. P. (1969). *Elements of Continuous Multivariate Analysis*. San Francisco: Addison-Wesley.

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society*, *39*, 1-38.

Dennis, J. E., & Schnabel, R. B. (1983). *Numerical methods for unconstrained optimization and nonlinear equations.* Englewood Cliffs, NJ: Prentice Hall.

Diamond, W. J. (1981). *Practical experimental design.* Belmont, CA: Wadsworth.

StatSoft®

Copyright © StatSoft, 1995

Dijkstra, T. K. (1990). Some properties of estimated scale invariant covariance structures. *Psychometrika*, *55*, 327-336.

Dixon, W. J. (1954). Power under normality of several non-parametric tests. *Annals of Mathematical Statistics*, *25*, 610-614.

Dixon, W. J., & Massey, F. J. (1983). *Introduction to statistical analysis* (4th ed.). New York: McGraw-Hill.

Dodd, B. (1979). Lip reading in infants: Attention to speech presented in- and out-of-synchrony. *Cognitive Psychology*, *11*, 478-484.

Dodge, Y. (1985). *Analysis of experiments with missing data*. New York: Wiley.

Dodge, Y., Fedorov, V. V., & Wynn, H. P. (1988). *Optimal design and analysis of experiments.* New York: North-Holland.

Duncan, A. J. (1974). *Quality control and industrial statistics*. Homewood, IL: Richard D. Irwin.

Duncan, O. D., Haller, A. O., & Portes, A. (1968). Peer influence on aspiration: A reinterpretation. *American Journal of Sociology*, *74*, 119-137.

Durbin, J. (1970). Testing for serial correlation in least-squares regression when some of the regressors are lagged dependent variables. *Econometrica*, *38*, 410-421.

Durbin, J., & Watson, G. S. (1951). Testing for serial correlations in least squares regression. II. *Biometrika*, *38*, 159-178.

Dykstra, O. Jr. (1971). The augmentation of experimental data to maximize |X'X|. *Technometrics*, *13*, 682-688.

Eason, E. D., & Fenton, R. G. (1974). A comparison of numerical optimization methods for engineering design. *ASME Paper 73-DET-17*.

Efron, B. (1982). *The jackknife, the bootstrap, and other resampling plans.* Philadelphia, PA: Society for Industrial and Applied Mathematics.

Elandt-Johnson, R. C., & Johnson, N. L. (1980). *Survival models and data analysis*. New York: Wiley.

Elliott, D. F., & Rao, K. R. (1982). *Fast transforms: Algorithms, analyses, applications.* New York: Academic Press.

Enslein, K., Ralston, A., & Wilf, H. S. (1977). *Statistical methods for digital computers.* New York: Wiley.

Everitt, B. S. (1977). *The analysis of contingency tables.* London: Chapman & Hall.

Everitt, B. S. (1984). *An introduction to latent variable models*. London: Chapman and Hall.

Ewan, W. D. (1963). When and how to use Cu-sum charts. *Technometrics*, *5*, 1-32.

Feigl, P., & Zelen, M. (1965). Estimation of exponential survival probabilities with concomitant information. *Biometrics*, *21*, 826-838.

Fetter, R. B. (1967). *The quality control system*. Homewood, IL: Richard D. Irwin.

Fienberg, S. E. (1977). *The analysis of cross-classified categorical data*. Cambridge, MA: MIT Press.

Finn, J. D. (1974). *A general model for multivariate analysis.* New York: Holt, Rinehart & Winston.

Finn, J. D. (1977). Multivariate analysis of variance and covariance. In K. Enslein, A. Ralston, and H. S. Wilf (Eds.), *Statistical methods for digital computers. Vol. III*. (pp. 203-264). New York: Wiley.

Finney, D. J. (1971). *Probit analysis*. Cambridge, MA: Cambridge University Press.

Fisher, R. A. (1936). The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, *7*, 179-188.

Fletcher, R. (1969). *Optimization*. New York: Academic Press.

Copyright © StatSoft, 1995

Fletcher, R., & Powell, M. J. D. (1963). A rapidly convergent descent method for minimization. *Computer Journal*, *6*, 163-168.

Fletcher, R., & Reeves, C. M. (1964). Function minimization by conjugate gradients. *Computer Journal*, *7*, 149-154.

Fomby, T.B., Hill, R.C., & Johnson, S.R. (1984). *Advanced econometric methods*. New York: Springer-Verlag.

Fraser, C., & McDonald, R. P. (1988). COSAN: Covariance structure analysis. *Multivariate Behavioral Research*, *23*, 263-265.

Friedman, M. (1937). The use of ranks to avoid the assumption of normality implicit in the analysis of variance. *Journal of the American Statistical Association*, *32*, 675-701.

Friedman, M. (1940). A comparison of alternative tests of significance for the problem of *m* rankings. *Annals of Mathematical Statistics*, *11*, 86-92.

Fries, A., & Hunter, W. G. (1980). Minimum aberration $2^{(k-p)}$ designs. *Technometrics*, *22*, 601-608.

Frost, P. A. (1975). Some properties of the Almon lag technique when one searches for degree of polynomial and lag. *Journal of the American Statistical Association*, *70*, 606-612.

Fuller, W. A. (1976). *Introduction to statistical time series*. New York: Wiley.

Gale, N., & Halperin, W. C. (1982). A case for better graphics: The unclassed choropleth map. *The American Statistician*, *36*, 330-336.

Galil, Z., & Kiefer, J. (1980). Time- and space-saving computer methods, related to Mitchell's DETMAX, for finding D-optimum designs. *Technometrics*, *22*, 301-313.

Gara, M. A., & Rosenberg, S. (1979). The identification of persons as supersets and subsets in free-response personality descriptions. *Journal of Personality and Social Psychology*, *37*, 2161-2170.

Gara, M. A., & Rosenberg, S. (1981). Linguistic factors in implicit personality theory. *Journal of Personality and Social Psychology*, *41*, 450-457.

Gardner, E. S., Jr. (1985). Exponential smoothing: The state of the art. *Journal of Forecasting*, *4*, 1-28.

Garvin, D. A. (1987). Competing on the eight dimensions of quality. *Harvard Business Review*, November/December, 101-109.

Gbur, E., Lynch, M., & Weidman, L. (1986). An analysis of nine rating criteria on 329 U. S. metropolitan areas. *Proceedings of the Section on Statistical Graphics, American Statistical Association*, 104-109.

Gedye, R. (1968). *A manager's guide to quality and reliability*. New York: Wiley.

Gehan, E. A. (1965a). A generalized Wilcoxon test for comparing arbitrarily singly-censored samples. *Biometrika*, *52*, 203-223.

Gehan, E. A. (1965b). A generalized two-sample Wilcoxon test for doubly-censored data. *Biometrika*, *52*, 650-653.

Gehan, E. A., & Siddiqui, M. M. (1973). Simple regression methods for survival time studies. *Journal of the American Statistical Association*, *68*, 848-856.

Gehan, E. A., & Thomas, D. G. (1969). The performance of some two sample tests in small samples with and without censoring. *Biometrika*, *56*, 127-132.

Gerald, C. F., & Wheatley, P. O. (1989). *Applied numerical analysis* (4th ed.). Reading, MA: Addison Wesley.

StatSoft®

Copyright © StatSoft, 1995

Gibbons, J. D. (1976). *Nonparametric methods for quantitative analysis.* New York: Holt, Rinehart, & Winston.

Gibbons, J. D. (1985). *Nonparametric statistical inference* (2nd ed.). New York: Marcel Dekker.

Gifi, A. (1990). *Nonlinear multivariate analysis.* New York: Wiley.

Gill, P. E., & Murray, W. (1972). Quasi-Newton methods for unconstrained optimization. *Journal of the Institute of Mathematics and its Applications*, *9*, 91-108.

Gill, P. E., & Murray, W. (1974). *Numerical methods for constrained optimization.* New York: Academic Press.

Glass, G. V., & Stanley, J. (1970). *Statistical methods in education and Psychology.* Englewood Cliffs, NJ: Prentice-Hall.

Glasser, M. (1967). Exponential survival with covariance. *Journal of the American Statistical Association*, *62*, 561-568.

Gnanadesikan, R., Roy, S., & Srivastava, J. (1971). *Analysis and design of certain quantitative multiresponse experiments*. Oxford: Pergamon Press, Ltd.

Golub, G. H., & Van Loan, C. F. (1983). *Matrix computations.* Baltimore: Johns Hopkins University Press.

