

STATISTICA: **ОБЗОР МЕТОДОВ АНАЛИЗА И** **РУКОВОДСТВО ПОЛЬЗОВАТЕЛЯ**

Оглавление

ОБЗОР МЕТОДОВ <i>STATISTICA</i>	1
Модуль Основные статистики и таблицы	8
Модуль Множественная регрессия	17
Модуль Общий Дисперсионный анализ (ДА).....	23
Модуль Непараметрическая статистика.....	31
Модуль Подгонка распределений	37
ПОШАГОВЫЕ ПРИМЕРЫ.....	43
Пример 1: Корреляции (простой пример)	45
Пример 2: Дисперсионный анализ (ДА) (более сложный пример)	68
ИНТЕРФЕЙС ПОЛЬЗОВАТЕЛЯ.....	77
Основные характеристики	79
Поддержка многозадачного режима	80
1. Интерактивный пользовательский интерфейс	83
2. Язык SVB и работа с пакетом <i>STATISTICA</i> из других приложений	94
3. Интерфейс пользователя, ориентированный на Web	96
ТРИ КАНАЛА ВЫВОДА РЕЗУЛЬТАТОВ АНАЛИЗА	99
1. Рабочие книги.....	102
2. Отчеты.....	104
3. Автономные окна	106
ДОКУМЕНТЫ СИСТЕМЫ <i>STATISTICA</i>.....	107
Рабочие книги.....	109
Электронные таблицы (мультимедийные таблицы)	113
Отчеты.....	119
Графики.....	122
Макросы (программы на языке <i>SVB</i>)	123

ГРАФИКИ	125
Обзор	127
Настройка параметров графиков	128
Создание графиков с помощью <i>SVB</i>	133
Благодарности, ссылки, стандарты	135
Основные типы графиков.....	136
Графики исходных данных	137
Графики блоковых данных	140
Графики из меню Графика	142
Другие специализированные графики	144
НАСТРОЙКА СИСТЕМЫ STATISTICA.....	145
Настройка интерактивного пользовательского интерфейса.....	148
Настройка документов	149
Временные и постоянные настройки	150
Общие стандартные настройки	151
Настройка параметров графиков.....	152
Поддержка различных конфигураций <i>STATISTICA</i>	153
Настройка конфигураций для нескольких пользователей.....	153
STATISTICA VISUAL BASIC	155
Запись макросов	158
Среда разработки <i>SVB</i>	160
Выполнение макросов <i>SVB</i>	162
Объекты и документы ActiveX (технические замечания)	163
Поддержка OLE.....	164
Поддержка DDE	164
STATISTICA QUERY – ЗАПРОС STATISTICA.....	165
Обзор	167
Краткие пошаговые инструкции	168
Обработка данных с удаленных серверов “на месте”	169
ОБЗОР ФУНКЦИОНАЛЬНЫХ ВОЗМОЖНОСТЕЙ СИСТЕМЫ STATISTICA.....	171
ПОЛУЧЕНИЕ ПОДРОБНОЙ СПРАВКИ	179
НОВЫЕ ВОЗМОЖНОСТИ STATISTICA 6	185
НОВЕЙШИЕ МЕТОДЫ АНАЛИЗА ДАННЫХ	189
Методология “Шесть Сигма” и <i>STATISTICA</i>	191
<i>STATISTICA</i> Data Miner	197

СЕМЕЙСТВО ПРОДУКТОВ *STATISTICA*205



1

ГЛАВА

ОБЗОР МЕТОДОВ *STATISTICA*

The screenshot displays the STATISTICA software interface. The menu bar includes: **Файл**, **Правка**, **Вид**, **Вставка**, **Формат**, **Анализ**, **Графика**, **Сервис**, **Данные**, **Окно**, **Справка**.

The toolbar contains icons for file operations (Open, Save, Print, Copy, Paste) and analysis actions (Add to Workbook, Add to Report).

The font settings are set to **Arial**, size **10**. The text is bolded (**B**), italicized (*I*), and underlined (U).

The left sidebar shows a list of statistical methods:

- Основное статистики и таблицы
- Множественная регрессия
- Дисперсионный анализ (ДА)
- Непараметрическая статистика
- Подгонка распределений
- Углубленные методы анализа
- Многомерный разведочный анализ
- Промышленная статистика и Шесть Сигма
- Анализ мощности, оценка объ
- Нейронные сети
- Добыча данных
- Добыча данных в управлении
- Текстовая добыча и сканиров
- Блочные статистики
- STATISTICA Visual Basic
- Вероятностный калькулятор

The main window displays a data table titled **Данные: Adstudy.sta (25v * 50c)**. The table content is as follows:

	Advertising Effectiveness	
	1	2
	GENDER	ADVERT
R. Rafuse	MALE	PEPSI
T. Leiker	MALE	COKE
E. Bizot	FEMALE	COKE
K. French	MALE	PEPSI
E. Van Landuyt	MALE	PEPSI
K. Harrell	FEMALE	COKE

The bottom status bar shows: **Нажмите F1 для вызова Электронног**, **H1,П1**, **MALE**, **Выбор:0**, and **Шесть сигма**.

STATISTICA



Основные статистики и таблицы: Таблица данных1

Быстрый

- Описательные статистики
- Парные и частные корреляции
- t-критерий для независимых выборок
- t-критерий для независимых переменных
- t-критерий для зависимых выборок
- t-критерий для однофакторного ДА
- Группировка и однофакторный ДА
- Таблицы частот
- Таблицы сопряженности, флагов и заголовков
- Таблицы многомерных откликов
- Другие критерии значимости
- Вероятностный калькулятор

Непараметрическая статистика: Таблица дан...

Быстрый

- Таблицы 2 x 2, хи/квadrat, Макнемара, точ...
- Сравнение наблюдаемых и ожидаемых частот
- Корреляция Спирмена, тау Кендалла, гамма
- Сравнение двух независимых групп
- Сравнение нескольких независимых групп
- Сравнение двух зависимых переменных
- Сравнение нескольких зависимых переменных
- Q критерий Кохрена
- Обычные описательные статистики (медиана, мода...)

Множественная регрессия: Таблица данных1

Быстрый

Зависимые: нет
Независимые: нет

Файл данных: Исходные данные

Подгонка распределений

Быстрый

Непрерывные распределения: Нормальное, Равномерное, Экспоненциальное, Гамма, Логнормальное, Хи квадрат, Другие ...

Дискретные распределения: Биномиальное, Пуассона, Геометрическое, Бернулли

Базовые модули STATISTICA

Обобщенные линейные и нелинейные модели: Таблица данных1

Быстрый

Вид анализа: Однофакторный ДА, Главные эффекты, Факторный ДА, Гнездовой план, Простая регрессия, Множественная регрессия, Факторная регрессия, Полиномиальная регрессия, Регрессия поверхности отклика

Задание анализа: Диалог, Мастер анализа, Редактор кода

Распределение: Нормальное, Пуассона, Гамма, Биномиальное, Полиномиальное

Общие линейные модели (GLM): Таблица данных1

Быстрый

Вид анализа: Однофакторный ДА, Главные эффекты, Факторный ДА, Гнездовой план, Большие баланс. планы, Повторные измерения, Простая регрессия, Множественная регрессия, Факторная регрессия, Полиномиальная регрессия, Регрессия поверхности отк., Регрессия поверхности см., Ковариационный анализ, Неоднородные коэф. наклона, Однородные коэф. наклона, Общие линейные модели

Анализ временных рядов: Таблица данных1

Переменные: нет

Методы: АРСС и автокорреляция, АРСС с интерв., Экспоненциальное сглажив., Фурье (спектральн.)

Общие регрессионные модели: Таблица данных1

Быстрый

Вид анализа: Простая регрессия, Множественная регрессия, Факторная регрессия, Полиномиальная регрессия, Регрессия поверхности отклика, Регрессия поверхности смеси, Однофакторный ДА, Главные эффекты, Факторный ДА

Множественная нелинейная регрессия: Таблица данных1

Быстрый

Переменные: нет

Удаление ПД: Построчное, Попарное, Замена средним

Углубленные методы анализа

Многомерный разведочный анализ

Надежность и позиционный анализ: 10Items.sla

Факторный анализ: Таблица данных1

Деревья классификации: Таблица данных1

Многомерное шкалирование: Nations

Главные компоненты: Таблица данных1

Канонический анализ: Factor

Анализ соответствий (АС): задание таблицы

Дискриминантный анализ: Таблица данных1

Общие модели дискриминантного анализа: Таблица данных1

Методы кластеризации: Таблица данных1

Промышленная статистика

Карты контроля качества: Таблица данных1

Планирование экспериментов (ПЭ): Таблица данных1

Процедуры анализа производственных процессов: Таблица данных1

Анализ мощности

Анализ мощности/интервальное оценивание: Таблица данных1

Оценка мощности критерия

Оценка объема выборки

Интервальное оценивание

Функции распределения

Вычислять мощность для заданной альтернативы

Построить графики мощности как функции объема выборки, ошибки первого рода α или величины тестируемого эффекта.

Одно выборочный t-критерий используется для сравнения выборочного среднего с заданным значением

Одно среднее, t-критерий

Два средних, t-критерий, независимые выборки

Два средних, t-критерий, зависимые выборки

Несколько средних, спланированный контраст

Несколько средних, однофакторный ДА

Несколько средних, двухфакторный ДА

Одна дисперсия, хи-квадрат критерий

Однородность дисперсий: F-критерий

Критерий значимости коэффициента корреляции для двух переменных

Квадраты коэфф. множественной корреляции

Одна пропорция, Z, хи-квадрат критерий

Две пропорции, Z-критерий

Две пропорции, критерий Макнемара

Анализ выживаемости - логранговый критерий

Анализ выживаемости - эксп. кум. период

Анализ выживаемости - эксп. кум. период, с инкл.

Моделирование структурными уравнениями

Нейронные сети

Нейронные Сети: Таблица данных1

Входные, выходные переменные

Зависимые: нет

Независимые: нет

Тип переменных

Непрерывные: нет

Категориальные: нет

Коды выбора: нет

Быстрый

Дополнительно

Сети/Ансамбли

Тип задачи

Регрессия

Классификация

Временные ряды

Кластерный анализ

Инструмент

Мастер решений

Конструктор сетей

Создает и тестирует сети для анализа данных и прогноза. Конструирует набор сетей для решения выбранной задачи.

Нет данных, опция недоступна.

Перемешать

Задать коды выбора

Обучающая: [] Используется для обучения

Контрольная: [] Используется для оценивания ошибки

Тестовая: [] Используется для сравнения альтернативных моделей

Игнорировать: [] Игнорируемые наблюдения исключаются из анализа

Выводить сообщение, если обнаружены пропущенные данные

ОБЗОР МЕТОДОВ *STATISTICA*

STATISTICA состоит из более чем 10 000 вычислительных процедур, сгруппированных в отдельные модули по основным направлениям анализа.

В каждом модуле вы можете выполнить определенный способ обработки, не обращаясь к процедурам из других модулей. Но также вы можете легко переключаться из одного модуля в другой при выполнении сложных исследований.

В целом систему *STATISTICA* можно представить в следующем виде:

Базовые модули *STATISTICA*. Обширный выбор основных статистик в сочетании с мощностью, производительностью и простотой использования технологии *STATISTICA*. Позволяют провести первичный анализ данных и подготовить аналитический отчет.

***STATISTICA* Углубленные методы анализа.** Большой набор самых современных инструментов для моделирования и прогнозирования, включающий возможность автоматического выбора оптимальных моделей и расширенные интерактивные средства визуализации. Позволяет строить сложные объяснительные модели зависимостей и прогноза с помощью классических моделей временных рядов.

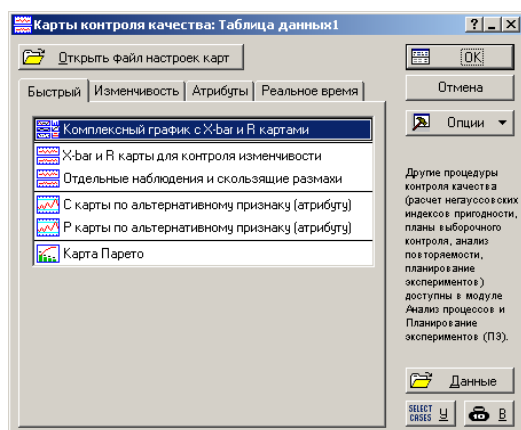
***STATISTICA* Многомерный разведочный анализ.** Предоставляет широкий выбор разведочных технологий анализа различных типов данных в сочетании с богатыми интерактивными средствами визуализации. Включает методы кластерного, факторного, дискриминантного, канонического анализа, многомерного шкалирования, деревьев классификации, анализа соответствий, анализа главных компонент. Позволяет проводить исследования многомерных данных, решать задачи классификации с минимальными ограничениями на число переменных и объем данных.

***STATISTICA* Нейронные сети.** *Впервые на русском языке!* Предоставляет самый обширный выбор нейросетевых методов с автоматическим мастером решения задач. Дополнительно предлагается генератор Си-кода.

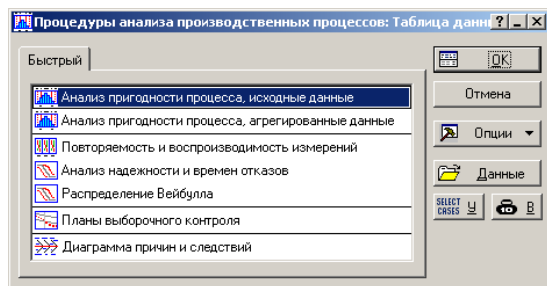
***STATISTICA* Анализ мощности.** Предельно точный и удобный для пользователя специализированный инструмент для всестороннего анализа статистической мощности и вычисления объема выборки.

Промышленная статистика и Шесть Сигма

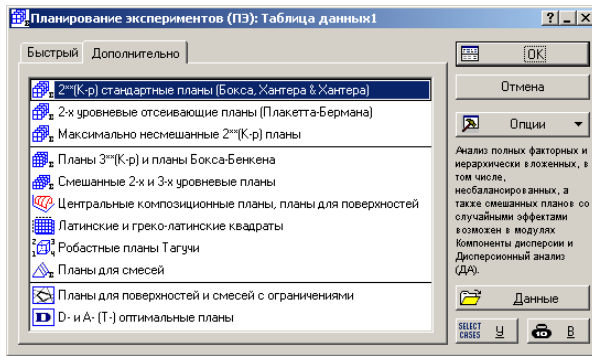
***STATISTICA* Карты контроля качества.** Предоставляет полностью настраиваемые (в частности, доступные из других приложений), простые в использовании графики контроля качества, карты Шухарта, а также набор средств автоматизации для упрощения решения ежедневных задач управления качеством и прогнозирования.



***STATISTICA* Анализ процессов.** Удобный пакет для анализа пригодности процессов, повторяемости и воспроизводимости измерений, а также других приложений контроля и повышения качества, включая разнообразные инструменты поддержки методологии Шесть Сигма.

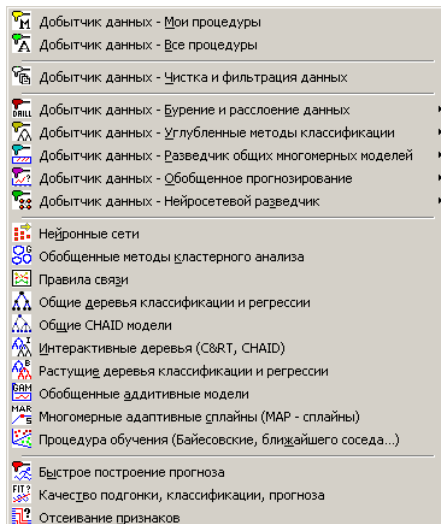


***STATISTICA* Планирование экспериментов.** Огромный выбор методов планирования экспериментов и связанных с ними графических средств, включая интерактивные профили желательности.



Корпоративные системы STATISTICA

STATISTICA Enterprise-Wide Data Mining System (Data Miner). Уникальная система добычи данных, построения прогнозов, классификаций и зависимостей, включая самые современные средства бурения и расслоения, чистки и фильтрации данных, растущие деревья классификации и регрессии, ассоциативные правила, процедуры обучения и др.

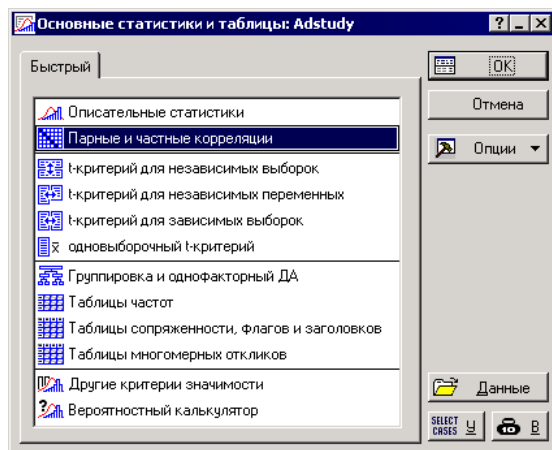


STATISTICA Enterprise-Wide Data Analysis System (SEDAS). Интегрированная многопользовательская система для решения общих задач анализа данных и для бизнес-приложений в области маркетинга, финансов и т.д.

STATISTICA Enterprise-Wide SPC System (SEWSS). Предназначена для локальных и глобальных корпоративных приложений по контролю и улучшению качества, включая методику Шесть Сигма.

Базовые модули STATISTICA

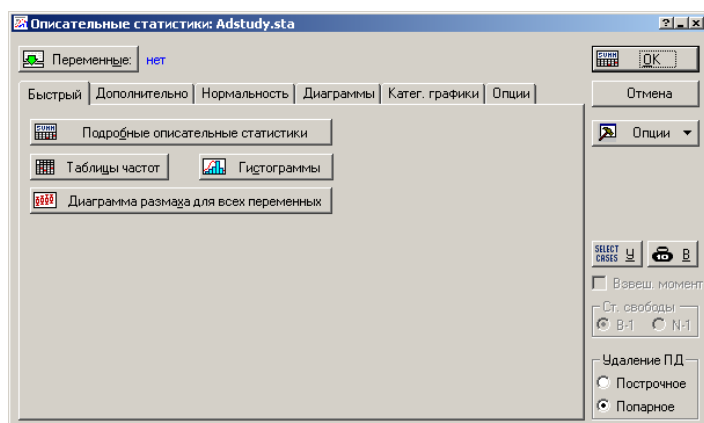
Модуль Основные статистики и таблицы



*Если вы хотите провести описательный анализ данных, разбить данные различными способами на группы, просмотреть эти группы визуально и вычислить описательные статистики для группированных данных, а также вычислить парные и частные корреляции, запустите модуль **Основные статистики и таблицы**.*

Обычно с этого модуля начинается работа в системе. С ним работает большинство пользователей, потому что он позволяет подготовить изысканный аналитический отчет.

Описательные статистики



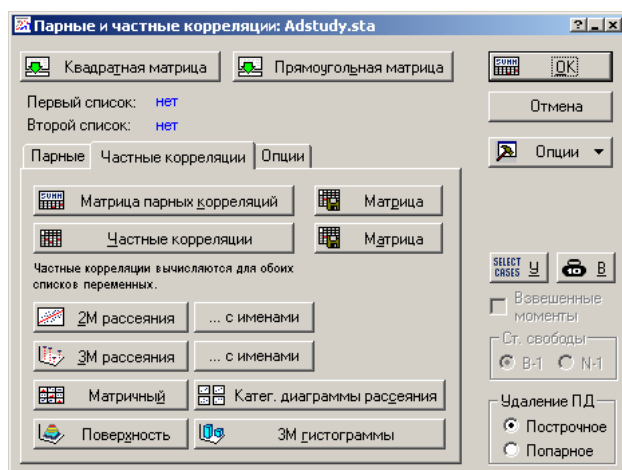
Выберите эту процедуру для вычисления основных описательных статистик.

Описательные статистики вычисляются отдельно для каждой переменной в файле данных, а также для групп переменных. Они предоставляют исследователю основную описательную информацию относительно распределения переменной: среднее, минимальное и максимальное значение, различные меры изменчивости или вариабельности дисперсий, а также характеристики формы распределения (асимметрия, эксцесс). Определения и формулы см. в книге “Искусство анализа данных на компьютере”, [2].

Для вычисления описательных статистик нужно выбрать переменные в текущем файле данных (можно выбрать все переменные одним щелчком кнопки). Далее STATISTICA построит таблицу описательных статистик, расположив их в отдельной строке для каждой переменной.

Меры изменчивости (вариабельности) включают стандартное отклонение и стандартную ошибку. Кроме того, доступны различные критерии нормальности.

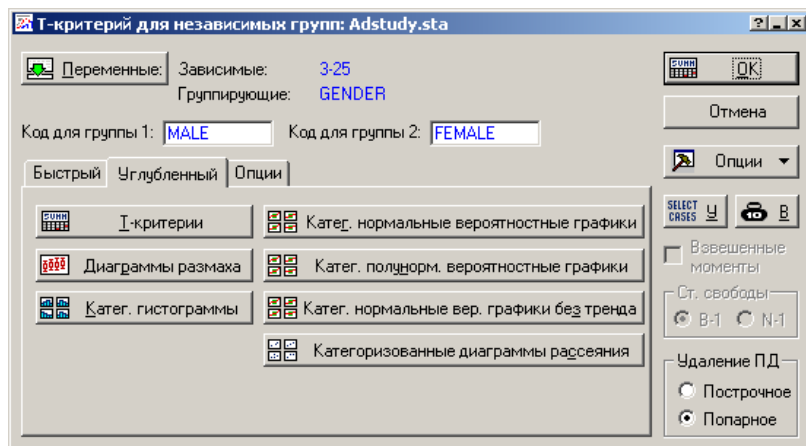
Парные и частные корреляции



Выберите эту процедуру для вычисления коэффициента корреляции Пирсона. Модуль позволяет вычислить квадратные и прямоугольные корреляционные матрицы, а также расширенные матрицы, куда входит число пар n (наблюдений, по которым вычислены корреляции) и уровни значимости. (Определения и формулы см. в книге “Искусство анализа данных на компьютере”, [2].)

Заметим, что непараметрические корреляции (например, ранговая корреляция **Спирмена R**, статистика **tau Кендалла**, **Гамма статистика** и т. д.) могут быть вычислены в модуле **Непараметрическая статистика**, различные меры расстояния можно вычислить в модуле **Кластерный анализ**.

t-критерий для независимых выборок



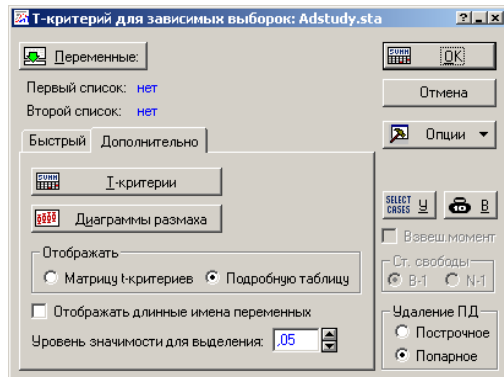
Эта процедура позволяет сравнить средние двух независимых выборок (например, число лейкоцитов у мужчин и у женщин).

t-критерий Стьюдента является наиболее часто используемым методом обнаружения различия между средними двух выборок. Например, *t*-критерий можно использовать для сравнения средних показателей группы пациентов, принимавших определенное лекарство, с контрольной группой, где принималось безвредное лекарство.

Теоретически, *t*-критерий может применяться даже для выборок небольшого размера (например, 10) и в случаях, если переменные нормально распределены (внутри групп), а дисперсии наблюдений в группах не слишком различны.

Предположение о нормальности можно проверить, исследуя распределение визуально с помощью гистограммы или применяя какой-либо критерий нормальности.

Равенство дисперсий в двух группах можно проверить с помощью *F*-критерия (который включен в таблицу вывода *t*-критерия) или с помощью критерия Левена. Также в STATISTICA доступен критерий Брауна-Форсайта.

t-критерий для зависимых выборок

Эта процедура позволяет сравнить средние двух зависимых выборок (например, кровяное давление одной и той же выборки пациентов до и после приема определенного лекарства).

t-критерий для зависимых выборок очень полезен в тех довольно часто возникающих на практике ситуациях, когда важный источник внутригрупповой вариации (или ошибки) может быть легко определен и исключен из анализа. Например, это относится к экспериментам, в которых две сравниваемые группы основываются на одной и той же совокупности наблюдений (субъектов), которые тестировались дважды (например, до и после лечения, до и после приема лекарства). Таким образом, исследуемые выборки являются зависимыми, т.к. сравниваются одни и те же субъекты.

В подобных экспериментах значительная часть внутригрупповой изменчивости (вариации) в обеих группах может быть объяснена индивидуальными различиями субъектов.

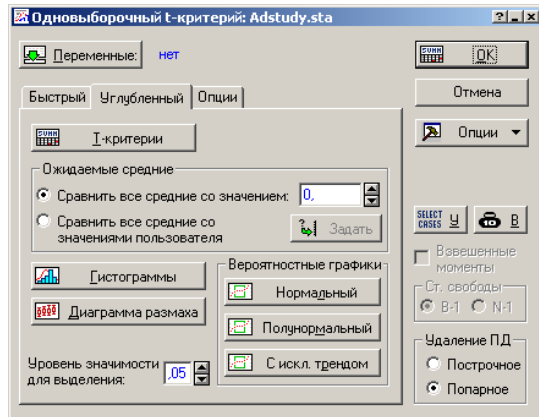
Заметим, что на самом деле, такая ситуация не слишком отличается от той, когда сравниваемые группы совершенно независимы, где индивидуальные отличия также вносят вклад в дисперсию ошибки. Однако в случае независимых выборок, вы ничего не сможете поделать с этим, т.к. не сможете определить (или “удалить”) 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 приема лекарства” и “после приема лекарства”) для каждого субъекта. Вычитая первые значения из вторых (для каждого субъекта) и анализируя затем только эти “чистые (парные) разности”, вы исключите ту часть вариации, которая является результатом различия в

исходных уровнях индивидуумов. Именно так и проводятся вычисления в t -критерии для зависимых выборок.

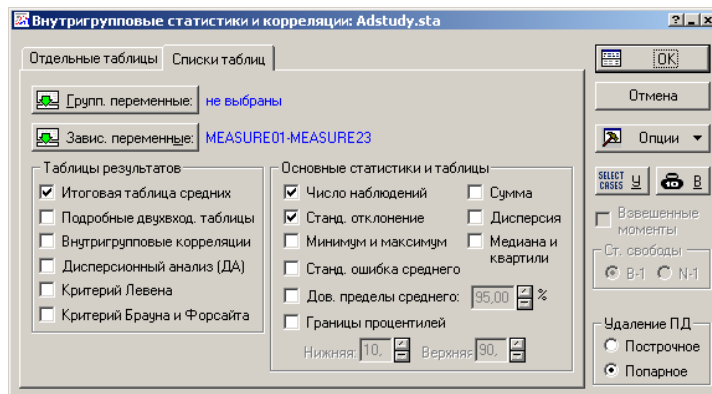
В сравнении с t -критерием для независимых выборок, такой подход дает всегда лучший результат (критерий становится более чувствительным).

Одновыборочный t -критерий



Эта процедура позволяет проверить гипотезу о равенстве выборочного среднего некоторому заданному числу, когда размер выборки небольшой и/или когда неизвестна выборочная дисперсия. В одновыборочных t -критериях наблюдаемое среднее (вычисленное по реализации выборки) сравнивается с ожидаемым средним выборки (т. е. с некоторым теоретическим средним). При этом выборочная дисперсия оценивается по наблюдаемой реализации выборки.

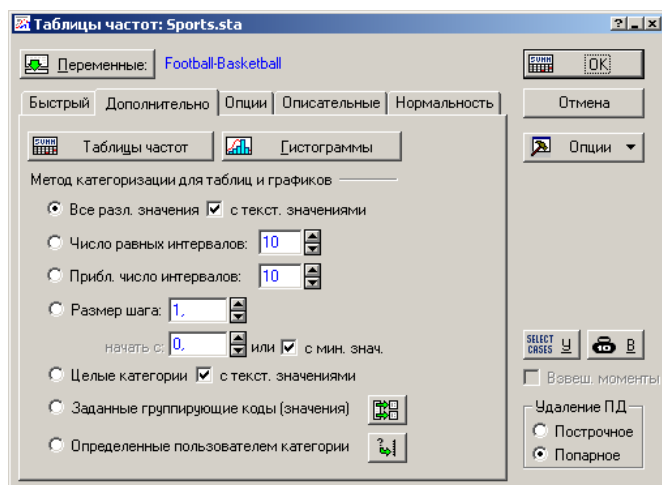
Группировка и однофакторный ДА



Эта процедура вычислит различные описательные статистики для данных, разбитых на группы одной или несколькими группирующими переменными (например, переменными **Пол** или **Область**). Для выбранных переменных также производится однофакторный дисперсионный анализ.

Вы можете исследовать очень большое число групп (более 200; такого размера планы встречаются, например, в сельском хозяйстве и промышленности).

Таблицы частот

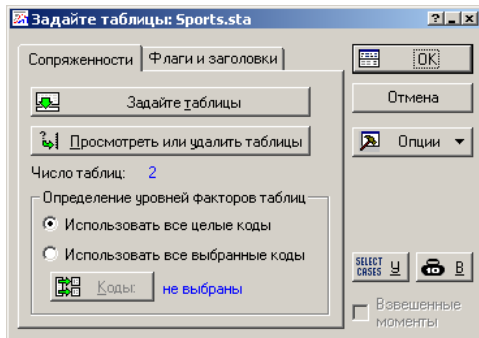


Эта процедура позволяет вам построить таблицы частот (и гистограммы).

Частоты или одноходовые таблицы представляют собой простейший метод анализа категориальных переменных. Часто их используют как одну из процедур разведочного анализа, чтобы посмотреть, каким образом различные группы данных распределены в выборке. При этом исходные данные (измеренные в любой подходящей шкале) представляются в виде частот наблюдений, попавших в некоторые определенные исследователем категории или классы.

STATISTICA обеспечивает различные опции для определения категорий в таблицах частот (например, целые интервалы, определенные коды и т.д.). Пользователь также может разбить данные в соответствии с условиями логической группировки, которые могут быть легко введены в систему.

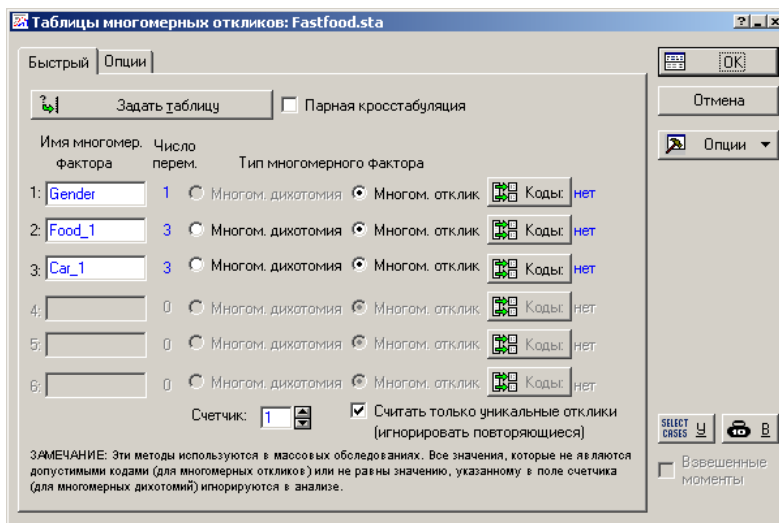
Таблицы сопряженности, таблицы флагов и заголовков



Эта процедура позволяет вам кросстабулировать данные и построить различные типы таблиц сопряженности. Доступен самый широкий выбор статистик для двуходовых таблиц. Таблицы флагов и заголовков или, кратко, таблицы заголовков позволяют отобразить несколько двуходовых таблиц сопряженности в сжатом виде.

Заметим, что модуль **Логлинейный анализ** и модуль **Анализ соответствий** позволяет построить многоходовые таблицы сопряженности и выполнять сложный анализ, например, построить модель связи.

Таблицы многомерных откликов



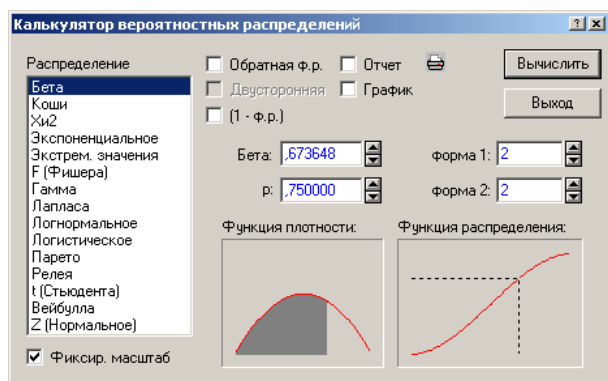
Модуль **Основные статистики и таблицы** предоставляет разнообразные возможности построения таблиц для переменных с многомерными откликами, а также для многомерных дихотомий.

Обычно группирующие переменные или факторы делят выборку на непересекающиеся группы, например, класс мужчин и женщин. Понятно, что достаточно только одной группирующей переменной, чтобы закодировать пол субъекта. Однако в некоторых исследованиях категории не являются взаимоисключающими.

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 (классы), очевидно, не являются взаимно исключающими. Действительно, респондент может отметить 3 различных напитка как наиболее предпочтительные. Следовательно, если наблюдение - это субъект, то для трех различных группирующих переменных это наблюдение является общим.

Такие группирующие переменные называются переменными с *многомерными откликами* (многомерные дихотомии по существу, схожи с ними). Эти переменные легко анализировать в модуле **Основные статистики и таблицы**.

Вероятностный калькулятор



Калькулятор вероятностных распределений (вероятностный калькулятор) позволяет вычислить для различных распределений верхние или нижние квантили (для симметричных распределений также можно вычислить двусторонние значения вероятности).

Одной из полезных функций вероятностного калькулятора является вычисление критических значений, а также уровней значимости различных статистик, имеющих определенное распределение (обычно это распределение Фишера F, распределение Стьюдента t или нормальное распределение z).

На интерактивной графической панели выводятся изображения функции плотности и функции распределения выбранного пользователем закона распределения с заданными параметрами.

Более формально действия вероятностного калькулятора можно описать следующим образом.

С каждым распределением случайной величины, принимающей действительные значения, 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 ($P\{\xi \leq x\} = p, p \in [0, 1]$).

Вероятностный калькулятор позволяет по любому x вычислить p , а 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).

Другие критерии значимости

Процедура позволяет вычислить наиболее общие критерии значимости. В частности, проверить значимость различия средних, коэффициентов корреляции, процентилей (для независимых выборок).

Модуль Множественная регрессия



Если вы хотите построить зависимость между несколькими переменными, подобрать простую линейную модель и оценить ее адекватность, воспользуйтесь модулем **Множественная регрессия**.

Регрессионные модели важны в задачах прогнозирования, когда по значениям независимых переменных предсказывается значение зависимой переменной.

Общее назначение

Общее назначение множественной (многомерной) регрессии состоит в анализе связи между несколькими независимыми переменными (называемыми также *регрессорами* или *предикторами*) и зависимой переменной (*откликом*).

Например, агент по продаже недвижимости может составить список, в котором записаны данные о домах в определенном районе: размер дома, число спален, средний доход населения в этом районе в соответствии с данными переписи, субъективная оценка привлекательности дома. Как только эта информация собрана для различных домов, было бы интересно посмотреть, связаны ли эти характеристики дома с ценой, по которой он был продан.