Goodman, L .A., & Kruskal, W. H. (1972). Measures of association for cross-classifications IV: Simplification of asymptotic variances. *Journal of the American Statistical Association*, *67*, 415-421.

Goodman, L. A. (1954). Kolmogorov-Smirnov tests for psychological research. *Psychological Bulletin*, *51*, 160-168.

Goodman, L. A. (1971). The analysis of multidimensional contingency tables: Stepwise procedures and direct estimation methods for models building for multiple classification. *Technometrics*, *13*, 33-61.

Goodman, L. A., & Kruskal, W. H. (1954). Measures of association for cross-classifications. *Journal of the American Statistical Association*, *49*, 732-764.

Goodman, L. A., & Kruskal, W. H. (1959). Measures of association for cross-classifications II: Further discussion and references. *Journal of the American Statistical Association*, *54*, 123-163.

Goodman, L. A., & Kruskal, W. H. (1963). Measures of association for cross-classifications III: Approximate sampling theory. *Journal of the American Statistical Association*, *58*, 310-364.

Grant, E. L., & Leavenworth, R. S. (1980). *Statistical quality control* (5th ed.). New York: McGraw-Hill.

Green, P. E., & Carmone, F. J. (1970). *Multidimensional scaling and related techniques in marketing analysis.* Boston: Allyn & Bacon.

Greenhouse, S. W., & Geisser, S. (1958). Extension of Box's results on the use of the *F* distribution in multivariate analysis. *Annals of Mathematical Statistics*, *29*, 95-112.

Greenhouse, S. W., & Geisser, S. (1959). On methods in the analysis of profile data. *Psychometrika*, *24*, 95-112.

Gross, A. J., & Clark, V. A. (1975). *Survival distributions: Reliability applications in the medical sciences*. New York: Wiley.

Gruvaeus, G., & Wainer, H. (1972). Two additions to hierarchical cluster analysis. *The British Journal of Mathematical and Statistical Psychology*, *25*, 200-206.

Guttman, L. (1954). A new approach to factor analysis: the radex. In P. F. Lazarsfeld (Ed.), *Mathematical thinking in the social sciences.* New York: Columbia University Press.

Guttman, L. (1968). A general nonmetric technique for finding the smallest coordinate space for a configuration of points. *Pyrometrical*, *33*, 469-506.

Copyright © StatSoft, 1995

Haberman, S. J. (1972). Loglinear fit for contingency tables. *Applied Statistics*, *21*, 218-225.

Haberman, S. J. (1974). *The analysis of frequency data*. Chicago: University of Chicago Press.

Hahn, G. J., & Shapiro, S. S. (1967). *Statistical models in engineering*. New York: Wiley.

Hakstian, A. R., Rogers, W. D., & Cattell, R. B. (1982). The behavior of numbers of factors rules with simulated data. *Multivariate Behavioral Research*, *17*, 193-219.

Hald, A. (1952). *Statistical theory with engineering applications.* New York: Wiley.

Harman, H. H. (1967). *Modern factor analysis.* Chicago: University of Chicago Press.

Harris, R. J. (1976). The invalidity of partitioned *U* tests in canonical correlation and multivariate analysis of variance. *Multivariate Behavioral Research*, *11*, 353-365.

Hart, K. M., & Hart, R. F. (1989). *Quantitative methods for quality improvement.* Milwaukee, WI: ASQC Quality Press.

Hartigan, J. A. (1975). *Clustering algorithms.* New York: Wiley.

Hartley, H. O. (1959). Smallest composite designs for quadratic response surfaces. *Biometrics*, *15*, 611-624.

Haviland, R. P. (1964). *Engineering reliability and long life design*. Princeton, NJ: Van Nostrand.

Hayduk, L. A. (1987). *Structural equation modelling with LISREL: Essentials and advances.* Baltimore: The Johns Hopkins University Press.

Hays, W. L. (1981). *Statistics* (3rd ed.). New York: CBS College Publishing.

Hays, W. L. (1988). *Statistics* (4th ed.). New York: CBS College Publishing.

Heiberger, R. M. (1989). *Computation for the analysis of designed experiments*. New York: Wiley.

Henley, E. J., & Kumamoto, H. (1980). *Reliability engineering and risk assessment*. New York: Prentice-Hall.

Hettmansperger, T. P. (1984). *Statistical inference based on ranks*. New York: Wiley.

Hibbs, D. (1974). Problems of statistical estimation and causal inference in dynamic time series models. In H. Costner (Ed.), *Sociological Methodology 1973/1974* (pp. 252-308). San Francisco: Jossey-Bass.

Hilton, T. L. (1969). *Growth study annotated bibliography. Princeton, NJ: Educational Testing Service Progress Report 69-11.*

Hochberg, J., & Krantz, D. H. (1986). Perceptual properties of statistical graphs. *Proceedings of the Section on Statistical Graphics, American Statistical Association*, 29-35.

Hocking, R. R., & Speed, F. M. (1975). A full rank analysis of some linear model problems. *Journal of the American Statistical Association*, *70*, 707-712.

Hoerl, A. E. (1962). Application of ridge analysis to regression problems. *Chemical Engineering Progress*, *58*, 54-59.

Hoff, J. C. (1983). *A practical guide to Box-Jenkins forecasting*. London: Lifetime Learning Publications.

Hogg, R. V., & Craig, A. T. (1970). *Introduction to mathematical statistics*. New York: Macmillan.

Holzinger, K. J., & Swineford, F. (1939). *A study in factor analysis: The stability of a bi-factor solution.* University of Chicago: Supplementary Educational Monographs, No. 48.

Hooke, R., & Jeeves, T. A. (1961). Direct search solution of numerical and statistical problems. *Journal of the Association for Computing Machinery*, *8*, 212-229.

Hotelling, H. (1947). Multivariate quality control. In Eisenhart, Hastay, and Wallis (Eds.), *Techniques of Statistical Analysis*. New York: McGraw-Hill.

StatSoft®

Copyright © StatSoft, 1995

Hotelling, H., & Pabst, M. R. (1936). Rank correlation and tests of significance involving no assumption of normality. *Annals of Mathematical Statistics*, *7*, 29-43.

Hsu, P. L. (1938). Contributions to the theory of Student's *t* test as applied to the problem of two samples. *Statistical Research Memoirs*, *2*, 1-24.

Huba, G. J., & Harlow, L. L. (1987). Robust structural equation models: implications for developmental psychology. *Child Development*, *58*, 147-166.

Huberty, C. J. (1975). Discriminant analysis. *Review of Educational Research*, *45*, 543-598.

Huynh, H., & Feldt, L. S. (1970). Conditions under which mean square ratios in repeated measures designs have exact *F*-distributions. *Journal of the American Statistical Association*, *65*, 1582-1589.

Ireland, C. T., & Kullback, S. (1968). Contingency tables with given marginals. *Biometrika*, *55*, 179-188.

Jaccard, J., Weber, J., & Lundmark, J. (1975). A multitrait-multimethod factor analysis of four attitude assessment procedures. *Journal of Experimental Social Psychology*, *11*, 149-154.

Jacobs, D. A. H. (Ed.). (1977). *The state of the art in numerical analysis.* London: Academic Press.

Jacoby, S. L. S., Kowalik, J. S., & Pizzo, J. T. (1972). *Iterative methods for nonlinear optimization problems.* Englewood Cliffs, NJ: Prentice-Hall.

James, L. R., Mulaik, S. A., & Brett, J. M. (1982). *Causal analysis. Assumptions, models, and data.* Beverly Hills, CA: Sage Publications.

Jardine, N., & Sibson, R. (1971). *Mathematical taxonomy.* New York: Wiley.

Jastrow, J. (1892). On the judgment of angles and position of lines. *American Journal of Psychology*, *5*, 214-248.

Jenkins, G. M., & Watts, D. G. (1968). *Spectral analysis and its applications.* San Francisco: Holden-Day.

Jennrich, R. I, & Sampson, P. F. (1968). Application of stepwise regression to non-linear estimation. *Technometrics*, *10*, 63-72.

Jennrich, R. I. (1970). An asymptotic χ^2 test for the equality of two correlation matrices. *Journal of the American Statistical Association*, *65*, 904-912.

Jennrich, R. I. (1977). Stepwise regression. In K. Enslein, A. Ralston, & H.S. Wilf (Eds.), *Statistical methods for digital computers*. New York: Wiley.

Jennrich, R. I. (1977). Stepwise discriminant analysis. In K. Enslein, A. Ralston, & H.S. Wilf (Eds.), *Statistical methods for digital computers*. New York: Wiley.