Может оказаться, что число спальных комнат является лучшим предиктором для цены продажи дома в некотором районе, чем привлекательность дома (субъективная оценка). Также могут 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:

$$\text{Salary} = .5 * \text{Resp} + .8 * \text{No_Super}$$

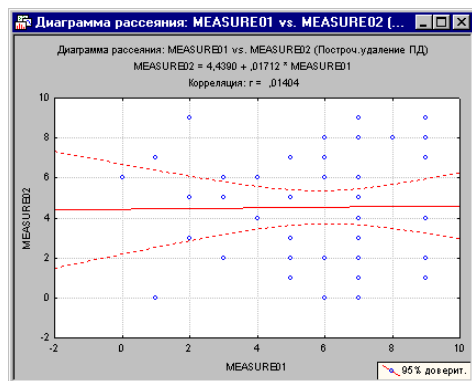
Как только эта так называемая линия регрессии определена, аналитик имеет возможность построить график ожидаемой (предсказанной) оплаты труда и реальных обязательств компании по выплате жалования.

Таким образом, аналитик может определить, какие позиции недооценены (лежат ниже линии регрессии), какие оплачиваются слишком высоко (лежат выше линии регрессии), а какие оплачены адекватно.

Вычислительные аспекты

Общая вычислительная задача, которую требуется решать при анализе методом множественной регрессии, состоит в подгонке гиперплоскости к некоторому набору точек.

В простейшем случае, когда имеется одна зависимая и одна независимая переменная, это можно увидеть на диаграмме рассеяния.



Метод наименьших квадратов. Программа строит линию регрессии так, чтобы минимизировать квадраты отклонений этой линии от наблюдаемых точек. Поэтому на эту общую процедуру иногда ссылаются как на оценивание по *методу наименьших квадратов*.

Метод взвешенных наименьших квадратов. В некоторых случаях при проведении регрессионного анализа желательно использовать различные веса наблюдений и вычислить оценки коэффициентов регрессии по *методу взвешенных наименьших квадратов*. Этот метод обычно применяется, когда дисперсия остатков неоднородна при различных значениях независимых переменных.

Уравнение регрессии. Прямая линия на плоскости (в пространстве двух измерений) задается уравнением $Y = a + b * X$, то есть переменная Y может быть выражена через константу (a) и угловой коэффициент (b), умноженный на переменную X . Константу иногда называют также свободным членом, а угловой коэффициент – *коэффициентом регрессии* или *B-коэффициентом*.

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 **Множественная регрессия**. В общем случае, процедуры множественной регрессии оценивают параметры линейного уравнения вида:

$$Y = a + b_1 * X_1 + b_2 * X_2 + \dots + b_p * X_p$$

Однозначный прогноз и частная корреляция. Регрессионные коэффициенты представляют независимые вклады каждой независимой переменной в предсказание зависимой переменной. К примеру, переменная X_1 коррелирует с переменной Y после учета влияния всех других независимых переменных. Этот тип корреляции называется частной корреляцией. Поясним данное понятие на примере.

Вы можете обнаружить значимую отрицательную корреляцию в популяции между длиной волос и ростом человека (невысокие люди обладают более длинными волосами). На первый взгляд это может показаться странным. Но если добавить переменную *Пол* в уравнение множественной регрессии, эта корреляция, скорее всего, исчезнет.

Произойдет это из-за того, что средняя длина волос у женщин больше, чем у мужчин. При этом они также в среднем ниже мужчин. Таким образом, после удаления разницы по полу посредством ввода предиктора *Пол* 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*. Другими

словами, при учете переменной **Пол** частная корреляция между длиной волос и ростом нулевая.

Если одна величина коррелирована с другой, то это может быть отражением того факта, что они обе коррелированы с третьей величиной или с совокупностью величин. (Более подробно см. книгу “Искусство анализа данных на компьютере”, [2].)

Предположения, ограничения и обсуждения 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 часто не может быть подтверждено.

К счастью, процедуры множественного регрессионного анализа в незначительной степени подвержены воздействию малых отклонений от этого предположения. Тем не менее, 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 преобразования переменных (с помощью программы на языке *STATISTICA* Visual Basic или *формул электронных таблиц*) 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**).

Предположение нормальности. Во множественной регрессии предполагается, что остатки (предсказанные значения минус наблюдаемые) распределены нормально (т.е. подчиняются закону нормального распределения). И хотя большинство тестов (в особенности *F*-тест) довольно устойчивы по отношению к отклонениям от этого предположения, следует всегда обращать внимание на распределения представляющих интерес переменных, прежде чем делать окончательные выводы.

В модуле **Множественная регрессия** вы можете построить гистограмму и нормальный вероятностный график остатков для визуального анализа их распределения.

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. 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 ненадежными и, скорее всего, невоспроизводимыми для желающих повторить это исследование.

Мультиколлинеарность и плохая обусловленность матрицы. Проблема мультиколлинеарности является общей для многих методов корреляционного анализа. Представим, что имеется два предиктора (переменные X) для роста субъекта: вес в фунтах и вес в унциях. Очевидно, что иметь оба предиктора совершенно излишне. Вес является одной и той же переменной, измеряется он в фунтах или унциях. Попытка определить, какая из двух мер является лучшим предиктором, выглядит довольно глупо.

Поэтому при попытке выполнить множественный регрессионный анализ с ростом в качестве зависимой переменной и двумя мерами веса в качестве независимых переменных, STATISTICA выдаст сообщение о *плохой обусловленности матрицы*.

Модуль **Множественная регрессия** содержит много статистических индикаторов избыточности (*толерантность*, *получастное R* и т.д.), а также некоторые средства борьбы с избыточностью (например, метод *гребневой регрессии*).

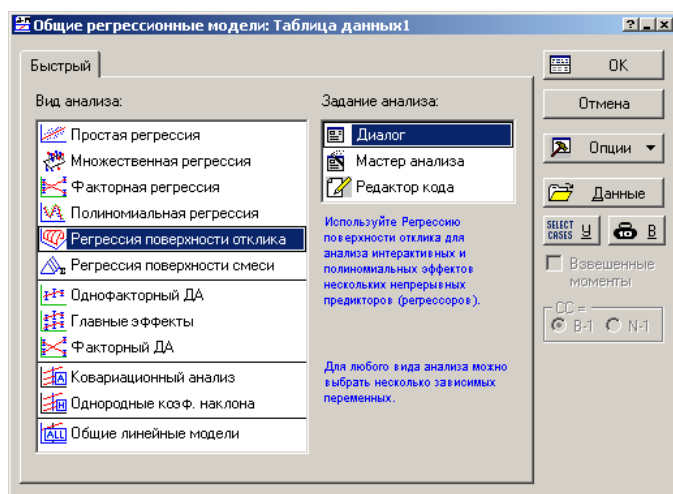
Важность анализа остатков. Хотя большинство предположений множественной регрессии нельзя в точности проверить, исследователь может обнаружить отклонения от этих предположений.

В частности, выбросы (т. е. экстремальные наблюдения) могут вызвать серьезное смещение оценок, “сдвигая” линию регрессии в определенном направлении и, тем самым, вызывая смещение регрессионных коэффициентов.

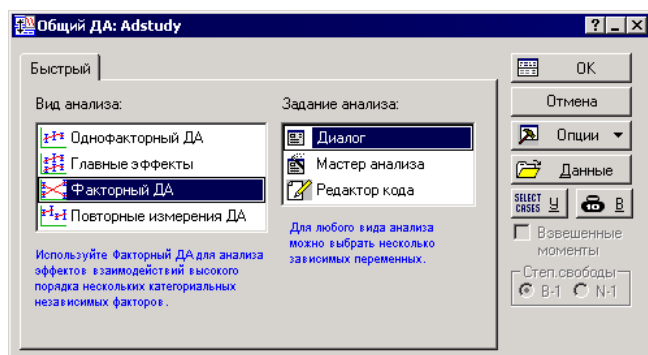
Часто исключение всего одного экстремального наблюдения приводит к совершенно другому результату. Поэтому одна из задач при создании интерфейса модуля **Множественная регрессия** состояла в упрощении анализа остатков для пользователя, чтобы отклонения от предположений, угрожающие обоснованности результатов анализа, могли быть легко обнаружены.

Замечание

Развитием регрессии являются общие регрессионные модели *STATISTICA*.



Модуль Общий Дисперсионный анализ (ДА)



Если вам известны факторы, которые воздействуют на измеряемые вами данные, и вы хотите оценить степень этого воздействия, понять, существенно оно или нет, выберите модуль **Общий Дисперсионный анализ**.

Термин ANOVA (Analysis of Variance) переводится как Анализ вариаций или Дисперсионный анализ (ДА). Термин MANOVA (Multiple Analysis of Variance) переводится как Многомерный дисперсионный анализ (МДА).

Основные идеи

Основной целью дисперсионного анализа является исследование значимости различия между средними.

ДА основывается на развитии t -критерия. Ключевым является наличие нескольких выборок.

Если вы просто сравниваете средние в двух выборках, дисперсионный анализ даст тот же результат, что и 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).

Может показаться странным, что процедура сравнения средних называется *дисперсионным* анализом. В действительности, это связано с тем, что при исследовании статистической значимости различия между средними двух (или нескольких) групп мы на самом деле анализируем выборочные дисперсии.

Разбиение суммы квадратов

В основе дисперсионного анализа лежит разделение дисперсии на части или компоненты. Для выборки объема N выборочная дисперсия вычисляется как сумма квадратов отклонений от выборочного среднего, деленная на N-1 (объем выборки минус единица).

Рассмотрим следующий набор данных:

	Группа 1 (X)	Группа 2 (Y)
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	

$$\text{Среднее: } \bar{X} = \frac{1}{3} * (X1 + X2 + X3), \bar{Y} = \frac{1}{3} * (Y1 + Y2 + Y3).$$

$$\text{Сумма квадратов (SS): } \sum_{i=1}^3 (Xi - \text{Среднее})^2, \sum_{i=1}^3 (Yi - \text{Среднее})^2.$$

$$\text{Общее среднее} = \frac{1}{3} * (X1 + X2 + X3) + \frac{1}{3} * (Y1 + Y2 + Y3).$$

$$\text{Общая сумма квадратов} = \sum_{i=1}^3 (Xi - \text{Общее среднее})^2 + \sum_{i=1}^3 (Yi - \text{Общее среднее})^2.$$

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 (от английского Sum of Squares - Сумма Квадратов), исходя из общего среднего этих двух выборок, то получим величину 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 суммами квадратов.

Если использовать для анализа этих данных модуль **Общий Дисперсионный анализ**, то будет получена следующая таблица, называемая таблицей дисперсионного анализа:

	Главный Эффект				
	SS	ст. св.	MS	F	p
Эффект	24.0	1	24.0	24.0	.008
Ошибка	4.0	4	1.0		

SS – суммы квадратов.

ст. св. – число степеней свободы.

MS – средние квадраты, $MS = SS/ст. св.$

F – статистика Фишера, $F = MS_{\text{эффект}}/MS_{\text{ошибка}}$.

p – уровень статистической значимости.

Как видно из таблицы, общая сумма квадратов SS, равная 28, разбита на компоненты: сумму квадратов, обусловленную внутригрупповой изменчивостью ($2+2 = 4$; см. вторую строку), и сумму квадратов, обусловленную различием средних значений между группами ($28-(2+2) = 24$; см. первую строку).

SS ошибок и SS эффекта. Внутригрупповая изменчивость обычно называется остаточной компонентой или дисперсией *ошибки*. Это значит, что обычно при проведении эксперимента она может быть предсказана или объяснена.

С другой стороны, *SS эффекта* можно объяснить различиями между средними значениями в группах. Иными словами, принадлежность к некоторой группе объясняет межгрупповую изменчивость, т. к. нам известно, что эти группы обладают разными средними значениями.

Проверка значимости. Проверка значимости в дисперсионном анализе основана на сравнении компоненты дисперсии, обусловленной межгрупповым разбросом (называемым *средним квадратом эффекта*), и компоненты дисперсии, обусловленной внутригрупповым разбросом (называемым *средним квадратом ошибки*).

Если верна нулевая гипотеза (равенство средних в двух популяциях), то можно ожидать, что различие выборочных средних из-за чисто случайной изменчивости будет сравнительно небольшим. Поэтому при нулевой гипотезе внутригрупповая дисперсия будет практически совпадать с общей дисперсией, подсчитанной без учета групповой принадлежности.

Полученные внутригрупповые дисперсии можно сравнить с помощью *F*-критерия, проверяющего, действительно ли отношение дисперсий значимо больше 1. В рассмотренном примере критерий показывает, что различие между средними статистически значимо.

Основная логика дисперсионного анализа. Подводя итоги, можно сказать, что целью дисперсионного анализа является проверка статистической значимости между средними (для групп или переменных).

Эта проверка проводится с помощью разбиения суммы квадратов на компоненты, 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 существенном различии между средними.

Многофакторный дисперсионный анализ

В рассмотренном выше простом примере (см. *Разбиение суммы квадратов*) можно сразу вычислить *t*-критерий для независимых выборок, используя соответствующую опцию модуля **Основные статистики и таблицы**. Полученные результаты, естественно, совпадут с результатами дисперсионного анализа. Однако модуль **Общий Дисперсионный анализ** содержит гораздо более гибкие и мощные технические средства, позволяющие исследовать планы практически неограниченной сложности.

Множественные факторы. 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 существенно более эффективен, и мы можем извлечь больше информации из малых выборок.

Управление факторами. Предположим, что в рассмотренном выше примере анализа двух выборок мы добавим еще один фактор, например, фактор **Пол**. Пусть каждая группа теперь состоит из 3 мужчин и 3 женщин. План этого эксперимента можно представить в виде таблицы 2 на 2:

	Экспериментальная группа 1	Экспериментальная группа 2
Мужчины	2 3 1	6 7 5
Среднее	2	6
Женщины	4 5 3	8 9 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),
- (2) изменчивость, связанная с принадлежностью к экспериментальной группе,
- (3) изменчивость, обусловленная полом объектов наблюдения.

(Отметим, что существует еще один возможный источник изменчивости - взаимодействие факторов. Его мы обсудим позднее.)

Что произойдет, если мы не будем рассматривать фактор **Пол** и вычислим обычный *t*-критерий? Если мы вычислим сумму квадратов, игнорируя **Пол** (т. е. объединяя объекты разного пола в одну группу при вычислении внутригрупповой дисперсии; в этом случае сумма квадратов для каждой группы будет равна 10, а общая сумма квадратов будет равна $10+10 = 20$), то получим большее значение внутригрупповой дисперсии, чем при более точном анализе с дополнительным разбиением на подгруппы по фактору **Пол** (при

этом внутригрупповые средние будут равны 2, а внутригрупповая сумма квадратов равна $2+2+2+2 = 8$).

Итак, при введении дополнительного фактора, остаточная дисперсия уменьшилась. Это связано с тем, что среднее значение для **Мужчин** меньше, чем среднее значение для **Женщин**. Это различие в средних увеличивает суммарную внутригрупповую изменчивость, если фактор пола не учитывается. Управление дисперсией ошибки увеличивает чувствительность (мощность) критерия.

На этом примере видно преимущество дисперсионного анализа по сравнению с обычным t -критерием для двух выборок. Дисперсионный анализ позволяет изучать каждый фактор, управляя значениями других факторов. Это, в действительности, и является основной причиной его большой статистической мощности (для получения значимых результатов требуется 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:

	Целеустремленные	Ленивые
Трудное задание	10	5
Легкое задание	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.

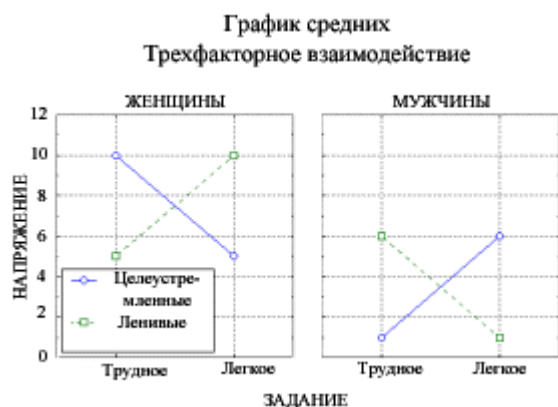
Анализируя результаты, правильнее было бы сказать, что над сложными заданиями работают упорнее только целеустремленные студенты, в то время как над легкими заданиями упорнее работают ленивые студенты.

Другими словами, характер студентов и сложность задания взаимодействуют в своем влиянии на затрачиваемое усилие. Это пример 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 сравнительно легко, взаимодействия высших порядков объяснить значительно сложнее. Представим себе, что в рассматриваемый выше пример введен фактор *Пол*, и мы получили следующую таблицу средних значений:

Женщины	Целеустремленные	Ленивые
Трудное задание	10	5
Легкое задание	5	10
Мужчины	Целеустремленные	Ленивые
Трудное задание	1	6
Легкое задание	6	1

Какие теперь выводы можно сделать из полученных результатов? Графики средних позволяют легко интерпретировать сложные эффекты. В модуле **Общий Дисперсионный анализ** вы можете построить эти графики практически одним щелчком мыши. Изображение на представленных ниже графиках 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.

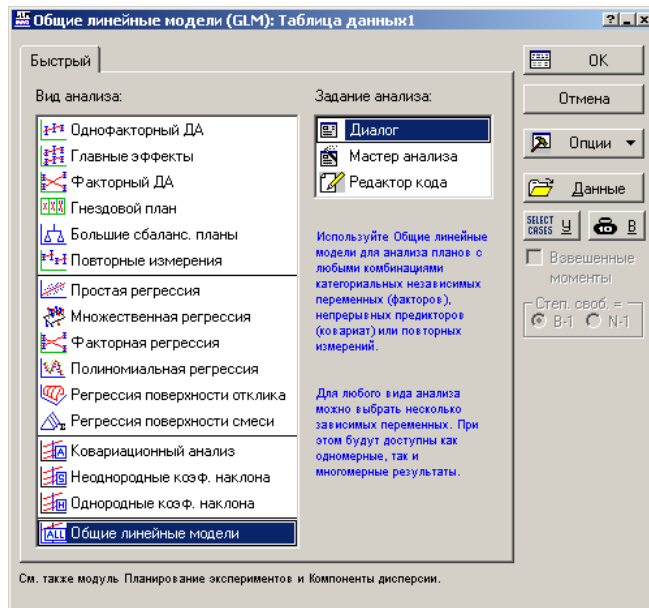
Для взаимодействия трех факторов из предыдущего параграфа можно сказать, что взаимодействие двух факторов (сложности задачи и характера студента) изменяется под воздействием фактора **Пол**. Если изучается взаимодействие четырех факторов, можно сказать, что взаимодействие трех факторов, изменяется под воздействием четвертого фактора, т.е. существуют различные типы взаимодействий на разных уровнях четвертого фактора.

Оказалось, что во многих областях исследований обычным явлением является взаимодействие пяти или даже большего количества факторов.

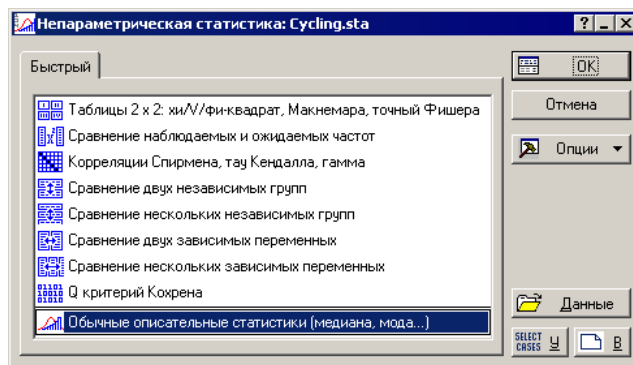
Замечание

Модуль **Общий Дисперсионный анализ** является подмножеством модуля **Общие линейные модели** системы STATISTICA (все планы, которые можно анализировать в

модуле **Общий Дисперсионный анализ**, также можно анализировать в модуле **Общие линейные модели**).



Модуль Непараметрическая статистика



Основное назначение

Для того чтобы понять идею *непараметрических* методов, следует сначала понять идею параметрической статистики.

Если вы знаете исходное распределение наблюдаемой переменной, то вы 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).

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%-ную критическую область t -распределения, то можно сделать обоснованный вывод, что жители Вышгорода действительно в среднем более высокие (чем в целом в популяции).

В рассмотренном примере использовался тот факт, что в повторных выборках равного объема средние значения (роста людей) будут иметь t -распределение (с определенным средним и дисперсией). Однако это верно лишь в том случае, если рассматриваемая переменная (рост) имеет *нормальное распределение*. Для многих изучаемых переменных это условие не выполнено. Например, случаи редких болезней не являются нормально распределенными в популяции, число автомобильных аварий также не является нормально распределенным.

Другим фактором, часто ограничивающим применимость критерия, основанного на предположении нормальности, является *объем или размер выборки*, доступной для анализа (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.

Использование критериев, основанных на предположении 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 А в два раза выше, чем успеваемость студента С? Является ли различие между средним баллом студентов В и А сравнимым с различием между студентами D и С?

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.

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, такие предположения часто неестественны, как в данном примере, и данные скорее просто ранжированы (измерены в порядковой шкале) или упорядочены, чем измерены точно.

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 точно, 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 *свободными от параметров* или *свободно распределенными методами*.

Краткий обзор непараметрических процедур ДА

По существу, для каждого параметрического критерия имеется, по крайней мере, один непараметрический аналог. Эти критерии можно отнести к одной из следующих групп:

- 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 необходимо сравнить относительно среднего значения некоторой изучаемой переменной, используется t -критерий для независимых выборок. Непараметрическими альтернативами этому критерию являются критерий серий Вальда-Вольфовица, U критерий Манна-Уитни и двухвыборочный критерий Колмогорова-Смирнова.

Если имеется несколько групп, то можно использовать дисперсионный анализ. Его непараметрическими аналогами являются 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 используется t -критерий для зависимых выборок. Альтернативными непараметрическими тестами являются критерий знаков и критерий Вилкоксона.

Если рассматриваемые переменные по природе своей категориальны, то подходящим является критерий хи-квадрат Макнемара.

Если рассматривается более двух переменных, относящихся к одной и той же выборке, то обычно используется дисперсионный анализ с повторными измерениями. Альтернативными непараметрическими методами являются ранговый дисперсионный анализ Фридмана и Q критерий Кохрена.

Зависимости между переменными. Для того чтобы оценить зависимость между двумя переменными, обычно вычисляют коэффициент корреляции. Непараметрическими эквивалентами стандартного коэффициента корреляции Пирсона являются статистика **Спирмена R** , статистика **Кендалла τ** и **Гамма статистика**.

Если две рассматриваемые переменные по природе своей категориальны или являются категоризованными, то подходящими непараметрическими критериями для проверки наличия связей (зависимостей) между ними будут критерий хи-квадрат, статистика фи коэффициент и точный критерий Фишера.

Описательные статистики. Если данные не являются нормально распределенными, а измерения, в лучшем случае, содержат ранжированную информацию, то вычисление обычных описательных статистик (например, среднего, стандартного отклонения) 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.

Модуль **Непараметрическая статистика** вычислит разнообразный набор мер положения (среднее, медиана, мода и т.д.) и рассеяния (дисперсия, среднее отклонение, квартильный размах и т.д.), позволяющих представить более полную “картину” данных.

Какой метод использовать

Нелегко дать простой совет, касающийся использования непараметрических процедур. Каждая непараметрическая процедура в модуле имеет свои достоинства и свои недостатки.

Например, двухвыборочный критерий Колмогорова-Смирнова чувствителен не только к различию в положении двух распределений, например, к различию средних, но также чувствителен и к форме распределения. Критерий 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, то всегда целесообразно применить различные непараметрические тесты. Возможно, результаты проверки (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 отклонения (например, выяснить, является ли определенная 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, $n > 100$), то имеет мало смысла использовать непараметрическую статистику.

Дело 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, всегда подходят для больших выборок.

Большинство критериев значимости многих непараметрических статистик основываются на асимптотической теории (больших выборок) поэтому соответствующие тесты часто не выполняются, если размер выборки становится слишком малым.

Ранговые корреляции

Чаще всего используются следующие коэффициенты непараметрической корреляции: статистика **Спирмена R**, статистика **Кендалла тау** и **Гамма статистика**.

Статистика Спирмена R. Статистика **R Спирмена** предполагает, что рассматриваемые переменные измерены как минимум в порядковой шкале, иными словами, наблюдения ранжированы.

Статистика Кендалла тау. Статистика **тау Кендалла** эквивалентна статистике **R Спирмена** при выполнении некоторых основных предположений. Критерии, основанные на этих статистиках, также сравнимы по мощности.

Тем не менее, обычно значения **R Спирмена** и **тау Кендалла** различны, что связано с существенным отличием этих статистик как по своей внутренней логике, так и по способу вычисления.

Соотношение между ними выражается следующим неравенством:

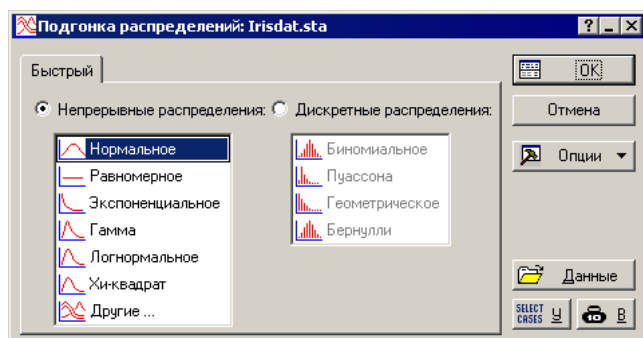
$$-1 \leq 3 * \text{тау Кендалла} - 2 * R \text{ Спирмена} \leq 1$$

Более важно, что **тау Кендалла** и **R Спирмена** по-разному интерпретируются. **R Спирмена** можно осмысливать как прямой аналог коэффициента корреляции Пирсона, но вычисленный по рангам (а не по исходным наблюдениям). **Тау Кендалла** представляет вероятность, точнее, разность между вероятностью того, что значения двух переменных располагаются в одном и том же порядке, и вероятностью 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 **тау Кендалла** в тех случаях, когда в данных имеется много совпадающих значений.

С точки зрения основных предположений, **статистика Гамма** эквивалентна статистике **R Спирмена** или **тау Кендалла**.

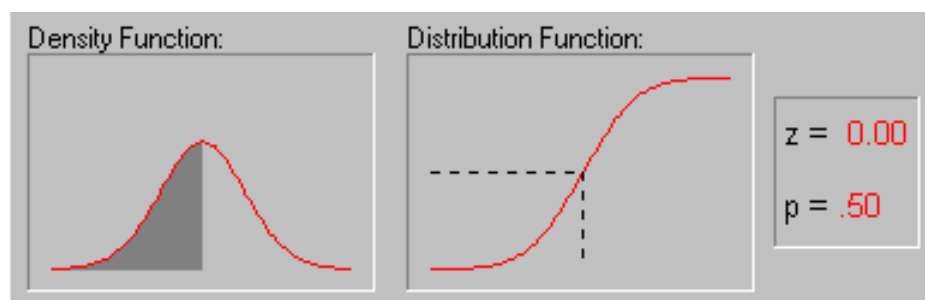
Модуль Подгонка распределений



Общая цель

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 человеческой популяции.



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.

Другим обычным приложением процедуры подгонки распределения является проверка гипотезы нормальности до того, как использовать какой-либо параметрический тест. Например, вы можете использовать критерий нормальности Колмогорова-Смирнова или критерий Шапиро-Уилка W для проверки нормальности.

Подгонка распределениями

Для прогноза часто необходимо знать закон распределения. Для определения закона, в большинстве случаев нужно подогнать наблюдаемое распределение под теоретическое. Это можно осуществить сравнением наблюдаемых частот в данных с ожидаемыми частотами в теоретическом распределении (т. е. используя критерий согласия *хи-квадрат*).

В некоторых исследовательских проектах можно сформулировать гипотезы относительно распределения рассматриваемой переменной. Например, переменные, чьи значения определены бесконечным числом независимых случайных событий, распределены нормально. Переменные, чьи значения являются результатом экстремально редкого события, имеют пуассоновское распределение.

Наиболее важны следующие семейства распределений, которые используются для описания продолжительности жизни или наработки до отказа: экспоненциальное (в том числе, линейно экспоненциальное), распределение Вейбулла экстремальных значений и распределение Гомперца.

Основные типы распределений

Нормальное распределение. Нормальное распределение (так называемая колоколообразная кривая, симметричная относительно среднего) - это теоретическая функция, наиболее часто используемая в статистике. Нормальное распределение дает хорошую модель при выполнении следующих условий:

1. данные с большой вероятностью принимают значения, близкие к центральному;
2. положительные и отрицательные отклонения от 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, болезни и т. д.). Таким образом, можно ожидать нормальное распределение для веса в популяции всех людей.

Плотность нормального распределения имеет вид:

$$f(x) = 1/[(2*\pi)^{1/2}*\sigma] * e^{-1/2*[(x-\mu)/\sigma]^2}, -\infty < x < \infty,$$

где μ - среднее, σ - стандартное отклонение, e - основание натурального логарифма, иногда называемое числом Эйлера (2.71...), π - константа Pi (3.14...).

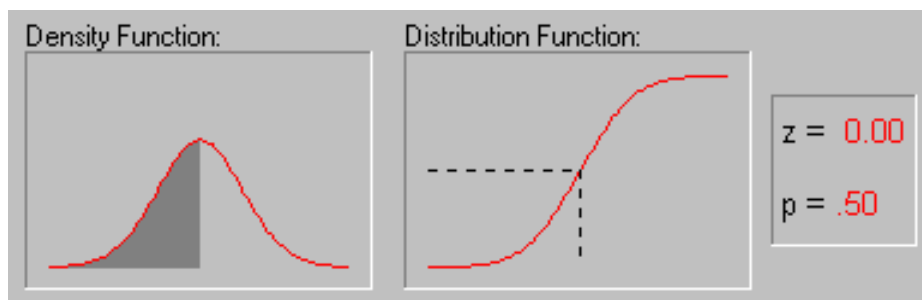


Рисунок выше показывает стандартное нормальное распределение (т. е. нормальное распределение со средним 0 и стандартным отклонением 1), которое часто используется в проверке гипотез.

Равномерное распределение. Равномерное распределение полезно для описания переменных, у которых значения плотности одинаковы на отрезке $[a;b]$, $a < b$.

$$f(x) = \begin{cases} \frac{1}{b-a}, & a < x < b \\ 0, & \text{иначе} \end{cases}$$

где $a < b$ - константы.

Экспоненциальное распределение. Если T - время между наступлениями редких событий, происходящих в среднем с интенсивностью λ , то величина T имеет экспоненциальное распределение с параметром λ .

Экспоненциальное распределение часто используется для описания интервалов времени между последовательными случайными событиями (например, интервалов времени между автомобилями, пересекающими перекресток, длительностей жизни электронных приборов или приходов покупателей в кассу бакалейного магазина).

Плотность экспоненциального распределения имеет вид:

$$f(x) = \lambda * e^{-\lambda x}, 0 \leq x < \infty, \lambda > 0,$$

где λ – параметр экспоненциальной функции (альтернативная параметризация имеет вид $b = 1/\lambda$), e - основание натурального логарифма, иногда называемое числом Эйлера (2.71...).

Гамма-распределение. Плотность экспоненциального распределения имеет моду 0. Во многих примерах *заранее* известно, что мода рассматриваемой случайной переменной не равна 0 (например, когда моделируется распределение длительности жизни электрических ламп). В этих случаях для описания распределения более подходит гамма-распределение.

Плотность гамма-распределения имеет вид:

$$f(x) = \{1/[b\Gamma(c)]\} * [x/b]^{c-1} * e^{-x/b}, 0 \leq x, c > 0,$$

где Γ – функция, c - параметр *формы*, b - параметр *масштаба*, e - основание натурального логарифма, иногда называемое числом Эйлера (2.71...).

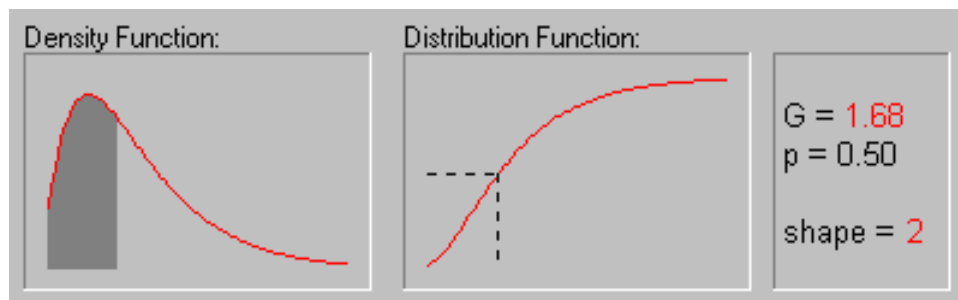


Рисунок выше показывает гамма-распределение, где параметр формы равен 2.

Логнормальное распределение. Логнормальное распределение часто используется в моделировании таких переменных, как персональные доходы, возраст новобрачных (точнее, первый раз вступающих в брак) или допустимое отклонение от стандарта вредных веществ в продуктах питания. В общем, если величина x имеет нормальное распределение, то величина $y = e^x$ имеет логнормальное распределение.

Плотность логнормального распределения имеет вид:

$$f(x) = 1/[x\sigma(2)^{1/2}] * e^{-[\log(x)-\mu]**2/2\sigma**2}, 0 < x < \infty, \mu \geq 0, \sigma > 0,$$

где μ – параметр масштаба, σ - параметр формы, e - основание натурального логарифма, иногда называемое числом Эйлера (2.71...).

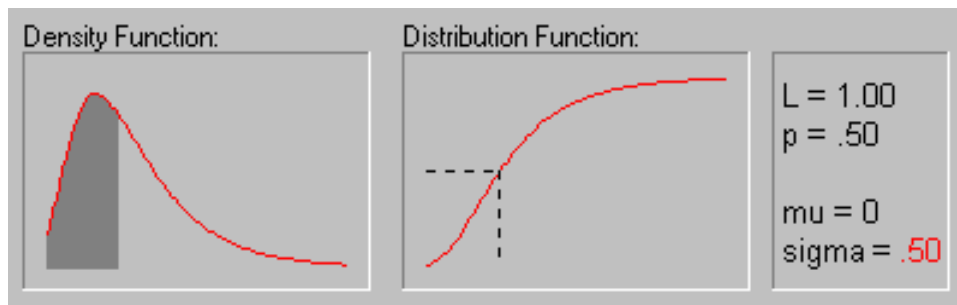


Рисунок выше показывает логнормальное распределение с $\mu = 0$ и $\sigma = .50$.

Хи-квадрат распределение. Распределение суммы квадратов ν независимых случайных переменных, каждая из которых имеет стандартное нормальное распределение, имеет следующий вид:

$$f(x) = \{1/[2^{\nu/2} * \Gamma(\nu/2)]\} * [x^{(\nu/2)-1} * e^{-x/2}], \nu = 1, 2, \dots, 0 < x,$$

где ν - степень свободы, e - основание натурального логарифма, иногда называемое числом Эйлера (2.71...), Γ - гамма-функция.