Jennrich, R. I., & Moore, R. H. (1975). Maximum likelihood estimation by means of nonlinear least squares. *Proceedings of the Statistical Computing Section*, American Statistical Association, 57-65.

Johnson, L. W., & Ries, R. D. (1982). *Numerical Analysis* (2nd ed.). Reading, MA: Addison Wesley.

Johnson, N. L. (1961). A simple theoretical approach to cumulative sum control charts. *Journal of the American Statistical Association*, *56*, 83-92.

Johnson, N. L., & Leone, F. C. (1962). Cumulative sum control charts - mathematical principles applied to their construction and use. *Industrial Quality Control*, *18*, 15-21.

Johnson, R. A. & Wichern, D. W. (1988). *Applied multivariate statistical analysis*. Englewood Cliffs, NJ: Prentice Hall.

Johnson, S. C. (1967). Hierarchical clustering schemes. *Psychometrika*, *32*, 241-254.

Copyright © StatSoft, 1995

Johnston, J. (1972). *Econometric methods*. New York: McGraw-Hill.

Jöreskog, K. G. (1973). A general model for estimating a linear structural equation system. In A. S. Goldberger and O. D. Duncan (Eds.), *Structural Equation Models in the Social Sciences.* New York: Seminar Press.

Jöreskog, K. G. (1974). Analyzing psychological data by structural analysis of covariance matrices. In D. H. Krantz, R. C. Atkinson, R. D. Luce, and P. Suppes (Eds.), *Contemporary Developments in Mathematical Psychology, Vol. II.* New York: W. H. Freeman and Company.

Jöreskog, K. G. (1978). Structural analysis of covariance and correlation matrices. *Psychometrika*, *43*, 443-477.

Jöreskog, K. G., & Lawley, D. N. (1968). New methods in maximum likelihood factor analysis. *British Journal of Mathematical and Statistical Psychology*, *21*, 85-96.

Jöreskog, K. G., & Sörbom, D. (1979). *Advances in factor analysis and structural equation models*. Cambridge, MA: Abt Books.

Jöreskog, K. G., & Sörbom, D. (1982). Recent developments in structural equation modeling. *Journal of Marketing Research*, *19*, 404-416.

Jöreskog, K. G., & Sörbom, D. (1984). *Lisrel VI. Analysis of linear structural relationships by maximum likelihood, instrumental variables, and least squares methods*. Mooresville, Indiana: Scientific Software.

Jöreskog, K. G., & Sörbom, D. (1989). *Lisrel 7. A guide to the program and applications.* Chicago: SPSS Inc.

Judge, G. G., Griffith, W. E., Hill, R. C., Luetkepohl, H., & Lee, T. S. (1985). *The theory and practice of econometrics*. New York: Wiley.

Juran, J. M. (1960). Pareto, Lorenz, Cournot, Bernnouli, Juran and others. *Industrial Quality Control*, *17*, 25.

Juran, J. M. (1962). *Quality control handbook*. New York: McGraw-Hill.

Juran, J. M., & Gryna, F. M. (1970). *Quality planning and analysis*. New York: McGraw-Hill.

Juran, J. M., & Gryna, F. M. (1980). *Quality planning and analysis* (2nd ed.). New York: McGraw-Hill.

Juran, J. M., & Gryna, F. M. (1988). *Juran's quality control handbook* (4th ed.). New York: McGraw-Hill.

Kachigan, S. K. (1986). *Statistical analysis: An interdisciplinary introduction to univariate & multivariate methods*. New York: Redius Press.

Kackar, R. M. (1985). Off-line quality control, parameter design, and the Taguchi method. *Journal of Quality Technology*, *17*, 176-188.

Kackar, R. M. (1986). Taguchi's quality philosophy: Analysis and commentary. *Quality Progress*, *19*, 21-29.

Kaiser, H. F. (1958). The varimax criterion for analytic rotation in factor analysis. *Pyrometrical*, *23*, 187-200.

Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement*, *20*, 141-151.

Kalbfleisch, J. D., & Prentice, R. L. (1980). *The statistical analysis of failure time data*. New York: Wiley.

Kane, V. E. (1986). Process capability indices. *Journal of Quality Technology*, *18*, 41-52.

Kaplan, E. L., & Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, *53*, 457-481.

Karsten, K. G., (1925). *Charts and graphs*. New York: Prentice-Hall.

StatSoft®

Copyright © StatSoft, 1995

Keeves, J. P. (1972). *Educational environment and student achievement.* Melbourne: Australian Council for Educational Research.

Kendall, M. G. (1948). *Rank correlation methods*. (1st ed.). London: Griffin.

Kendall, M. G. (1975). *Rank correlation methods* (4th ed.). London: Griffin.

Kendall, M. G. (1984). *Time Series*. New York: Oxford University Press.

Kendall, M., & Ord, J. K. (1990). *Time series* (3rd ed.). London: Griffin.

Kendall, M., & Stuart, A. (1979). *The advanced theory of statistics* (Vol. 2). New York: Hafner.

Kennedy, A. D., & Gehan, E. A. (1971). Computerized simple regression methods for survival time studies. *Computer Programs in Biomedicine*, *1*, 235-244.

Kennedy, W. J., & Gentle, J. E. (1980). *Statistical computing.* New York: Marcel Dekker, Inc.

Kenny, D. A. (1979). *Correlation and causality.* New York: Wiley.

Keppel, G. (1973). *Design and analysis: A researcher's handbook*. Engelwood Cliffs, NJ: Prentice-Hall.

Keppel, G. (1982). *Design and analysis: A researcher's handbook* (2nd ed.). Engelwood Cliffs, NJ: Prentice-Hall.

Keselman, H. J., Rogan, J. C., Mendoza, J. L., & Breen, L. L. (1980). Testing the validity conditions for repeated measures *F* tests. *Psychological Bulletin*, *87*, 479-481.

Khuri, A. I., & Cornell, J. A. (1987). *Response surfaces: Designs and analyses*. New York: Marcel Dekker, Inc.

Kiefer, J., & Wolfowitz, J. (1960). The equivalence of two extremum problems. *Canadian Journal of Mathematics*, *12*, 363-366.

Kim, J. O., & Mueller, C. W. (1978a). *Factor analysis: Statistical methods and practical issues*. Beverly Hills, CA: Sage Publications.

Kim, J. O., & Mueller, C. W. (1978b). *Introduction to factor analysis: What it is and how to do it.* Beverly Hills, CA: Sage Publications.

Kirk, D. B. (1973). On the numerical approximation of the bivariate normal (tetrachoric) correlation coefficient. *Psychometrika*, *38*, 259-268.

Kirk, R. E. (1968). *Experimental design: Procedures for the behavioral sciences*. (1st ed.). Monterey, CA: Brooks/Cole.

Kirk, R. E. (1982). *Experimental design: Procedures for the behavioral sciences*. (2nd ed.). Monterey, CA: Brooks/Cole.

Kivenson, G. (1971). *Durability and reliability in engineering design*. New York: Hayden.

Klecka, W. R. (1980). *Discriminant analysis*. Beverly Hills, CA: Sage.

Klein, L. R. (1974). *A textbook of econometrics*. Englewood Cliffs, NJ: Prentice-Hall.

Kline, P. (1979). *Psychometrics and psychology*. London: Academic Press.

Kline, P. (1986). *A handbook of test construction.* New York: Methuen.

Kmenta, J. (1971). *Elements of econometrics*. New York: Macmillan.

Knuth, Donald E. (1981). *Seminumerical algorithms.* 2nd ed., Vol 2 of: *The art of computer programming*. Reading, Mass.: Addison-Wesley.

Kolata, G. (1984). The proper display of data. *Science*, *226*, 156-157.

Kolmogorov, A. (1941). Confidence limits for an unknown distribution function. *Annals of Mathematical Statistics*, *12*, 461-463.

Korin, B. P. (1969). On testing the equality of *k* covariance matrices. *Biometrika*, *56*, 216-218.

Copyright © StatSoft, 1995

Kruskal, J. B. (1964). Nonmetric multidimensional scaling: A numerical method. *Pyrometrical*, *29*, 1-27, 115-129.

Kruskal, J. B., & Wish, M. (1978). *Multidimensional scaling*. Beverly Hills, CA: Sage Publications.

Kruskal, W. H. (1952). A nonparametric test for the several sample problem. *Annals of Mathematical Statistics*, *23*, 525-540.

Kruskal, W. H. (1975). Visions of maps and graphs. In J. Kavaliunas (Ed.), *Auto-carto II, proceedings of the international symposium on computer assisted cartography*. Washington, DC: U. S. Bureau of the Census and American Congress on Survey and Mapping.