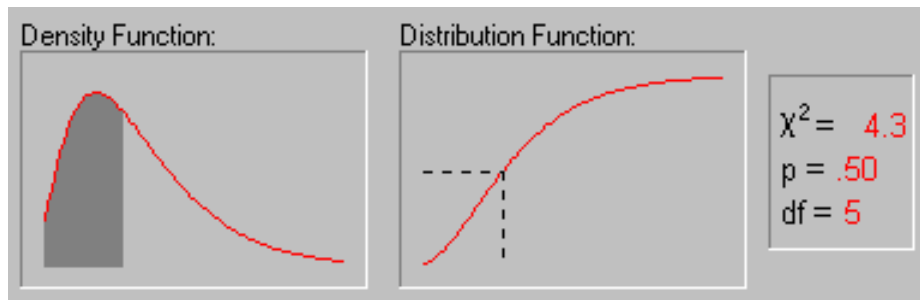


Рисунок показывает распределение *хи-квадрат* со степенью свободы 5.

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:

$$f(x) = [n!/(x!*(n-x)!)] * p^x * q^{n-x}, x = 0, 1, 2, \dots, n,$$

где p – вероятность наступления события, $q = 1 - p$, n – число независимых испытаний.

Распределение Пуассона. Распределение Пуассона иногда называют распределением редких событий. Примеры переменных, распределенных по закону Пуассона, дают число несчастных случаев, число фатальных дефектов в производственном процессе.

Распределение Пуассона определяется формулой:

$$f(x) = (\lambda^x * e^{-\lambda})/x!, \quad x = 0, 1, 2, \dots \text{ и } 0 < \lambda,$$

где λ (лямбда) - ожидаемое значение x (среднее), e - основание натурального логарифма, иногда называемое числом Эйлера (2.71...).

Геометрическое распределение. Если проводятся независимые испытания Бернулли и подсчитывается количество испытаний до наступления успеха, то это число имеет геометрическое распределение.

Геометрическое распределение определяется формулой:

$$f(x) = p * (1-p)^x, \quad x = 1, 2, \dots,$$

где p - вероятность того, что определенное событие (например, успех) произойдет.

Распределение Бернулли. Это распределение наилучшим образом описывает ситуации, где испытание имеет результатом успех либо неуспех. Например, при бросании монеты, при моделировании удачной или неудачной хирургической операции.

Распределение Бернулли определяется формулой:

$$f(x) = p^x * (1-p)^{1-x}, \quad x \in \{0, 1\},$$

где p - вероятность того, что определенное событие (например, успех) произойдет.

2

ГЛАВА

ПОШАГОВЫЕ ПРИМЕРЫ

Пример 1: Корреляции (простой пример).....45

Пример 2: Дисперсионный анализ (ДА) (более сложный пример)68



STATISTICA - Adstudy

Файл Правка Вид Вставка Формат Анализ Графика Сервис Данные Окно Справка

Данные: Adstudy (25v * 50)

Парные и частные корреляции: Adstudy

Квадратная матрица Прямоугольная матрица

Первый список: нет
Второй список: нет

Парные Частные корреляции Опции

Формат отображения корреляционных матриц

- Выделять значимые корреляции
- Отображать р-уровень и N
- Подробная таблица результатов

Отображать длинные имена переменных

Вычисления с повышенной точностью

Уровень значим. для выделения: 0.05

Опции

SELECT CASES

Вешенные моменты

Ст. свободы

В-1 N-1

Удаление П.Д.

Построчное Попарное

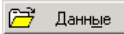

	1	2	3	4	5	6	7	8	9
Advertising									
GENDER	MALE								
R. Rafuse	MALE								
T. Leiker	MALE								
E. Bizot	FEMALE								
K. French	MALE								
E. Van Landuyt	MALE								
K. Harrell	FEMALE								
W. Noren	FEMALE								
S. Willden	MALE								
S. Kohut	FEMALE								
B. Madden	MALE								
M. Bowling	FEMALE								
J. Willcoxson	MALE								
J. Landrum	MALE								
M. Taylor	MALE								
N.S. Madden	FEMALE								
K. Ridgway	FEMALE	PEPSI	3	2	5	4			
L. Cunha	MALE	COKE	2	9	9	3	1	4	
F. Wind	FEMALE	PEPSI	1	0	7	5	2	4	

Добавить документ в отчет

H1,П1 MALE Выбор:OFF Вес:OFF CAP NUM REC



Настройка системы STATISTICA. Практически все параметры интерфейса *STATISTICA* могут быть настроены по вашему желанию. Например, вы можете настроить запуск системы *STATISTICA*, в частности, отменить заданный по умолчанию полноэкранный режим работы, убрать панели инструментов и т.д.

Выбор файла данных. Для этого примера нам потребуется файл **Adstudy.sta**. Чтобы открыть его, зайдите в папку, в которой располагается *STATISTICA*, и выберите папку Examples\Datasets. Вы можете открыть файл с помощью меню **Файл - Открыть**, кнопки  на любой стартовой панели, или кнопки  панели инструментов.


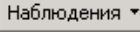
Электронные таблицы (мультимедийные таблицы). Файлы данных *STATISTICA* всегда отображаются в электронных таблицах (т.е. одна электронная таблица соответствует одному файлу данных).

Для отображения всех электронных таблиц в системе *STATISTICA* используется технология мультимедийных таблиц StatSoft, которая будет пояснена ниже. Благодаря этой технологии электронные таблицы могут содержать не только данные, практически неограниченные по объему, но также звук, видео, автоматически выполняемые сценарии ваших действий и различные пользовательские интерфейсы.

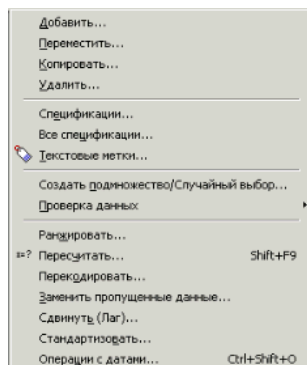
Вы можете открыть одновременно несколько электронных таблиц (при этом каждая электронная таблица будет связана со своей процедурой анализа).

Большую часть выводимых данных система *STATISTICA* также отображает в электронных таблицах (мультимедийные таблицы).

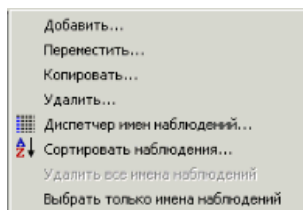
Заметим, если электронная таблица открыта, то становятся доступными опции управления данными в меню **Данные**.

На панели инструментов **Электронная таблица** расположены кнопки  (Переменные) и  (Наблюдения). При нажатии на эти кнопки появляются меню с перечнем команд для изменения файла данных (например, **Добавить**, **Переместить**, **Сдвинуть** переменные).

Меню **Переменные** ▾:

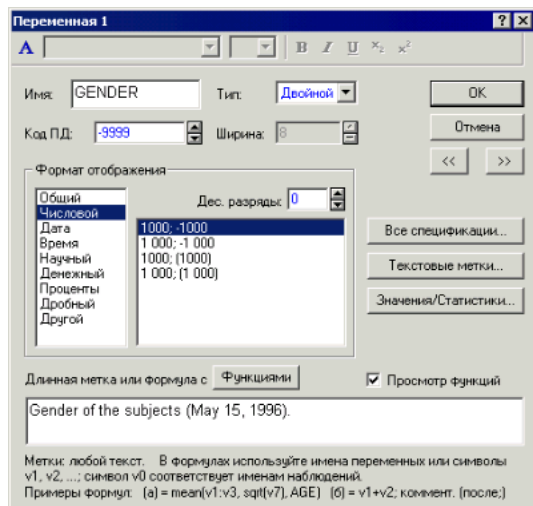


Меню **Наблюдения** ▾:



Все представленные выше опции описаны в *Электронном руководстве*.

Спецификации переменной. Заголовки столбцов в электронной таблице содержат имена переменных. Двойной щелчок на имени переменной открывает диалоговое окно спецификаций переменной **Переменная**.

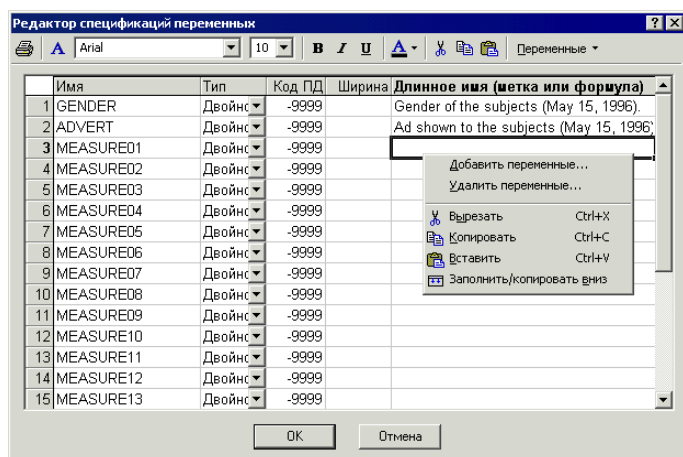


Формулы электронных таблиц. В этом окне можно изменить формат, имя переменной, ввести формулу для пересчета ее значений и т.д.

Если текст в поле **Длинная метка или формула с Функциями** начинается с символа равенства (=), то *STATISTICA* будет воспринимать его как формулу. При этом комментарии могут следовать за точкой с запятой (;).

Например, в поле **Длинная метка или формула с Функциями** можно ввести выражение $=(v2+v3+v4)/3$ или $=mean(v2:v4)$. Тогда для каждого наблюдения (строки) электронной таблицы текущие значения переменной заменятся на среднее значений второй, третьей и четвертой переменных.

Спецификации всех переменных можно просмотреть и отредактировать в “комбинированном” диалоговом окне **Редактор спецификаций переменных**, которое открывается с помощью кнопки **Все спецификации** в окне **Переменная**.



Быстрые меню в электронных таблицах. Полезной особенностью электронных таблиц является возможность вызова контекстных меню. Контекстные меню – это меню, которые вызываются нажатием правой кнопки мыши на соответствующем объекте (например, ячейке электронной таблицы).

Контекстное меню электронной таблицы содержит набор специальных операций по управлению данными и другие команды, связанные с текущей переменной (столбцом), наблюдением (строкой) и/или блоком ячеек.

The screenshot shows a window titled "Данные: Adstudy (25v * 50c)" containing a table with the following data:

	1	2	3	4	5	6
	GENDER	ADVERT	MEASURE01	MEASURE02	MEASURE03	MEASURE04
R. Rafuse	MALE	PEPSI	9	1	6	8
T. Leiker	MALE	COKE	6	7	1	8
E. Bizot	FEMALE	COKE	9	8	2	9
K. French	MALE	PEPSI	7			
E. Van Landuyt	MALE	PEPSI	7			
K. Harrell	FEMALE	COKE	6			
W. Noren	FEMALE	COKE	7			
W. Willden	MALE	PEPSI	9			
S. Kohut	FEMALE	PEPSI	7			
B. Madden	MALE	PEPSI	6			
M. Bowling	FEMALE	PEPSI	4			
J. Willcoxson	MALE	COKE	7			
J. Landrum	MALE	PEPSI	6			
M. Taylor	MALE	COKE	7			
N.S. Madden	FEMALE	PEPSI	6			
K. Ridgway	FEMALE	PEPSI	3			
L. Cunha	MALE	COKE	2			
F. Wind	FEMALE	PEPSI	1			
K. Judkasikam	FEMALE	COKE	0	6	2	3
B. Brinker	MALE	COKE	6	8	1	9
U. Kasetsart	MALE	PEPSI	9	2	7	7

A context menu is open over the table, showing options such as "Вырезать", "Копировать", "Вставить", and "Выделение ячеек".

Три способа управления выводом. Вы можете задать способ вывода результатов в системе *STATISTICA* (см. *Три канала вывода результатов анализа*, стр. 101). Когда вы проводите анализ, *STATISTICA* выводит результаты в виде мультимедийных таблиц (электронных таблиц) и графиков. Существует три основных канала вывода, в которые вы можете направить все результаты:

- Рабочие книги (см. стр. 102),
- Отчеты (см. стр. 104),
- Автономные окна (см. стр. 106).

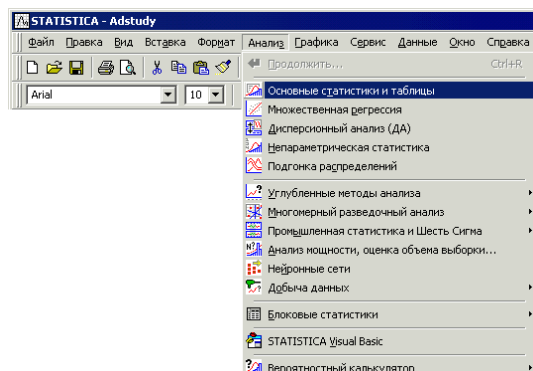
Эти три канала вывода могут использоваться в различных сочетаниях (например, могут быть выбраны одновременно Рабочая книга и Отчет), и каждый канал вывода может быть настроен различным образом.

Кроме того, все объекты вывода (электронные таблицы и графики) могут содержать другие внедренные и связанные объекты и документы.

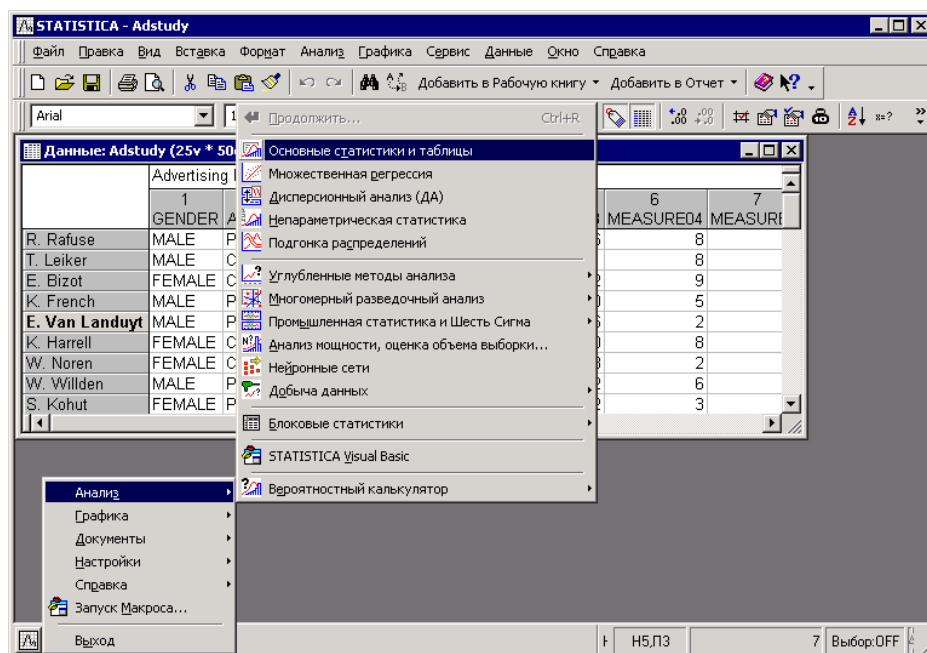
Таким образом, *STATISTICA* предоставляет возможность иерархической организации вывода.

Управление выводом результатов осуществляется с помощью вкладки **Диспетчер вывода** диалогового окна **Параметры** (доступного из меню **Файл - Диспетчер вывода** или **Сервис - Параметры**).

Вычисление матрицы корреляций. Вычислим матрицу корреляций для переменных из файла данных. Вначале вызовите стартовую панель **Основные статистики и таблицы**. Для этого выберите команду **Основные статистики и таблицы** в меню **Анализ**



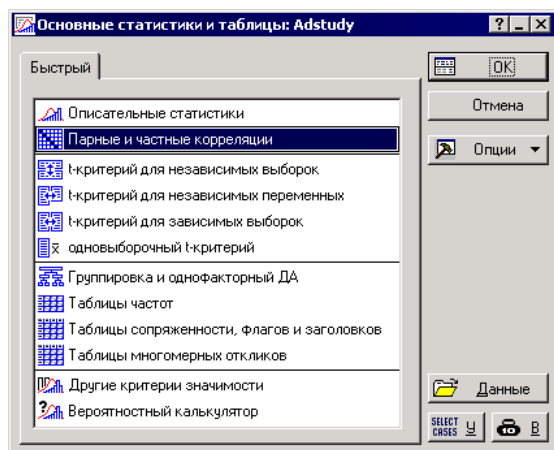
или нажмите кнопку  в нижнем левом углу экрана.



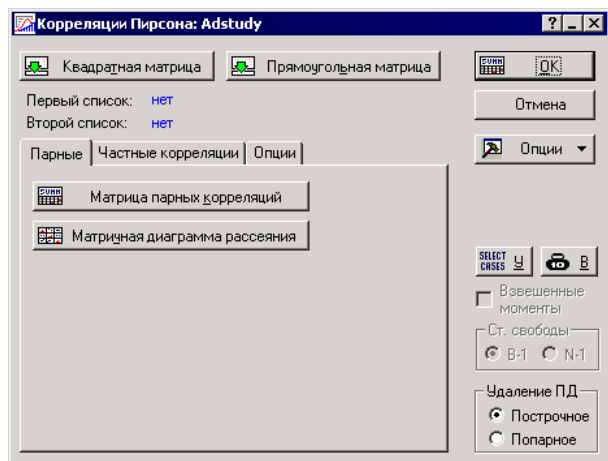
При вызове стартовой панели убедитесь, что в электронной таблице не выделен блок, т.е. группа ячеек (чтобы отменить выделение блока, щелкните мышкой на любой другой ячейке в электронной таблице).

Если блок выделен, *STATISTICA* будет проводить анализ именно этого блока и после нажатия кнопки **OK** или **Итоги** вместо того, чтобы попросить выбрать переменную, выведет корреляции для блока переменных.