Kruskal, W. H., & Wallis, W. A. (1952). Use of ranks in one-criterion variance analysis. *Journal of the American Statistical Association*, *47*, 583-621.

Ku, H. H., & Kullback, S. (1968). Interaction in multidimensional contingency tables: An information theoretic approach. *J. Res. Nat. Bur. Standards Sect. B*, 72, 159-199.

Ku, H. H., Varner, R. N., & Kullback, S. (1971). Analysis of multidimensional contingency tables. *Journal of the American Statistical Association*, *66*, 55-64.

Kullback, S. (1959). *Information theory and statistics.* New York: Wiley.

Kvålseth, T. O. (1985). Cautionary note about R^2. *The American Statistician*, *39*, 279-285.

Lagakos, S. W., & Kuhns, M. H. (1978). Maximum likelihood estimation for censored exponential survival data with covariates. *Applied Statistics*, *27*, 190-197.

Lance, G. N., & Williams, W. T. (1966). A general theory of classificatory sorting strategies. *Computer Journal*, *9*, 373.

Lance, G. N., & Williams, W. T. (1966). Computer programs for hierarchical polythetic classification ("symmetry analysis"). *Computer Journal*, *9*, 60.

Larsen, W. A., & McCleary, S. J. (1972). The use of partial residual plots in regression analysis. *Technometrics*, *14*, 781-790.

Lawless, J. F. (1982). *Statistical models and methods for lifetime data*. New York: Wiley.

Lawley, D. N., & Maxwell, A. E. (1971). *Factor analysis as a statistical method.* New York: American Elsevier.

Lawley, D. N., & Maxwell, A. E. (1971). *Factor analysis as a statistical method* (2nd. ed.). London: Butterworth & Company.

Lee, E. T. (1980). *Statistical methods for survival data analysis.* Belmont, CA: Lifetime Learning.

Lee, E. T., & Desu, M. M. (1972). A computer program for comparing *K* samples with right-censored data. *Computer Programs in Biomedicine*, *2*, 315-321.

Lee, E. T., Desu, M. M., & Gehan, E. A. (1975). A Monte-Carlo study of the power of some two-sample tests. *Biometrika*, *62*, 425-532.

Lee, S., & Hershberger, S. (1990). A simple rule for generating equivalent models in covariance structure modeling. *Multivariate Behavioral Research*, *25*, 313-334.

Lehmann, E. L. (1975). *Nonparametrics: Statistical methods based on ranks.* San Francisco: Holden-Day.

Lilliefors, H. W. (1967). On the Kolmogorov-Smirnov test for normality with mean and variance unknown. *Journal of the American Statistical Association, 64*, 399-402.

Lindeman, R. H., Merenda, P. F., & Gold, R. (1980). *Introduction to bivariate and multivariate analysis*. New York: Scott, Foresman, & Co.

Lindman, H. R. (1974). *Analysis of variance in complex experimental designs*. San Francisco: W. H. Freeman & Co.

Linfoot, E. H. (1957). An informational measure of correlation. *Information and Control*, *1*, 50-55.

StatSoft®

Copyright © StatSoft, 1995

Linn, R. L. (1968). A Monte Carlo approach to the number of factors problem. *Psychometrika*, *33*, 37-71.

Lipson, C., & Sheth, N. C. (1973). *Statistical design and analysis of engineering experiments*. New York: McGraw-Hill.

Lloyd, D. K., & Lipow, M. (1977). *Reliability: Management, methods, and mathematics.* New York: McGraw-Hill.

Loehlin, J. C. (1987). *Latent variable models: An introduction to latent, path, and structural analysis.* Hillsdale, NJ: Erlbaum.

Long, J. S. (1983a). *Confirmatory factor analysis.* Beverly Hills: Sage.

Long, J. S. (1983b). *Covariance structure models: An introduction to LISREL.* Beverly Hills: Sage.

Longley, J. W. (1967). An appraisal of least squares programs for the electronic computer from the point of view of the user. *Journal of the American Statistical Association*, *62*, 819-831.

Longley, J. W. (1984). *Least squares computations using orthogonalization methods*. New York: Marcel Dekker.

Lord, F. M. (1957). A significance test for the hypothesis that two variables measure the same trait except for errors of measurement. *Psychometrika*, *22*, 207-220.

Lorenz, M. O. (1904). Methods of measuring the concentration of wealth. *American Statistical Association Publication*, *9*, 209-219.

Lucas, J. M. (1976). The design and use of cumulative sum quality control schemes. *Journal of Quality Technology*, *8*, 45-70.

Lucas, J. M. (1982). Combined Shewhart-CUSUM quality control schemes. *Journal of Quality Technology*, *14*, 89-93.

Maddala, G. S. (1977) *Econometrics*. New York: McGraw-Hill.

Maiti, S. S., & Mukherjee, B. N. (1990). A note on the distributional properties of the Jöreskog-Sörbom fit indices. *Psychometrika*, *55*, 721-726.

Makridakis, S. G. (1983). Empirical evidence versus personal experience. *Journal of Forecasting*, *2*, 295-306.

Makridakis, S. G. (1990). *Forecasting, planning, and strategy for the 21st century.* London: Free Press.

Makridakis, S. G., & Wheelwright, S. C. (1978). *Interactive forecasting: Univariate and multivariate methods* (2nd ed.). San Francisco, CA: Holden-Day.

Makridakis, S. G., & Wheelwright, S. C. (1989). *Forecasting methods for management* (5th ed.). New York: Wiley.

Makridakis, S. G., Wheelwright, S. C., & McGee, V. E. (1983). *Forecasting: Methods and applications* (2nd ed.). New York: Wiley.

Makridakis, S., Andersen, A., Carbone, R., Fildes, R., Hibon, M., Lewandowski, R., Newton, J., Parzen, R., & Winkler, R. (1982). The accuracy of extrapolation (time series) methods: Results of a forecasting competition. *Journal of Forecasting*, *1*, 11-153.

Malinvaud, E. (1970). *Statistical methods of econometrics*. Amsterdam: North-Holland Publishing Co.

Mandel, B. J. (1969). The regression control chart. *Journal of Quality Technology*, *1*, 3-10.

Mann, H. B., & Whitney, D. R. (1947). On a test of whether one of two random variables is stochastically larger than the other. *Annals of Mathematical Statistics*, *18*, 50-60.

Mann, N. R., Schafer, R. E., & Singpurwalla, N. D. (1974). *Methods for statistical analysis of reliability and life data*. New York: Wiley.

Mantel, N. (1966). Evaluation of survival data and two new rank order statistics arising in its consideration. *Cancer Chemotherapy Reports*, *50*, 163-170.

Copyright © StatSoft, 1995

Mantel, N. (1967). Ranking procedures for arbitrarily restricted observations. *Biometrics*, *23*, 65-78.

Mantel, N. (1974). Comment and suggestion on the Yates continuity correction. *Journal of the American Statistical Association*, *69*, 378-380.

Mantel, N., & Haenszel, W. (1959). Statistical aspects of the analysis of data from retrospective studies of disease. *Journal of the National Cancer Institute*, *22*, 719-748.

Marascuilo, L. A., & McSweeney, M. (1977). *Nonparametric and distribution free methods for the social sciences.* Monterey, CA: Brooks/Cole.

Marple, S. L., Jr. (1987). *Digital spectral analysis.* Englewood Cliffs, NJ: Prentice-Hall.

Mardia, K. V., Kent, J. T., and Bibby, J. M. (1979). *Multivariate analysis*. New York: Academic Press.

Marsaglia, G. (1962). Random variables and computers. In J. Kozenik (Ed.), *Information theory, statistical decision functions, random processes: Transactions of the Third Prague Conference*. Prague: Czechoslovak Academy of Sciences.

Mason, R. L., Gunst, R. F., & Hess, J. L. (1989). *Statistical design and analysis of experiments with applications to engineering and science.* New York: Wiley.

Massey, F. J., Jr. (1951). The Kolmogorov-Smirnov test for goodness of fit. *Journal of the American Statistical Association*, *46*, 68-78.

Matsueda, R. L., & Bielby, W. T. (1986). Statistical power in covariance structure models. In N. B. Tuma (Ed.), *Sociological methodology.* Washington, DC: American Sociological Association.

McArdle, J. J. (1978). A structural view of structural models. Paper presented at the *Winter Workshop on Latent Structure Models Applied to Developmental Data, University of Denver, December, 1978.*

McArdle, J. J., & McDonald, R. P. (1984). Some algebraic properties of the Reticular Action Model for moment structures. *British Journal of Mathematical and Statistical Psychology*, *37*, 234-251.