На Стартовой панели **Основные статистики и таблицы**

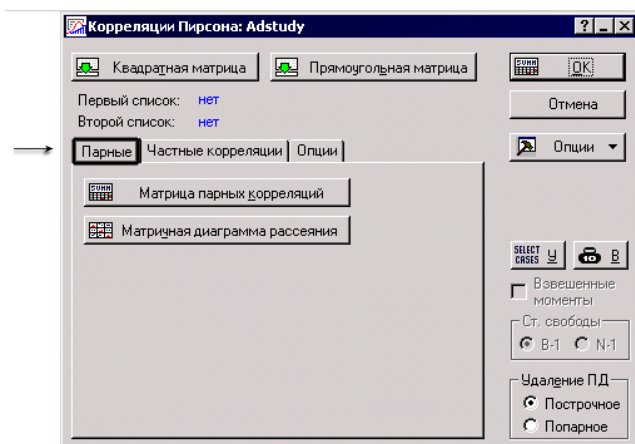


выберите двойным щелчком пункт **Парные и частные корреляции** (или выберите этот пункт и просто нажмите кнопку **OK**). Появится диалоговое окно **Корреляции Пирсона**.



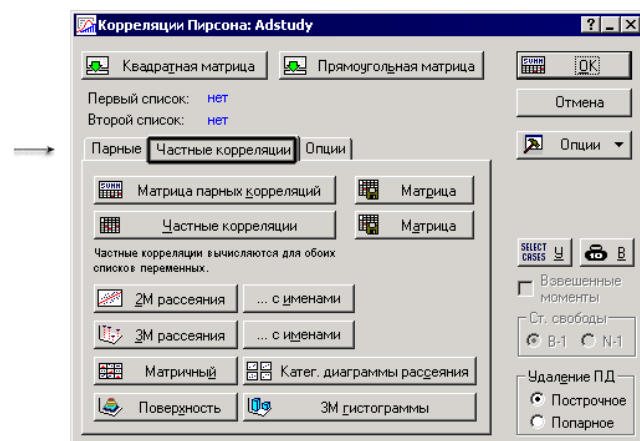
Быстрый и углубленный анализ. Как и большинство диалоговых окон анализа, окно **Корреляции Пирсона** содержит набор опций. Как правило, предлагается не менее двух типов анализа.

Вкладка **Парные** этого окна содержит опции, позволяющие вычислить парные корреляции нескольких переменных и отобразить результаты в различных форматах, например, вывести уровень значимости и т.д.



Вкладка **Частные корреляции** содержит опции для вычисления как парных, так и частных корреляций, а также опции сохранения матриц, получения некоторых статистик или графиков. (См. также книгу “Искусство анализа данных на компьютере”, [2].)

В зависимости от выбранного анализа могут быть доступны также дополнительные вкладки.



Заметим, что в некоторых случаях доступна только вкладка **Быстрый**. Вы можете ознакомиться с описанием опций выбранной вкладки, нажав клавишу F1 или кнопку **?** в верхнем правом углу диалогового окна.

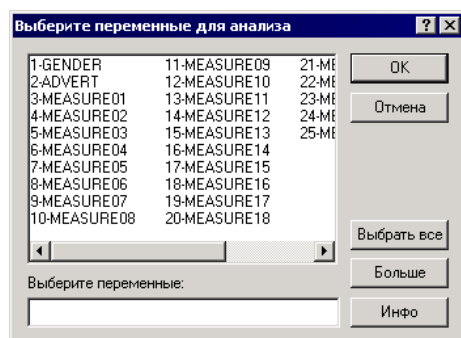
Система “контекстной подсказки”. Все диалоговые окна системы *STATISTICA* поддерживают “контекстную подсказку”. Если вы не знаете, что делать дальше, просто нажмите **OK** или кнопку **Итоги**, и программа сама сделает следующий логический шаг. При этом, если какой-либо этап был пропущен, программа попросит вас ввести недостающую информацию (например, переменные для анализа).

Кнопка Переменные. Диалоговое окно *Корреляции Пирсона* является стандартным диалоговым окном анализа. Оно содержит возможности для выбора переменных, которые подлежат анализу, а также предлагает различные варианты получения итоговых статистик.

Каждое диалоговое окно статистического анализа в системе *STATISTICA* содержит, по крайней мере, одну кнопку **Переменные**, с помощью которой можно выбрать нужную переменную. Для этого нужно нажать данную кнопку.

Диалоговое окно выбора переменной. После нажатия кнопки **Квадратная матрица** (или кнопки **OK**, если вы еще не выбрали переменные) появится диалоговое окно выбора переменной.

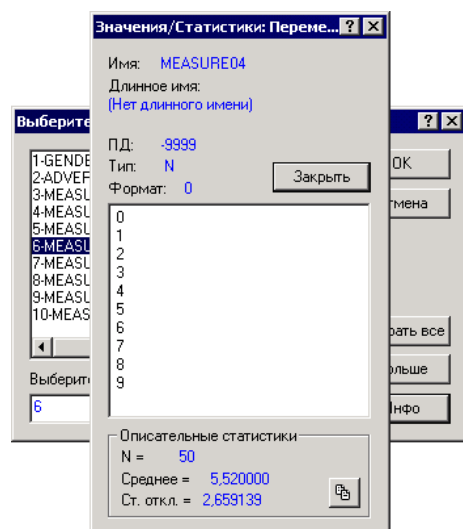
(Еще раз обратите внимание, что при “случайном” выделении блока в электронной таблице переменные этого блока будут выбраны автоматически, и при нажатии кнопки **OK** по умолчанию будет вычислена матрица корреляций для выделенного блока переменных.)



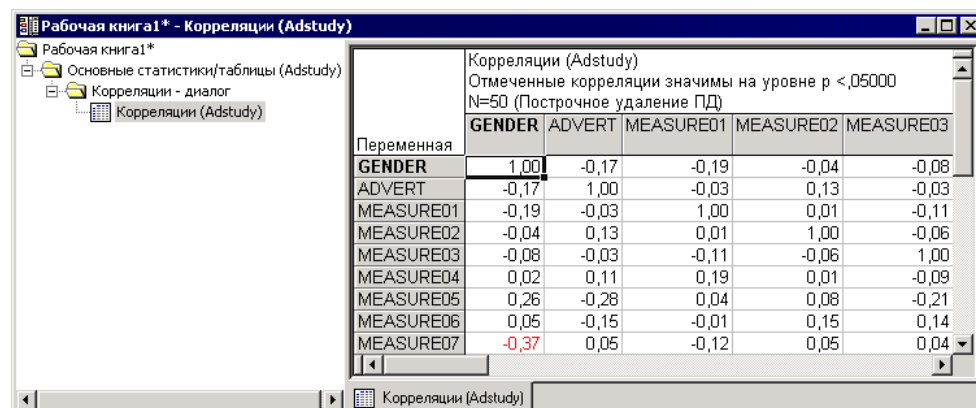
Диалоговое окно выбора переменной поддерживает разные способы выделения переменных (включая стандартные Windows соглашения о выделении при нажатых клавишах SHIFT и CTRL произвольных или непрерывных списков переменных соответственно).

Это диалоговое окно также предлагает различные варианты просмотра содержания файла данных. Например, нажав кнопку **Больше**, вы можете развернуть список

переменных, чтобы посмотреть длинные метки или формулы. Нажав кнопку **Инфо**, вы можете просмотреть отсортированный список всех значений и описательных статистик переменной.



В данном примере нажмите кнопку **Выбрать все**, а затем **ОК**. Вы вернетесь в диалоговое окно **Корреляции Пирсона**. Щелкните на кнопке **ОК** для получения матрицы корреляций выбранных переменных.




Заметим, что вместо кнопки **ОК** вы можете нажать **Матрица парных корреляций**. Также, в зависимости от ваших установок для вывода, таблица корреляций может быть добавлена в Отчет или в Автономное окно, как это было показано выше.

Таблицы результатов (мультимедийные таблицы). Кроме хранения данных, электронные таблицы используются в системе *STATISTICA* для вывода большинства численных оценок, полученных в ходе анализа данных.

Заметим, что имеется широкий набор средств для работы с электронными таблицами. Так, в нашем примере значимые корреляции выделяются цветом (по умолчанию - красным). Электронные таблицы могут содержать от одной строки до нескольких гигабайт и предлагают множество процедур для удобного просмотра результатов и их визуализации в виде стандартных или определенных пользователем графиков.

Электронные таблицы *STATISTICA* используют технологию мультимедийных таблиц StatSoft. Термин “мультимедийная” означает, что таблица может содержать не только данные практически неограниченного объема, но также звук, видео, автоматически выполняемые сценарии и различные пользовательские интерфейсы (см. стр. 113).

Параметры электронных таблиц. Как было отмечено ранее, большинство функций электронной таблицы доступно на панели инструментов **Таблица данных** и в контекстных меню (появляющихся при нажатии правой кнопки мыши на любой ячейке таблицы). Чтобы просмотреть описание этих функций, нажмите клавишу (F1) или кнопку  на панели инструментов.

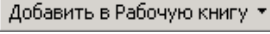
Вы можете изменить все аспекты формата отображения каждого столбца, отредактировать результаты, добавить пустые строки и столбцы для заметок или для результатов, полученных из других источников. Предусмотрены различные способы вывода таблиц результатов на печать (по умолчанию они печатаются в виде таблицы с высоким качеством и разрешением).

Кроме того, вы легко можете использовать в своем анализе результаты предыдущих исследований (например, вы можете использовать нашу матрицу корреляций для многомерного шкалирования). Для использования таблицы результатов в качестве исходной таблицы сделайте ее активной и выберите команду **Исходная таблица данных** в меню **Данные**.

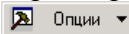
Рабочие книги анализа и другие параметры вывода. Все результаты могут быть отображены (и сохранены) в Автономных окнах, Отчетах или Рабочих книгах. Рабочие книги представляют собой принятый по умолчанию (и, возможно, наиболее универсальный) способ управления выводом из анализа (см. стр. 102 и 109).

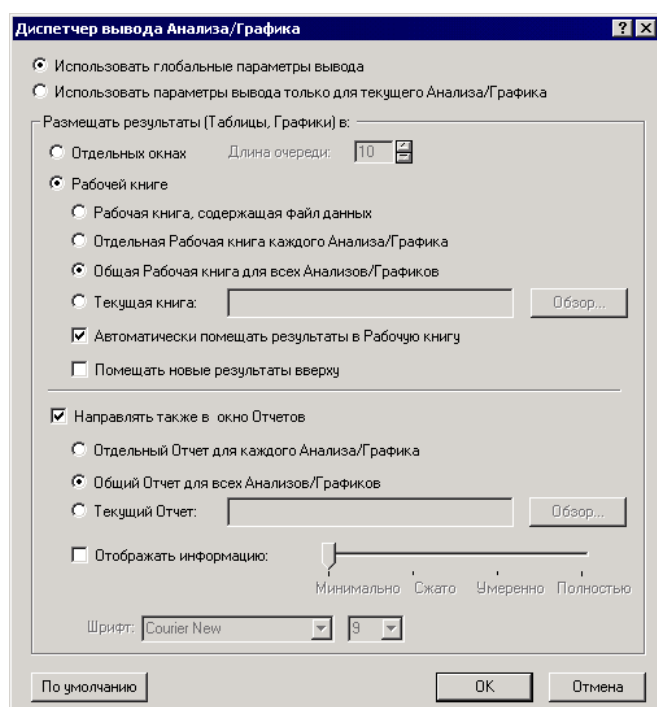
В зависимости от ваших установок в окне **Диспетчер вывода** (доступном через меню **Файл - Диспетчер вывода**, см. следующий параграф), результаты могут быть помещены в простую Рабочую книгу, которая содержит результаты всех видов анализа, или в

различные Рабочие книги, каждая из которых содержит результаты (электронные таблицы и графики) отдельного вида анализа.

Вы можете выбрать режим автоматического помещения результатов в Рабочую книгу. Но также вы можете добавлять электронные таблицы или графические окна в Рабочую книгу с помощью кнопки .

Диспетчер вывода. В зависимости от того, как вы хотите сохранять данные и результаты анализа, вы можете воспользоваться Рабочими книгами разных типов.

Чтобы изменить параметры представления результатов только для текущего анализа, нажмите кнопку  в диалоговом окне задания анализа или графика и в появившемся меню выберите пункт **Диспетчер вывода**. Откроется диалоговое окно **Диспетчер вывода Анализа/Графика**.



Если вы хотите изменить параметры вывода для всех видов анализа, выберите опцию **Использовать глобальные параметры вывода** в диалоговом окне **Диспетчер вывода Анализа/Графика** или же используйте общий **Диспетчер вывода**.


В последнем случае выберите вкладку **Диспетчер вывода** диалогового окна **Параметры**, доступного из меню **Файл - Диспетчер вывода** или из меню **Сервис - Параметры**.

Как и все Рабочие книги, отдельные документы (например, электронные таблицы, графики) или группы документов могут быть распечатаны на принтере, вырезаны, скопированы и удалены из Рабочей книги анализа. См. обзор *Рабочие книги STATISTICA* на стр. 109; см. также *Электронное руководство*.

Копирование и Копирование с заголовками. Электронная таблица может быть скопирована в буфер обмена с помощью команды **Копировать**, соответствующей нажатию клавиш CTRL+C. При этом копируется только содержимое выделенного блока. Также можно произвести копирование с помощью команды **Копировать с заголовками** в меню **Правка**. При этом копируется блок вместе с соответствующими именами переменных и наблюдений.

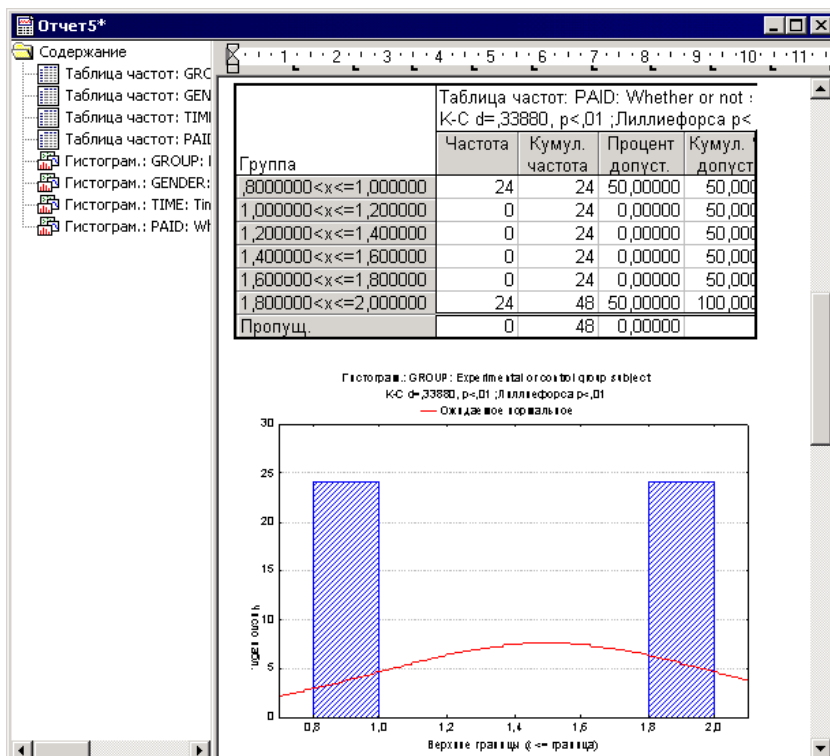
Если вставка осуществляется в документ Word, электронные таблицы представляют собой активные объекты *STATISTICA*, стандартные таблицы формата RTF или текст, разделенный символами табуляции (в зависимости от вашего выбора в диалоговом окне **Специальная вставка** редактора Word).

Печать таблиц данных. Чтобы распечатать электронную таблицу результатов, выберите команду **Печать** в меню **Файл** (или нажмите CTRL+P). Появится диалоговое окно **Печать**, в котором вы можете задать параметры печати.

Вы можете также использовать ускоренный способ, щелкнув кнопку **Печать**  на панели инструментов. При таком способе диалоговое окно **Печать** не отображается, и на печать выводится содержимое текущего документа.

Если вы хотите распечатать документ, содержащийся в Рабочей книге, убедитесь, что нужный документ выделен в Рабочей книге, и установите опцию **Выделенный документ** в диалоговом окне **Печать**. Вы можете также извлечь копию документа из Рабочей книги (переместите его из окна дерева Рабочей книги или используйте меню **Рабочая книга**) и затем распечатать его.


Дополнительные Отчеты для вывода. Рабочие книги предоставляют, возможно, наиболее гибкие возможности для управления выводом (см. стр. 102 и 109). В некоторых случаях полезно ведение автоматического журнала вывода (содержащего все электронные таблицы и/или графики) в виде традиционного Отчета текстового редактора Word, в который вы затем можете добавить комментарии и объекты (см. стр. 104 и стр. 119).



Для создания такого Отчета откройте **Диспетчер вывода** и установите флажок **Направлять результаты в Окно Отчетов**.

Далее выберите одну из опций: **Отдельный Отчет для каждого Анализа/Графика**, **Общий Отчет для всех Анализ/Графиков** или **Текущий Отчет**.

Как было отмечено выше, чтобы открыть **Диспетчер вывода** для общих изменений, выберите одноименную команду в меню **Файл** или команду **Параметры** в меню **Сервис**.

Для локальных изменений нажмите кнопку  **Опции** в любом диалоговом окне спецификаций анализа или графика, а затем выберите **Диспетчер вывода**.

Также с помощью диалога **Диспетчер вывода** вы можете изменять объем информации, отображаемой в итоговой электронной таблице.

Интерпретация результатов. Вернемся к нашему примеру и полученной ранее матрице корреляций.