McCleary, R., & Hay, R. A. (1980). *Applied time series analysis for the social sciences*. Beverly Hills, CA: Sage Publications.

McDonald, R. P. (1980). A simple comprehensive model for the analysis of covariance structures. *British Journal of Mathematical and Statistical Psychology*, *31*, 59-72.

McDonald, R. P. (1989). An index of goodness-of-fit based on noncentrality. *Journal of Classification*, *6*, 97-103.

McDonald, R. P., & Hartmann, W. M. (1992). A procedure for obtaining initial value estimates in the RAM model. *Multivariate Behavioral Research*, *27*, 57-76.

McDonald, R. P., & Mulaik, S. A. (1979). Determinacy of common factors: A nontechnical review. *Psychological Bulletin*, *86*, 297-306.

McDowall, D., McCleary, R., Meidinger, E. E., & Hay, R. A. (1980). *Interrupted time series analysis*. Beverly Hills, CA: Sage Publications.

McKenzie, E. (1984). General exponential smoothing and the equivalent ARMA process. *Journal of Forecasting*, *3*, 333-344.

McKenzie, E. (1985). Comments on 'Exponential smoothing: The state of the art' by E. S. Gardner, Jr. *Journal of Forecasting*, *4*, 32-36.

McLain, D. H. (1974). Drawing contours from arbitrary data points. *The Computer Journal*, *17*, 318-324.

McLean, R. A., & Anderson, V. L. (1984). *Applied factorial and fractional designs.* New York: Marcel Dekker.

StatSoft®

Copyright © StatSoft, 1995

McLeod, A. I., & Sales, P. R. H. (1983). An algorithm for approximate likelihood calculation of ARMA and seasonal ARMA models. *Applied Statistics*, 211-223 (Algorithm AS).

McNemar, Q. (1947). Note on the sampling error of the difference between correlated proportions or percentages. *Psychometrika*, *12*, 153-157.

McNemar, Q. (1969). *Psychological statistics* (4th ed.). New York: Wiley.

Mels, G. (1989). *A general system for path analysis with latent variables*. M. S. Thesis: Department of Statistics, University of South Africa.

Mendoza, J. L., Markos, V. H., & Gonter, R. (1978). A new perspective on sequential testing procedures in canonical analysis: A Monte Carlo evaluation. *Multivariate Behavioral Research*, *13*, 371-382.

Meredith, W. (1964). Canonical correlation with fallible data. *Psychometrika*, *29*, 55-65.

Miller, R. (1981). *Survival analysis*. New York: Wiley.

Milligan, G. W. (1980). An examination of the effect of six types of error perturbation on fifteen clustering algorithms. *Psychometrika*, *45*, 325-342.

Milliken, G. A., & Johnson, D. E. (1984). *Analysis of messy data: Vol. I. Designed experiments.* New York: Van Nostrand Reinhold, Co.

Mitchell, T. J. (1974a). Computer construction of "D-optimal" first-order designs. *Technometrics*, *16*, 211-220.

Mitchell, T. J. (1974b). An algorithm for the construction of "D-optimal" experimental designs. *Technometrics*, *16*, 203-210.

Mittag, H. J. (1993). *Qualitätsregelkarten.* München/Wien: Hanser Verlag.

Mittag, H. J., & Rinne, H. (1993). *Statistical methods of quality assurance.* London/New York: Chapman & Hall.

Monro, D. M. (1975). Complex discrete fast Fourier transform. *Applied Statistics*, *24*, 153-160.

Monro, D. M., & Branch, J. L. (1976). The chirp discrete Fourier transform of general length. *Applied Statistics*, *26*, 351-361.

Montgomery, D. C. (1976). *Design and analysis of experiments*. New York: Wiley.

Montgomery, D. C. (1985). *Statistical quality control*. New York: Wiley.

Montgomery, D. C. (1991) *Design and analysis of experiments* (3rd ed.). New York: Wiley.

Montgomery, D. C., & Wadsworth, H. M. (1972). Some techniques for multivariate quality control applications. *Technical Conference Transactions*. Washington, DC: American Society for Quality Control.

Montgomery, D. C., Johnson, L. A., & Gardiner, J. S. (1990). *Forecasting and time series analysis* (2nd ed.). New York: McGraw-Hill.

Mood, A. M. (1954). *Introduction to the theory of statistics.* New York: McGraw Hill.

Morris, M., & Thisted, R. A. (1986). Sources of error in graphical perception: A critique and an experiment. *Proceedings of the Section on Statistical Graphics, American Statistical Association*, 43-48.

Morrison, A. S., Black, M. M., Lowe, C. R., MacMahon, B., & Yuasa, S. (1973). Some international differences in histology and survival in breast cancer. *International Journal of Cancer*, *11*, 261-267.

Morrison, D. (1967). *Multivariate statistical methods.* New York: McGraw-Hill.

Moses, L. E. (1952). Non-parametric statistics for psychological research. *Psychological Bulletin*, *49*, 122-143.

Mulaik, S. A. (1972). *The foundations of factor analysis.* New York: McGraw Hill.

Copyright © StatSoft, 1995

Muth, J. F. (1960). Optimal properties of exponentially weighted forecasts. *Journal of the American Statistical Association*, *55*, 299-306.

Nachtsheim, C. J. (1979). *Contributions to optimal experimental design.* Ph.D. thesis, Department of Applied Statistics, University of Minnesota.

Nachtsheim, C. J. (1987). Tools for computer-aided design of experiments. *Journal of Quality Technology*, *19*, 132-160.

Nelder, J. A., & Mead, R. (1965). A Simplex method for function minimization. *Computer Journal*, *7*, 308-313.

Nelson, L. (1984). The Shewhart control chart - tests for special causes. *Journal of Quality Technology, 15*, 237-239.

Nelson, L. (1985). Interpreting Shewhart X-bar control charts. *Journal of Quality Technology, 17*, 114-116.

Nelson, W. (1982). *Applied life data analysis*. New York: Wiley.

Neter, J., Wasserman, W., & Kutner, M. H. (1985). *Applied linear statistical models: Regression, analysis of variance, and experimental designs.* Homewood, IL: Irwin.

Neter, J., Wasserman, W., & Kutner, M. H. (1989). *Applied linear regression models* (2nd ed.). Homewood, IL: Irwin.

Nisbett, R. E., Fong, G. F., Lehman, D. R., & Cheng, P. W. (1987). Teaching reasoning. *Science*, *238*, 625-631.

Noori, H. (1989). The Taguchi methods: Achieving design and output quality. *The Academy of Management Executive*, *3*, 322-326.

Nunally, J. C. (1970). *Introduction to psychological measurement*. New York: McGraw-Hill.

Nunnally, J. C. (1978). *Psychometric theory*. New York: McGraw-Hill.

Nussbaumer, H. J. (1982). *Fast Fourier transforms and convolution algorithms* (2nd ed.). New York: Springer-Verlag.

O'Brien, R. G., & Kaiser, M. K. (1985). MANOVA method for analyzing repeated measures designs: An extensive primer. *Psychological Bulletin*, *97*, 316-333.

O'Neill, R. (1971). Function minimization using a Simplex procedure. *Applied Statistics*, *3*, 79-88.

Okunade, A. A., Chang, C. F., & Evans, R. D. (1993). Comparative analysis of regression output summary statistics in common statistical packages. *The American Statistician*, *47*, 298-303.

Olds, E. G. (1949). The 5% significance levels for sums of squares of rank differences and a correction. *Annals of Mathematical Statistics*, *20*, 117-118.

Olson, C. L. (1976). On choosing a test statistic in multivariate analysis of variance. *Psychological Bulletin*, *83*, 579-586.

Ostle, B., & Malone, L. C. (1988). *Statistics in research: Basic concepts and techniques for research workers* (4th ed.). Ames, IA: Iowa State Press.

Ostrom, C. W. (1978). *Time series analysis: Regression techniques*. Beverly Hills, CA: Sage Publications.

Overall, J. E., & Speigel, D. K. (1969). Concerning least squares analysis of experimental data. *Psychological Bulletin*, *83*, 579-586.

Page, E. S. (1954). Continuous inspection schemes. *Biometrics*, *41*, 100-114.

Page, E. S. (1961). Cumulative sum charts. *Technometrics*, *3*, 1-9.

Palumbo, F. A., & Strugala, E. S. (1945). Fraction defective of battery adapter used in handie-talkie. *Industrial Quality Control*, *November*, 6-8.

Pankratz, A. (1983). *Forecasting with univariate Box-Jenkins models: Concepts and cases*. New York: Wiley.

StatSoft®

Copyright © StatSoft, 1995

Parzen, E. (1961). Mathematical considerations in the estimation of spectra: Comments on the discussion of Messers, Tukey, and Goodman. *Technometrics*, *3*, 167-190; 232-234.

Patil, K. D. (1975). Cochran's Q test: Exact distribution. *Journal of the American Statistical Association*, *70*, 186-189.

Peace, G. S. (1993). *Taguchi methods: A hands-on approach*. Milwaukee, WI: ASQC.

Pearson, K., (Ed.). (1968). *Tables of incomplete beta functions* (2nd ed.). Cambridge, MA: Cambridge University Press.

Pedhazur, E. J. (1973). *Multiple regression in behavioral research*. New York: Holt, Rinehart, & Winston.

Pedhazur, E. J. (1982). *Multiple regression in behavioral research* (2nd ed.). New York: Holt, Rinehart, & Winston.

Peressini, A. L., Sullivan, F. E., & Uhl, J. J., Jr. (1988). *The mathematics of nonlinear programming.* New York: Springer.

Peto, R., & Peto, J. (1972). Asymptotically efficient rank invariant procedures. *Journal of the Royal Statistical Society*, *135*, 185-207.

Phadke, M. S. (1989). *Quality engineering using robust design.* Englewood Cliffs, NJ: Prentice-Hall.

Piepel, G. F. (1988). Programs for generating extreme vertices and centroids of linearly constrained experimental regions. *Journal of Quality Technology, 20*, 125-139.

Pike, M. C. (1966). A method of analysis of certain class of experiments in carcinogenesis. *Biometrics, 22*, 142-161.

Pillai, K. C. S. (1965). On the distribution of the largest characteristic root of a matrix in multivariate analysis. *Biometrika*, *52*, 405-414.

Plackett, R. L., & Burman, J. P. (1946). The design of optimum multifactorial experiments. *Biometrika*, *34*, 255-272.

Porebski, O. R. (1966). Discriminatory and canonical analysis of technical college data. *British Journal of Mathematical and Statistical Psychology*, *19*, 215-236.

Powell, M. J. D. (1964). An efficient method for finding the minimum of a function of several variables without calculating derivatives. *Computer Journal*, *7*, 155-162.

Prentice, R. (1973). Exponential survivals with censoring and explanatory variables. *Biometrika*, *60*, 279-288.

Press, William, H., Flannery, B. P., Teukolsky, S. A., Vetterling, W. T. (1986). *Numerical recipies*. New York: Cambridge University Press.

Priestley, M. B. (1981). *Spectral analysis and time series.* New York: Academic Press.

Pyzdek, T. (1989). *What every engineer should know about quality control.* New York: Marcel Dekker.

Ralston, A., & Wilf, H.S. (Eds.). (1960). *Mathematical methods for digital computers.* New York: Wiley.

Ralston, A., & Wilf, H.S. (Eds.). (1967). *Mathematical methods for digital computers* (Vol. II). New York: Wiley.

Randles, R. H., & Wolfe, D. A. (1979). *Introduction to the theory of nonparametric statistics.* New York: Wiley.

Rao, C. R. (1951). An asymptotic expansion of the distribution of Wilks' criterion. *Bulletin of the International Statistical Institute*, *33*, 177-181.

Rao, C. R. (1965). *Linear statistical inference and its applications.* New York: Wiley.

Rhoades, H. M., & Overall, J. E. (1982). A sample size correction for Pearson chi-square in 2 x 2 contingency tables. *Psychological Bulletin*, *91*, 418-423.

Rinne, H., & Mittag, H. J. (1995). *Statistische Methoden der Qualitätssicherung (3rd. edition).* München/Wien: Hanser Verlag.

Copyright © StatSoft, 1995

Ripley, B. D. (1981). *Spacial statistics.* New York: Wiley.

Rogan, J. C., Keselman, J. J., & Mendoza, J. L. (1979). Analysis of repeated measurements. *British Journal of Mathematical and Statistical Psychology*, *32*, 269-286.

Rosenberg, S. (1977). New approaches to the analysis of personal constructs in person perception. In A. Landfield (Ed.), *Nebraska symposium on motivation* (Vol. 24). Lincoln, NE: University of Nebraska Press.

Rosenberg, S., & Sedlak, A. (1972). Structural representations of implicit personality theory. In L. Berkowitz (Ed.). *Advances in experimental social psychology* (Vol. 6). New York: Academic Press.

Roskam, E. E., & Lingoes, J. C. (1970). *MINISSA-I*: A Fortran IV program for the smallest space analysis of square symmetric matrices. *Behavioral Science*, *15*, 204-205.

Ross, P. J. (1988). *Taguchi techniques for quality engineering: Loss function, orthogonal experiments, parameter, and tolerance design*. Milwaukee, WI: ASQC.

Roy, J. (1958). Step-down procedure in multivariate analysis. *Annals of Mathematical Statistics*, *29*, 1177-1187.

Roy, J. (1967). *Some aspects of multivariate analysis.* New York: Wiley.

Roy, R. (1990). *A primer on the Taguchi method*. Milwaukee, WI: ASQC.

Royston, J. P. (1982). An extension of Shapiro and Wilk's W test for normality to large samples. *Applied Statistics, 31*, 115-124.

Rozeboom, W. W. (1979). Ridge regression: Bonanza or beguilement? *Psychological Bulletin*, *86*, 242-249.

Rozeboom, W. W. (1988). Factor indeterminacy: the saga continues. *British Journal of Mathematical and Statistical Psychology*, *41*, 209-226.

Runyon, R. P., & Haber, A. (1976). *Fundamentals of behavioral statistics*. Reading, MA: Addison-Wesley.

Ryan, T. P. (1989). *Statistical methods for quality improvement.* New York: Wiley.

Sandler, G. H. (1963). *System reliability engineering*. Englewood Cliffs, NJ: Prentice-Hall.

SAS Institute, Inc. (1982). *SAS user's guide: Statistics, 1982 Edition*. Cary, NC: SAS Institute, Inc.

Satorra, A., & Saris, W. E. (1985). Power of the likelihood ratio test in covariance structure analysis. *Psychometrika*, *50*, 83-90.

Saxena, K. M. L., & Alam, K. (1982). Estimation of the noncentrality parameter of a chi squared distribution. *Annals of Statistics*, *10*, 1012-1016.

Scheffé, H. (1953). A method for judging all possible contrasts in the analysis of variance. *Biometrica*, *40*, 87-104.

Scheffé, H. (1959). *The analysis of variance*. New York: Wiley.

Scheffé, H. (1963). The simplex-centroid design for experiments with mixtures. *Journal of the Royal Statistical Society*, *B25*, 235-263.

Scheffé, H., & Tukey, J. W. (1944). A formula for sample sizes for population tolerance limits. *Annals of Mathematical Statistics*, *15*, 217.

Scheines, R. (1994). Causation, indistinguishability, and regression. In F. Faulbaum, (Ed.), *SoftStat '93. Advances in statistical software 4*. Stuttgart: Gustav Fischer Verlag.

Schiffman, S. S., Reynolds, M. L., & Young, F. W. (1981). *Introduction to multidimensional scaling: Theory, methods, and applications*. New York: Academic Press.

StatSoft®

Copyright © StatSoft, 1995

Schmidt, P., & Muller, E. N. (1978). The problem of multicollinearity in a multistage causal alienation model: A comparison of ordinary least squares, maximum-likelihood and ridge estimators. *Quality and Quantity*, *12*, 267-297.

Schmidt, P., & Sickles, R. (1975). On the efficiency of the Almon lag technique. *International Economic Review*, *16*, 792-795.

Schmidt, P., & Waud, R. N. (1973). The Almon lag technique and the monetary versus fiscal policy debate. *Journal of the American Statistical Association*, *68*, 11-19.

Schnabel, R. B., Koontz, J. E., and Weiss, B. E. (1985). A modular system of algorithms for unconstrained minimization. *ACM Transactions on Mathematical Software*, *11*, 419-440.

Schneider, H. (1986). *Truncated and censored samples from normal distributions*. New York: Marcel Dekker.

Schönemann, P. H., & Steiger, J. H. (1976). Regression component analysis. *British Journal of Mathematical and Statistical Psychology*, *29*, 175-189.

Schrock, E. M. (1957). *Quality control and statistical methods*. New York: Reinhold Publishing.

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, *6*, 461-464.

Scott, D. W. (1979). On optimal and data-based histograms. *Biometrika*, *66*, 605-610.