Переменная	GENDER	ADVERT	MEASURE01	MEASURE02	MEASURE03
GENDER	1,00	-0,17	-0,19	-0,04	-0,08
ADVERT	-0,17	1,00	-0,03	0,13	-0,03
MEASURE01	-0,19	-0,03	1,00	0,01	-0,11
MEASURE02	-0,04	0,13	0,01	1,00	-0,06
MEASURE03	-0,08	-0,03	-0,11	-0,06	1,00
MEASURE04	0,02	0,11	0,19	0,01	-0,09
MEASURE05	0,26	-0,28	0,04	0,08	-0,21
MEASURE06	0,05	-0,15	-0,01	0,15	0,14
MEASURE07	-0,37	0,05	-0,12	0,05	0,04

В каждой ячейке корреляционной матрицы находится число в интервале от -1.00 до $+1.00$, показывающее степень зависимости между соответствующими переменными. Например, коэффициент корреляции между **Measure04** и **Measure01** равен $0,19$.

Напомним основные сведения корреляционного анализа (см. также книгу “Искусство анализа данных на компьютере”, [2]).

Чем больше по модулю значение коэффициента корреляции, тем сильнее зависимость; если значение положительно, зависимость “положительна” (большие значения одной переменной соответствуют большим значениям другой переменной, а меньшие значения одной переменной соответствуют меньшим значениям другой переменной).

Если значение отрицательно, то верно обратное (меньшие значения одной переменной соответствуют большим значениям другой переменной).

Если значение коэффициента корреляции равно 0 , то связь отсутствует. При значении коэффициента корреляции, равном $+1$ или -1 , имеется *линейная* связь между рассматриваемыми переменными (положительная или отрицательная). Иными словами, если рассматриваются переменные X и Y , коэффициент корреляции между которыми равен $+1$, то $Y = aX + b$, где a, b – константы, причем $a > 0$. Если коэффициент корреляции между переменными X и Y равен -1 , то $Y = aX + b$, где a, b – константы, причем $a < 0$.

Обратим внимание на *шкалу Чеддока*, которая позволяет дать качественную оценку значениям корреляций (в таблице коэффициенты берутся по абсолютной величине).

Шкала Чеддока имеет следующий вид:

Показания тесноты связи	0,1-0,3	0,3-0,5	0,5-0,7	0,7-0,9	0,9-0,99
Сила связи	слабая	умеренная	заметная	высокая	весьма высокая

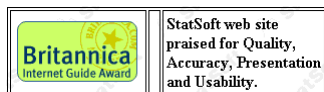
При значениях показателей тесноты связи, превышающих 0,7, зависимость является *высокой*, а при значениях более 0,9 – *весьма высокой*. Это означает, что более половины общей вариации объясняется влиянием изучаемого фактора. Последнее позволяет считать оправданным применение метода функционального анализа для изучения корреляционной связи, а построенные при этом математические модели, как правило, признаются пригодными для их практического использования.

При показателях тесноты связи ниже 0,7 коэффициент детерминации R^2 всегда будет меньше 0.5.

Заметим, что имеются так называемые непараметрические или ранговые корреляции (см. книгу “Искусство анализа данных на компьютере”, [2]).

Для получения более подробной информации о том, как интерпретировать коэффициенты корреляции, см. соответствующие разделы *Электронного руководства*.

Электронное руководство включает в себя полное содержание *Электронного учебника* компании StatSoft. Этот статистический ресурс был отмечен наградами информационным порталом *Encyclopedia Britannica*



за свое “высокое качество, точность, удобное преподнесение информации и практичность”.

Воспользуйтесь также книгой Боровикова В. П. “Программа *STATISTICA* для студентов и инженеров”, где популярно объяснены основные понятия анализа данных и дана их интерпретация.

Вы можете также обратиться к **Статистическому советнику** (см. стр. 67).

Для запуска *Электронного руководства* выберите пункт **Справка по STATISTICA** в меню **Справка**. Затем введите соответствующий термин (например, **Корреляции**) в поле **Искать следующие слова** вкладки **Поиск** *Электронного руководства*.

Нажмите кнопку **Показать** и выберите нужную тему в поле **Раздел** (в нашем случае **Корреляции - Вводный обзор**):

Корреляции - Вводный обзор

Корреляция представляет собой меру зависимости переменных. Наиболее известна корреляция Пирсона. При вычислении корреляции Пирсона предполагается, что переменные измерены, как минимум, в интервальной шкале. Некоторые другие коэффициенты корреляции (например, корреляция Спирмена или тетракорическая корреляция) могут быть вычислены для менее информативных шкал; эти корреляции доступны в других модулях. Коэффициенты корреляции изменяются в пределах от -1.00 до +1.00. Обратите внимание на крайние значения коэффициента корреляции. Значение корреляции -1.00 означает, что переменные связаны строгой линейной зависимостью с отрицательным коэффициентом наклона. Значение +1.00 означает, что переменные связаны строгой линейной зависимостью с положительным коэффициентом наклона. Отметим, что значение 0.00 означает отсутствие корреляции.

CORRS.stg

MEASURE1 vs. MEASURE2

GROUP: 1 $r = -.90$

GROUP: 2 $r = -.90$

GROUP: 3 $r = .00$

GROUP: 4 $r = .40$

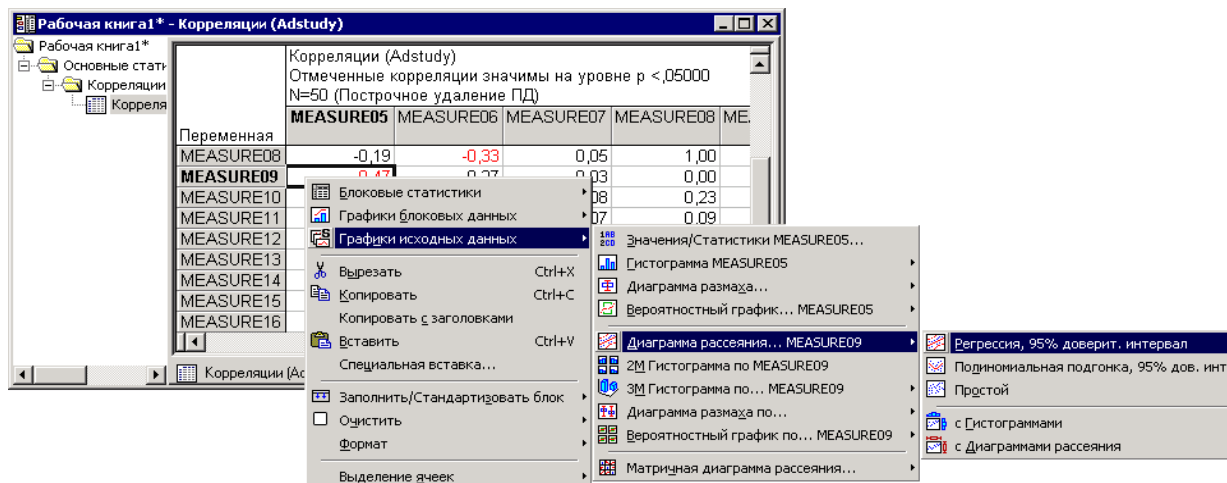
Одним из важных вопросов (которому часто не придают значения), обсуждаемым в *Электронном руководстве*, является использование диаграмм рассеивания для анализа корреляций.

Например, очень большой и статистически значимый коэффициент корреляции может получиться из-за наличия резко выделяющихся наблюдений (“выбросов”). В таком случае коэффициент корреляции (даже статистически значимый) не будет представлять интерес (например, не будет иметь “прогнозной силы”).

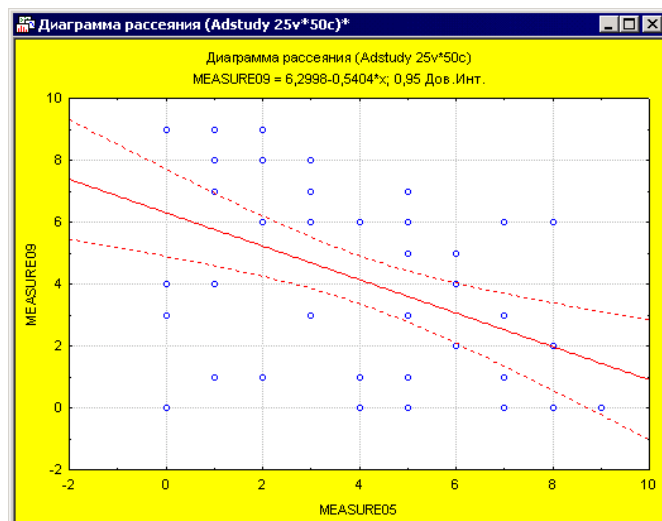
Следуя этому замечанию, а также советнику *Электронного статистического учебника*, рекомендуем вам всегда обращаться к диаграмме рассеивания, чтобы увидеть наглядное представление зависимости между переменными и понять природу коэффициентов частной корреляции.

Построение графиков для электронных таблиц. Вернемся к примеру. Вы можете наглядно представить корреляционную зависимость, например, между переменными *Measure09* и *Measure05*, с помощью диаграммы рассеивания.

Чтобы построить диаграмму рассеяния для этих двух переменных, щелкните правой кнопкой мыши по соответствующему коэффициенту корреляции (**-0,47**). В появившемся контекстном меню выберите пункт **Графики исходных данных**, а далее нужный тип графика, как показано ниже.

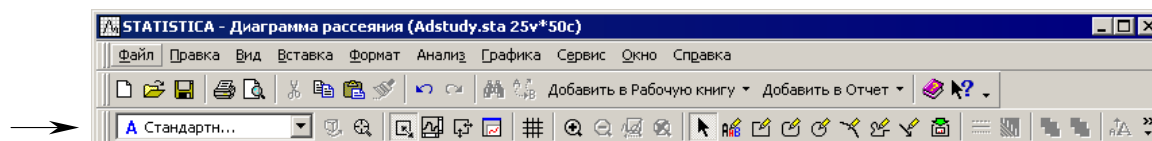


Требуемый график будет отображен на экране.



Как видно из графика, резко выделяющиеся наблюдения отсутствуют, т. е. мы больше можем не беспокоиться о выбросах (см. предыдущий параграф, а также раздел *Электронного руководства*, посвященный выбросам).

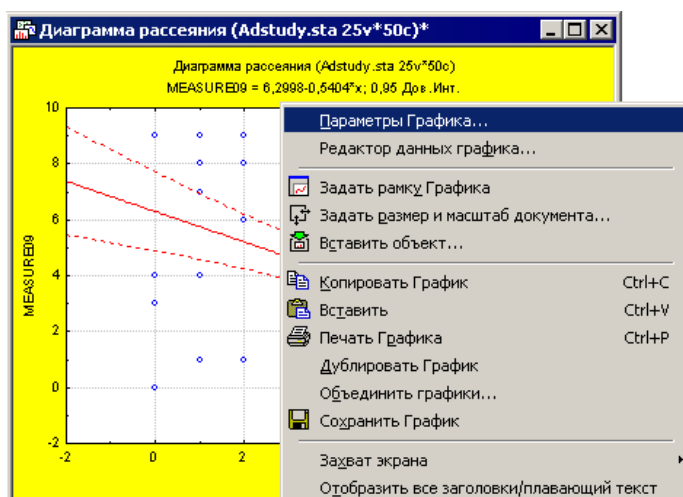
Настройка графика. Заметьте, что при появлении активного графического окна появилась и соответствующая ему панель инструментов. Панель **Графические инструменты** отличается от панелей инструментов электронных таблиц.



Она содержит множество возможностей настройки графика и инструментов для рисования. Все эти возможности также доступны из меню, и многие из них – из контекстных меню, появляющихся при щелчке правой кнопкой мыши на соответствующей части графика.

Заметим, что опции контекстных меню располагаются иерархично. Первые одна или две опции относятся только к выделенному элементу графика, в то время как нижние опции относятся к элементам, которые так или иначе с ним связаны.

Если щелкнуть где-нибудь на фоновой поверхности графического окна вне осей графика, появится меню общих настроек, как показано ниже.



Для получения более подробной информации по настройкам графиков см. стр. 128, а также *Электронное руководство*.

Теперь вернемся к таблице данных.

Разделение прокрутки в электронных таблицах. Электронные таблицы 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 электронной таблицы.

Когда курсор мыши указывает на разделитель, он изменяется и принимает вид двунаправленной стрелки \leftrightarrow или \updownarrow . Теперь можно нажать левую кнопку мыши и переместить разделитель в нужное положение.

	1	2	3	4	5	6
	GENDER	ADVERT	MEASURE01	MEASURE02	MEASURE03	MEASURE04
R. Rafuse	MALE	PEPSI	9	1	6	
T. Leiker	MALE	COKE	6	7	1	
E. Bizot	FEMALE	COKE	9	8	2	
K. French	MALE	PEPSI	7	9	0	
E. Van Landuyt	MALE	PEPSI	7	1	6	
K. Harrell	FEMALE	COKE	6	0	0	
W. Noren	FEMALE	COKE	7	4	3	
W. Willden	MALE	PEPSI	9	9	2	
S. Kohut	FEMALE	PEPSI	7	8	2	
B. Madden	MALE	PEPSI	6	6	2	
M. Bowling	FEMALE	PEPSI	4	6	6	
J. Willcoxson	MALE	COKE	7	3	3	

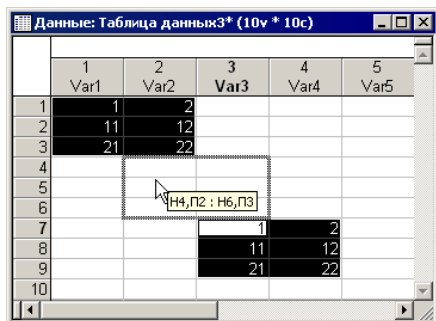
Вы можете легко изменять мышью позицию разделителя.

	1	2	3	4	5	6
	GENDER	ADVERT	MEASURE01	MEASURE02	MEASURE03	MEASURE04
M. Bowling	FEMALE	PEPSI	4	6	6	
J. Willcoxson	MALE	COKE	7	3	3	
J. Landrum	MALE	PEPSI	6	2	3	
M. Taylor	MALE	COKE	7	2	4	
N.S. Madden	FEMALE	PEPSI	6	2	7	
K. Ridowav	FEMALE	PEPSI	3	2	5	
W. Noren	FEMALE	COKE	7	4	3	
W. Willden	MALE	PEPSI	9	9	2	
S. Kohut	FEMALE	PEPSI	7	8	2	
B. Madden	MALE	PEPSI	6	6	2	
M. Bowling	FEMALE	PEPSI	4	6	6	
J. Willcoxson	MALE	COKE	7	3	3	

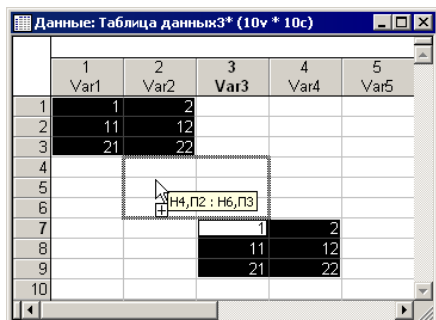
Обратите внимание, что разделенные по вертикали окна прокручиваются синхронно, когда вы прокручиваете таблицу горизонтально, а окна, разделенные по горизонтали, прокручиваются синхронно при вертикальной прокрутке.

Информацию о выделении блоков данных в разделенных окнах, а также об изменении скорости выделения блоков данных можно найти в разделе *Как расширить блок в Таблице данных за пределы экрана?* в *Электронном руководстве*.

“Перетащить и отпустить”. Система *STATISTICA* предлагает полный набор возможностей метода “перетащить и отпустить”, которые доступны при работе со стандартными электронными таблицами (стиля Excel). Например, чтобы переместить выделенный блок, подведите курсор мыши к его границе (курсor мыши превратится из креста в стрелку) и перетащите блок на новое место.



Чтобы скопировать блок, необходимо при перетаскивании копии блока на новое место держать нажатой клавишу CTRL. Заметьте, что при этом рядом с курсором мыши появится знак “плюс” (+). Этот знак информирует о том, что производится именно копирование, а не перемещение (см. рисунок ниже).



Чтобы вставить блок между строками или столбцами, подведите курсор мыши к границе выделения (курсor мыши превратится в стрелку) и перетащите блок, держа нажатой клавишу SHIFT.

Когда курсор мыши указывает на границу между строками, между ними отображается линия вставки, и когда вы отпустите кнопку мыши, блок будет вставлен между этими строками. При этом будут созданы новые наблюдения.

Когда курсор указывает на границу между столбцами, линия вставки отображается между столбцами, и когда вы отпустите кнопку мыши, блок будет вставлен между этими столбцами. При этом будут созданы новые переменные.

Если при перетаскивании блока также нажать клавишу CTRL, то рядом со стрелкой курсора мыши появится плюс (см. рисунок ниже), а вместо перемещения и вставки будет произведено копирование и вставка.

	1 Var1	2 Var2	3 Var3	4 Var4
1	1	2		
2	11	12		
3	21	22		
4				
5				
6				
7				
8				
9				

	1 Var1	2 Var2	3 Var1	4 Var2
1	1	2		
2	11	12		
3	21	22		
4				
5				
6			1	2
7			11	12
8			21	22
9				

Кроме того, можно экстраполировать ряд значений в пределах выделенного блока (режим *Автозаполнение*), перемещая *Ручку заполнения* (маленький черный квадратик, расположенный в нижнем правом углу на границе выделенного блока).

	1 Var1	2 Var2	3 Var3	4 Var4	5 Var5
1	1	MON	JAN	1.500	
2	2	TUE	FEB	1.570	
3	3	WED	MAR	1.530	
4					
5					
6					
7					
8					
9					
10					

	1 Var1	2 Var2	3 Var3	4 Var4	5 Var5
1	1	MON	JAN	1.5	
2	2	TUE	FEB	1.57	
3	3	WED	MAR	1.53	
4	4	THU	APR	1.563	
5	5	FRI	MAY	1.578	
6	6	SAT	JUN	1.593	
7	7	SUN	JUL	1.608	
8					
9					
10					

Электронное руководство. Чтобы получить дополнительную информацию о некоторых функциях системы, нажмите клавишу F1, когда выделена соответствующая команда или пункт меню.