Searle, S. R. (1987). *Linear models for unbalanced data*. New York: Wiley.

Searle, S. R., Casella, G., & McCullock, C. E. (1992). *Variance components*. New York: Wiley.

Seber, G. A. F., & Wild, C. J. (1989). *Nonlinear regression.* New York: Wiley.

Sebestyen, G. S. (1962). *Decision making processes in pattern recognition.* New York: Macmillan.

Sen, P. K., & Puri, M. L. (1968). On a class of multivariate multisample rank order tests, II: Test for homogeneity of dispersion matrices. *Sankhya*, *30*, 1-22.

Shapiro, A., & Browne, M. W. (1983). On the investigation of local identifiability: A counter example. *Psychometrika*, *48*, 303-304.

Shapiro, S. S., Wilk, M. B., & Chen, H. J. (1968). A comparative study of various tests of normality. *Journal of the American Statistical Association, 63*, 1343-1372.

Shewhart, W. A. (1931). *Economic control of quality of manufactured product*. New York: D. Van Nostrand.

Shewhart, W. A. (1939). *Statistical method from the viewpoint of quality*. Washington, DC: The Graduate School Department of Agriculture.

Shirland, L. E. (1993). *Statistical quality control with microcomputer applications.* New York: Wiley.

Shiskin, J., Young, A. H., & Musgrave, J. C. (1967). *The X-11 variant of the census method II seasonal adjustment program.* Technical paper no. 15. Washington, DC: Bureau of the Census.

Shumway, R. H. (1988). *Applied statistical time series analysis.* Englewood Cliffs, NJ: Prentice Hall.

Siegel, A. E. (1956). Film-mediated fantasy aggression and strength of aggressive drive. *Child Development*, *27*, 365-378.

Siegel, S. (1956). *Nonparametric statistics for the behavioral sciences.* New York: McGraw-Hill.

Siegel, S., & Castellan, N. J. (1988). *Nonparametric statistics for the behavioral sciences* (2nd ed.) New York: McGraw-Hill.

Simkin, D., & Hastie, R. (1986). Towards an information processing view of graph perception. *Proceedings of the Section on Statistical Graphics, American Statistical Association*, 11-20.

Copyright © StatSoft, 1995

Sinha, S. K., & Kale, B. K. (1980). *Life testing and reliability estimation*. New York: Halstead.

Smirnov, N. V. (1948). Table for estimating the goodness of fit of empirical distributions. *Annals of Mathematical Statistics*, *19*, 279-281.

Smith, D. J. (1972). *Reliability engineering*. New York: Barnes & Noble.

Smith, K. (1953). Distribution-free statistical methods and the concept of power efficiency. In L. Festinger and D. Katz (Eds.), *Research methods in the behavioral sciences* (pp. 536-577). New York: Dryden.

Sneath, P. H. A., & Sokal, R. R. (1973). *Numerical taxonomy.* San Francisco: W. H. Freeman & Co.

Snee, R. D. (1975). Experimental designs for quadratic models in constrained mixture spaces. *Technometrics*, *17*, 149-159.

Snee, R. D. (1979). Experimental designs for mixture systems with multi-component constraints. *Communications in Statistics - Theory and Methods*, *A8(4)*, 303-326.

Snee, R. D. (1985). Computer-aided design of experiments - some practical experiences. *Journal of Quality Technology*, *17*, 222-236.

Snee, R. D. (1986). An alternative approach to fitting models when re-expression of the response is useful. *Journal of Quality Technology, 18*, 211-225.

Sokal, R. R., & Mitchener, C. D. (1958). A statistical method for evaluating systematic relationships. *University of Kansas Science Bulletin*, *38*, 1409.

Sokal, R. R., & Sneath, P. H. A. (1963). *Principles of numerical taxonomy*. San Francisco: W. H. Freeman & Co.

Spirtes, P., Glymour, C., & Scheines, R. (1993). *Causation, prediction, and search.* Lecture Notes in Statistics, V. 81. New York: Springer-Verlag.

Spjotvoll, E., & Stoline, M. R. (1973). An extension of the *T*-method of multiple comparison to include the cases with unequal sample sizes. *Journal of the American Statistical Association*, *68*, 976-978.

Springer, M. D. (1979). *The algebra of random variables*. New York: Wiley.

Spruill, M. C. (1986). Computation of the maximum likelihood estimate of a noncentrality parameter. *Journal of Multivariate Analysis*, *18*, 216-224.

Steiger, J. H. (1979). Factor indeterminacy in the 1930's and in the 1970's; some interesting parallels. *Psychometrika*, *44*, 157-167.

Steiger, J. H. (1980a). Tests for comparing elements of a correlation matrix. *Psychological Bulletin*, *87*, 245-251.

Steiger, J. H. (1980b). Testing pattern hypotheses on correlation matrices: Alternative statistics and some empirical results. *Multivariate Behavioral Research*, *15*, 335-352.

Steiger, J. H. (1988). Aspects of person-machine communication in structural modeling of correlations and covariances. *Multivariate Behavioral Research*, *23*, 281-290.

Steiger, J. H. (1989). *EzPATH: A supplementary module for SYSTAT and SYGRAPH.* Evanston, IL: SYSTAT, Inc.

Steiger, J. H. (1990). Some additional thoughts on components and factors. *Multivariate Behavioral Research*, *25*, 41-45.

Steiger, J. H., & Browne, M. W. (1984). The comparison of interdependent correlations between optimal linear composites. *Psychometrika*, *49*, 11-24.

Steiger, J. H., & Hakstian, A. R. (1982). The asymptotic distribution of elements of a correlation matrix: Theory and application. *British Journal of Mathematical and Statistical Psychology*, *35*, 208-215.

StatSoft®

Copyright © StatSoft, 1995

Steiger, J. H., & Lind, J. C. (1980). Statistically-based tests for the number of common factors. Paper presented at the annual Spring Meeting of the Psychometric Society in Iowa City, IA. May 30, 1980.

Steiger, J. H., & Schönemann, P. H. (1978). A history of factor indeterminacy. In S. Shye, (Ed.), *Theory Construction and Data Analysis in the Social Sciences.* San Francisco: Jossey-Bass.

Steiger, J. H., Shapiro, A., & Browne, M. W. (1985). On the multivariate asymptotic distribution of sequential chi-square statistics. *Psychometrika*, *50*, 253-264.

Stelzl, I. (1986). Changing causal relationships without changing the fit: Some rules for generating equivalent LISREL models. *Multivariate Behavioral Research*, *21*, 309-331.

Stenger, F. (1973). Integration formula based on the trapezoid formula. *Journal of the Institute of Mathematics and Applications*, *12*, 103-114.

Stevens, J. (1986). *Applied multivariate statistics for the social sciences.* Hillsdale, NJ: Erlbaum.

Stevens, W. L. (1939). Distribution of groups in a sequence of alternatives. *Annals of Eugenics*, *9*, 10-17.

Stewart, D. K., & Love, W. A. (1968). A general canonical correlation index. *Psychological Bulletin*, *70*, 160-163.

Steyer, R. (1992). *Theorie causale regressionsmodelle* [Theory of causal regression models]. Stuttgart: Gustav Fischer Verlag.

Steyer, R. (1994). Principles of causal modeling: a summary of its mathematical foundations and practical steps. In F. Faulbaum, (Ed.), *SoftStat '93. Advances in statistical software 4.* Stuttgart: Gustav Fischer Verlag.

Taguchi, G. (1987). *Jikken keikakuho* (3rd ed., Vol I & II). Tokyo: Maruzen. English translation edited by D. Clausing. *System of experimental design*. New York: UNIPUB/Kraus International

Tanaka, J. S., & Huba, G. J. (1985). A fit index for covariance structure models under arbitrary GLS estimation. *British Journal of Mathematical and Statistical Psychology*, *38*, 197-201.

Tanaka, J. S., & Huba, G. J. (1989). A general coefficient of determination for covariance structure models under arbitrary GLS estimation. *British Journal of Mathematical and Statistical Psychology*, *42*, 233-239.

Tatsuoka, M. M. (1970). *Discriminant analysis*. Champaign, IL: Institute for Personality and Ability Testing.

Tatsuoka, M. M. (1971). *Multivariate analysis*. New York: Wiley.

Tatsuoka, M. M. (1976). Discriminant analysis. In P. M. Bentler, D. J. Lettieri, and G. A. Austin (Eds.), *Data analysis strategies and designs for substance abuse research.* Washington, DC: U.S. Government Printing Office.

Thorndyke, R. L., & Hagen, E. P. (1977). *Measurement and evaluation in psychology and education*. New York: Wiley.

Thurstone, L. L. (1947). *Multiple factor analysis.* Chicago: University of Chicago Press.

Timm, N. H. (1975). *Multivariate analysis with applications in education and psychology.* Monterey, CA: Brooks/Cole.

Timm, N. H., & Carlson, J. (1973). *Multivariate analysis of non-orthogonal experimental designs using a multivariate full rank model*. Paper presented at the American Statistical Association Meeting, New York.

Timm, N. H., & Carlson, J. (1975). Analysis of variance through full rank models. *Multivariate behavioral research monographs*, No. *75-1*.

Tribus, M., & Sconyi, G. (1989). An alternative view of the Taguchi approach. *Quality Progress*, *22*, 46-48.

Copyright © StatSoft, 1995

Trivedi, P. K., & Pagan, A. R. (1979). Polynomial distributed lags: A unified treatment. *Economic Studies Quarterly*, *30*, 37-49.

Tucker, L. R., Koopman, R. F., & Linn, R. L. (1969). Evaluation of factor analytic research procedures by means of simulated correlation matrices. *Psychometrika*, *34*, 421-459.

Tufte, E. R. (1983). *The visual display of quantitative information*. Cheshire, CT: Graphics Press.

Tufte, E. R. (1990). *Envisioning information*. Cheshire, CT: Graphics Press.

Tukey, J. W. (1953). *The problem of multiple comparisons*. Unpublished manuscript, Princeton University.

Tukey, J. W. (1967). An introduction to the calculations of numerical spectrum analysis. In B. Harris (Ed.), *Spectral analysis of time series*. New York: Wiley.

Tukey, J. W. (1977). *Exploratory data analysis*. Reading, MA: Addison-Wesley.

Tukey, J. W. (1984). *The collected works of John W. Tukey.* Monterey, CA: Wadsworth.

Tukey, P. A. (1986). A data analyst's view of statistical plots. *Proceedings of the Section on Statistical Graphics, American Statistical Association*, 21-28.

Tukey, P. A., & Tukey, J. W. (1981). Graphical display of data sets in 3 or more dimensions. In V. Barnett (Ed.), *Interpreting multivariate data*. Chichester, U.K.: Wiley.

Vale, C. D., & Maurelli, V. A. (1983). Simulating multivariate nonnormal distributions. *Psychometrika*, *48*, 465-471.

Vandaele, W. (1983). *Applied time series and Box-Jenkins models*. New York: Academic Press.

Vaughn, R. C. (1974). *Quality control*. Ames, IA: Iowa State Press.

Velicer, W. F., & Jackson, D. N. (1990). Component analysis vs. factor analysis: some issues in selecting an appropriate procedure. *Multivariate Behavioral Research*, *25*, 1-28.

Velleman, P. F., & Hoaglin, D. C. (1981). *Applications, basics, and computing of exploratory data analysis*. Belmont, CA: Duxbury Press.

Wainer, H. (1995). Visual revelations. *Chance*, *8*, 48-54.

Wald, A. (1947). *Sequential analysis*. New York: Wiley.

Walker, J. S. (1991). *Fast Fourier transforms.* Boca Raton, FL: CRC Press.

Wallis, K. F. (1974). Seasonal adjustment and relations between variables. *Journal of the American Statistical Association*, *69*, 18-31.

Wang, C. M., & Gugel, H. W. (1986). High-performance graphics for exploring multivariate data. *Proceedings of the Section on Statistical Graphics, American Statistical Association*, 60-65.

Ward, J. H. (1963). Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, *58*, 236.

Wei, W. W. (1989). *Time series analysis: Univariate and multivariate methods.* New York: Addison-Wesley.

Welstead, S. T. (1994). *Neural network and fuzzy logic applications in C/C++*. New York: Wiley.

Wescott, M. E. (1947). Attribute charts in quality control. *Conference Papers, First Annual Convention of the American Society for Quality Control.* Chicago: John S. Swift Co.

Wheaton, B., Múthen, B., Alwin, D., & Summers G. (1977). Assessing reliability and stability in panel models. In D. R. Heise (Ed.), *Sociological Methodology.* New York: Wiley.

StatSoft®

Copyright © StatSoft, 1995

Wheeler, D. J., & Chambers, D.S. (1986). *Understanding statistical process control*. Knoxville, TN: Statistical Process Controls, Inc.

Wherry, R. J. (1984). *Contributions to correlational analysis*. New York: Academic Press.

Whitney, D. R. (1948). *A comparison of the power of non-parametric tests and tests based on the normal distribution under non-normal alternatives.* Unpublished doctoral dissertation, Ohio State University.

Whitney, D. R. (1951). A bivariate extension of the *U* statistic. *Annals of Mathematical Statistics*, *22*, 274-282.

Wiggins, J. S., Steiger, J. H., and Gaelick, L. (1981). Evaluating circumplexity in models of personality. *Multivariate Behavioral Research*, *16*, 263-289.

Wilcoxon, F. (1945). Individual comparisons by ranking methods. *Biometrica Bulletin*, *1*, 80-83.

Wilcoxon, F. (1947). Probability tables for individual comparisons by ranking methods. *Biometrics*, *3*, 119-122.

Wilcoxon, F. (1949). *Some rapid approximate statistical procedures.* Stamford, CT: American Cyanamid Co.

Wilde, D. J., & Beightler, C. S. (1967). *Foundations of optimization.* Englewood Cliffs, NJ: Prentice-Hall.

Wilks, S. S. (1946). *Mathematical statistics.* Princeton, NJ: Princeton University Press.

Williams, W. T., Lance, G. N., Dale, M. B., & Clifford, H. T. (1971). Controversy concerning the criteria for taxonometric strategies. *Computer Journal*, *14*, 162.

Wilson, G. A., & Martin, S. A. (1983). An empirical comparison of two methods of testing the significance of a correlation matrix. *Educational and Psychological Measurement*, *43*, 11-14.

Winer, B. J. (1962). *Statistical principles in experimental design*. New York: McGraw-Hill.

Winer, B. J. (1971). *Statistical principles in experimental design* (2nd ed.). New York: McGraw-Hill.

Wolynetz, M. S. (1979a). Maximum likelihood estimation from confined and censored normal data. *Applied Statistics*, *28*, 185-195.

Wolynetz, M. S. (1979b). Maximum likelihood estimation in a linear model from confined and censored normal data. *Applied Statistics*, *28*, 195-206.

Wonnacott, R. J., & Wonnacot, T. H. (1970). *Econometrics*. New York: Wiley.

Woodward, J. A., Bonett, D. G., & Brecht, M. L. (1990). *Introduction to linear models and experimental design*. New York: Harcourt, Brace, Jovanovich.

Woodward, J. A., & Overall, J. E. (1975). Multivariate analysis of variance by multiple regression methods. *Psychological Bulletin*, *82*, 21-32.

Woodward, J. A., & Overall, J. E. (1976). Calculation of power of the *F* test. *Educational and Psychological Measurement*, *36*, 165-168.

Woodward, J. A., Douglas, G. B., & Brecht, M. L. (1990). *Introduction to linear models and experimental design*. New York: Academic Press.

Yokoyama, Y., & Taguchi, G. (1975). *Business data analysis: Experimental regression analysis.* Tokyo: Maruzen.

Youden, W. J., & Zimmerman, P. W. (1936). Field trials with fiber pots. *Contributions from Boyce Thompson Institute, 8*, 317-331.

Young, F. W, & Hamer, R. M. (1987). *Multidimensional scaling: History, theory, and applications.* Hillsdale, NJ: Erlbaum

Copyright © StatSoft, 1995

Young, F. W., Kent, D. P., & Kuhfeld, W. F. (1986). Visuals: Software for dynamic hyper-dimensional graphics. *Proceedings of the Section on Statistical Graphics, American Statistical Association*, 69-74.

Younger, M. S. (1985). *A first course in linear regression* (2nd ed.). Boston: Duxbury Press.

Yuen, C. K., & Fraser, D. (1979). *Digital spectral analysis*. Melbourne: CSIRO/Pitman.

Zippin, C., & Armitage, P. (1966). Use of concomitant variables and incomplete survival information in the estimation of an exponential survival parameter. *Biometrics*, *22*, 665-672.

Zupan, J. (1982). *Clustering of large data sets*. New York: Research Studies Press.

Zwick, W. R., & Velicer, W. F. (1986). Comparison of five rules for determining the number of components to retain. *Psychological Bulletin*, *99*, 432-442.

StatSoft®

Copyright © StatSoft, 1995