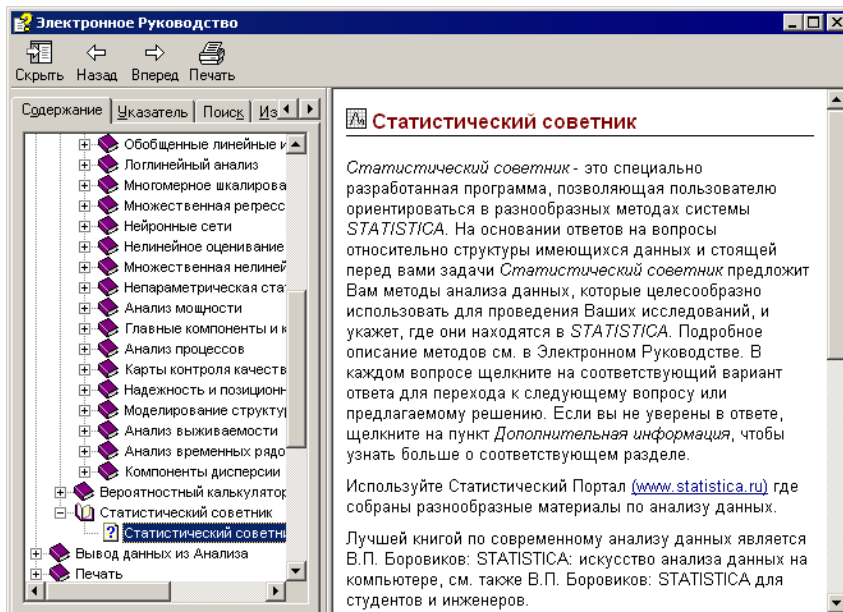
Система *STATISTICA* содержит *Электронное руководство*, предоставляющее справочную информацию по всем процедурам и функциям программы, доступную в контекстно-зависимом режиме. Справка открывается при нажатии клавиши F1 или кнопки справки **?** в строке заголовка всех диалоговых окон. Она содержит свыше 100 мегабайт документации в сжатом виде.

Найти необходимую информацию, пользуясь *Электронным руководством*, можно быстрее, чем просматривая традиционные справочники в напечатанном виде. Это достигается благодаря использованию динамической организации *Электронного руководства* с помощью гиперссылок, организационных вкладок (например, **Содержание**, **Указатель**, **Поиск** и **Избранное**), а также различных возможностей настройки справочной системы.

Заметим также, что в строке состояния в нижней части окна системы *STATISTICA* также отображается краткое пояснение к командам меню или кнопкам панелей инструментов соответственно при выделении пункта меню или нажатии кнопки.

Статистический советник. *Статистический советник* является частью *Электронного руководства*.

После выбора пункта **Статистический советник** в меню **Справка**, система *STATISTICA* задаст вам несложные вопросы о характере стоящей перед вами проблемы и типе исходных данных. Затем советник выдаст список наиболее подходящих для вашей проблемы процедур и объяснит, где их найти в системе *STATISTICA*.



С помощью гиперссылок можно непосредственно перейти из разделов **Статистического советника** к подробному описанию соответствующих статистических методов и процедур в главе *Вводный обзор Электронного учебника StatSoft*.

Пример 2: Дисперсионный анализ (ДА) (более сложный пример)

Цель дисперсионного анализа. Дисперсионный анализ (ДА), изобретенный Фишером в 1920-х годах, интерпретируется как метод разделения влияния на наблюдаемые значения зависимой переменной различных подмножеств параметров (факторов). Важность такого разделения во многих областях исследования делает дисперсионный анализ одним из важнейших и востребованных методов анализа данных.

В экспериментальных исследованиях в качестве параметров часто выступают воздействия некоторых “обработок” зависимой переменной y .

Так, например, при проведении сельскохозяйственных экспериментов (откуда и возникла соответствующая терминология), y может быть урожаем пшеницы или картофеля с участка определенного размера, а изучаемая “обработка” состоит во введении на участке некоторого удобрения в течение сельскохозяйственного сезона.

Естественно, чтобы получить разумные результаты, в эксперимент должны включаться как *обработанные*, так и *необработанные* удобрениями участки. Такого рода эксперимент описывается в терминах *общей линейной модели* посредством введения индикаторной переменной x , которая равна *единице*, если воздействие *присутствует*, и *нулю*, если воздействия нет.

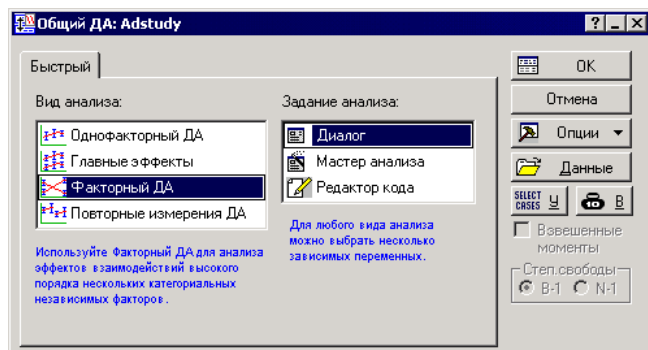
Аналогичная задача возникает в биомедицине, например, при создании и испытании новых лекарств.

Простейшим является случай, называемый обычно *однофакторным анализом* или классификацией по одному признаку, когда наблюдения разбиты на группы в соответствии с возможным различием математических ожиданий.

Ключевым в ДА является F-критерий Фишера и разбиение суммы квадратов $y'u$ на сумму компонент, каждая из которых отвечает подмножеству параметров линейной модели.

F-критерий Фишера является естественным развитием t-критерия Стьюдента на случай более двух выборок или групп, например, мужчины-женщины, здоров-нездоров и т.д. (см., например, книгу “Искусство анализа данных на компьютере”, [2]).

Запуск модуля ДА. Для того чтобы запустить модуль **Дисперсионный анализ**, выберите команду **Дисперсионный анализ (ДА)** в меню **Анализ**. Появится стартовая панель **Общий ДА**.

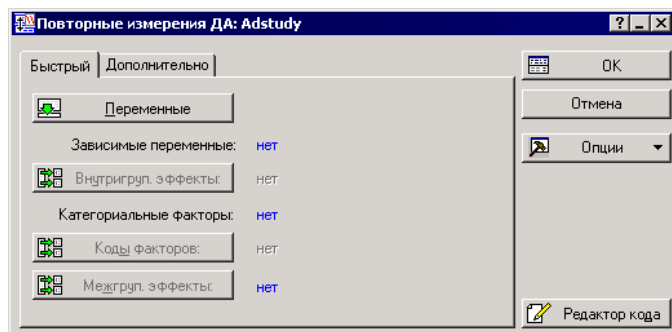


Это диалоговое окно используется для задания простого анализа (например, с помощью **Однофакторного ДА** создаются планы с одним межгрупповым фактором) и более сложных процедур (например, с помощью **Повторных измерений ДА** создаются планы с межгрупповыми и внутренними факторами).

План. Для этого примера плана 2 x 2 x 3 (2 межгрупповых фактора и один фактор с повторными измерениями) откройте файл данных **Adstudy.sta**.

Выберите **Повторные измерения ДА** в поле **Вид анализа**. Выберите **Диалог** в поле **Задание анализа**.

Затем нажмите кнопку **OK** на стартовой панели **Общий ДА** для отображения диалогового окна **Повторные измерения ДА**.

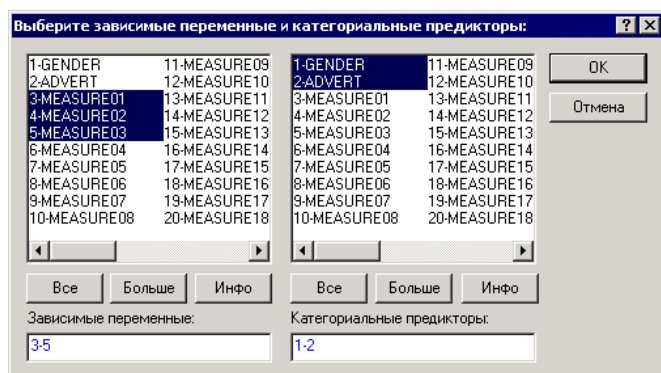


Задание плана (переменных). Первый межгрупповой фактор **Gender – Пол** имеет 2 уровня: **Male – Мужчина** и **Female – Женщина**. Второй межгрупповой фактор **Advert – Предпочтение** также имеет 2 уровня: **Pepsi** и **Coke**. Эти факторы являются

пересекающимися, т.к. среди респондентов, предпочитающих *Pepsi* или *Coke*, есть и *Male – Мужчины*, и *Female – Женщины*.

Каждый респондент был опрошен три раза и значения зависимой переменной на полученных трех уровнях фактора (повторных измерений) *Response (Отклик)* были сохранены в переменных: *Measure01*, *Measure02* и *Measure03*.

После нажатия кнопки *Переменные* в диалоговом окне *Повторные измерения ДА* откроется стандартное окно выбора переменных. Задайте переменные *Measure01 - Measure03* в качестве зависимых переменных (в списке *Зависимые переменные*), а переменные *Gender – Пол* и *Advert – Предпочтение* – в качестве факторов (в списке *Категориальные предикторы*).



Затем нажмите кнопку **ОК**, чтобы вернуться в предыдущее окно.

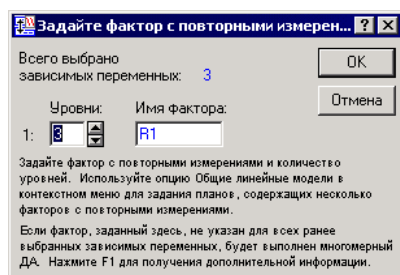
План повторных измерений. Заметим, что план эксперимента может быть представлен в виде:

	Межгрупповой Фактор #1: <i>Gender</i>	Межгрупповой Фактор #2: <i>Advert</i>	Фактор повторных измерений: <i>Отклик</i>		
			Уровень #1:	Уровень #2:	Уровень #3:
			<i>Measure01</i>	<i>Measure02</i>	<i>Measure03</i>
Субъект 1	Male	Pepsi	9	1	6
Субъект 2	Male	Coke	6	7	1
Субъект 3	Female	Coke	9	8	2
	⋮	⋮	⋮	⋮	⋮

Задание факторов повторных измерений. Теперь, после того как мы задали минимально необходимую информацию для проведения анализа, мы могли бы просто нажать **ОК** для проведения оценивания и просмотра полученных результатов.

Однако в этом примере нам нужно указать программе, что зависимые переменные содержат значения одного фактора повторных измерений, который имеет три уровня. Если не указать это специально, программа будет рассматривать три выбранные зависимые переменные как различные и проведет многомерный дисперсионный анализ.

Для того чтобы определить интересующий нас фактор повторных измерений, нажмите кнопку **Внутригруп. эффекты**. После этого появится диалоговое окно **Задайте фактор с повторными измерениями**.



Заметим, что в нашем случае программа автоматически предлагает выбрать один фактор повторных измерений с тремя уровнями (по умолчанию он назван **R1**).

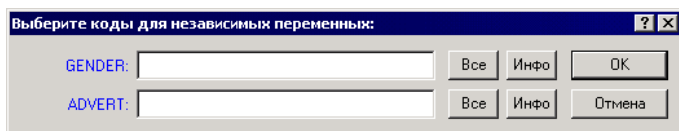
С помощью этого диалогового окна вы можете только задать один внутригрупповой фактор (фактор повторных измерений). Для задания множественных внутригрупповых факторов используйте модуль **Общие линейные модели** (содержащийся в прилагаемом пакете **Углубленные методы анализа**).

Более подробное описание планов повторных измерений можно найти в *Электронном руководстве*, нажав клавишу F1 или щелкнув мышью по кнопке **?** в этом окне.

В этом примере измените название фактора (замените **R1** на **RESPONSE**), а затем нажмите кнопку **OK**, чтобы закрыть окно.

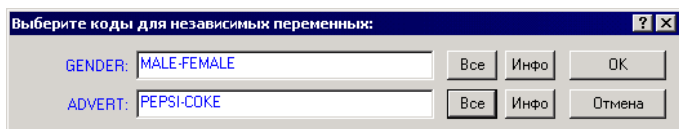
Коды, определяющие уровни межгрупповых факторов. Поскольку программа автоматически использует в качестве кодов межгрупповых факторов все встречающиеся значения соответствующих переменных, в нашем примере для этого не требуется никаких дополнительных действий [например, мы могли бы пояснить программе, что переменная **Gender – Пол** имеет два уровня: **1** и **2** (или **Male – Мужчина** и **Female – Женщина**)].

Если же задание кодов межгрупповых факторов все-таки необходимо, это можно сделать в окне **Выберите коды для независимых переменных**, для перехода в которое нажмите кнопку **Коды факторов**.



Это диалоговое окно предлагает несколько способов ввода кодов. Например, перед выбором кодов можно просмотреть значения отдельных переменных с помощью кнопки **Инфо**, просмотреть файл и указать в поле допустимые значения (например, **Gender – Пол** и **Advert – Предпочтение**).

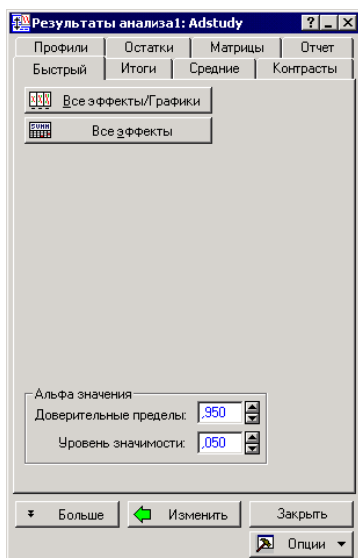
Если поле ввода, соответствующее некоторой переменной, оставлено незаполненным, после нажатия **OK** система автоматически заполнит это поле всеми ее встречающимися значениями,



и закроет данное диалоговое окно.


Проведение анализа. После нажатия кнопки **OK** в диалоговом окне **Повторные измерения ДА** система произведет анализ заданного плана и выведет полученные результаты в окне **Результаты анализа**.

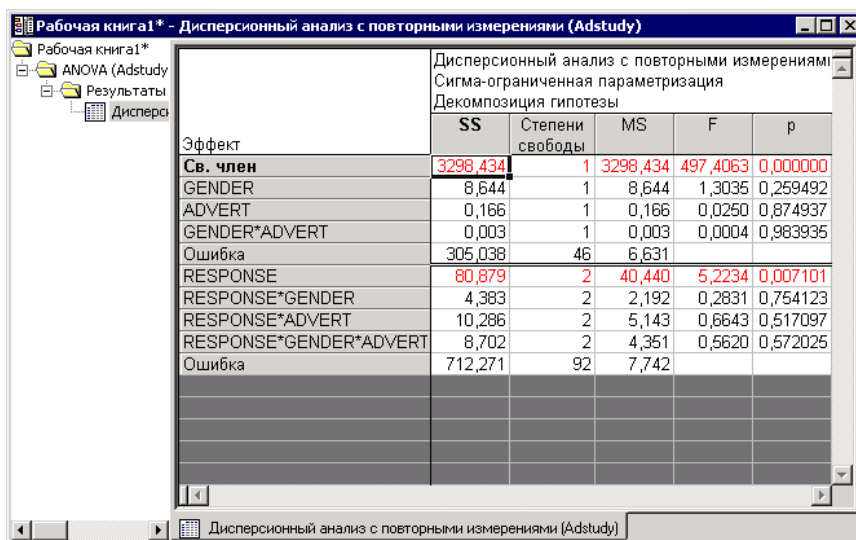
С помощью этого окна пользователь может обратиться к различным описательным графикам и таблицам результатов.



Окно содержит вкладки, позволяющие быстро выбирать требуемые результаты. Например, если вы хотите осуществить спланированные сравнения, выберите вкладку **Контрасты**. Для просмотра статистик остатков выберите вкладку **Остатки**.

Для нашего вводного примера мы будем пользоваться только результатами, доступными во вкладке **Быстрый**.

Просмотр результатов дисперсионного анализа. Начнем с просмотра итоговой таблицы дисперсионного анализа всех эффектов, которую можно построить, нажав кнопку **Все эффекты** ().



Эффект	SS	Степени свободы	MS	F	p
Св. член	3298,434	1	3298,434	497,4063	0,000000
GENDER	8,644	1	8,644	1,3035	0,259492
ADVERT	0,166	1	0,166	0,0250	0,874937
GENDER*ADVERT	0,003	1	0,003	0,0004	0,983935
Ошибка	305,038	46	6,631		
RESPONSE	80,879	2	40,440	5,2234	0,007101
RESPONSE*GENDER	4,383	2	2,192	0,2831	0,754123
RESPONSE*ADVERT	10,286	2	5,143	0,6643	0,517097
RESPONSE*GENDER*ADVERT	8,702	2	4,351	0,5620	0,572025
Ошибка	712,271	92	7,742		

В нашем примере только эффект фактора **RESPONSE – Отклик** (если не рассматривать **Св. член**) является значимым ($p = .007$). Этот результат может быть объяснен поведением среднего значения эффекта **RESPONSE – Отклик** (для получения подробной информации обратитесь к разделу *Дисперсионный анализ - Вводный обзор Электронного руководства*). Для того чтобы понять, что это значит, мы рассмотрим маргинальные средние этого эффекта.

Для возврата в диалоговое окно **Результаты анализа** (т.е. для “возобновления” анализа) нажмите сочетание клавиш CTRL+R, или выберите команду **Продолжить** в меню **Анализ**, или нажмите кнопку **Результаты анализа** на панели **Анализ**.

Когда появится диалоговое окно **Результаты анализа**, нажмите кнопку **Все эффекты/Графики** для просмотра средних значений отдельных эффектов.

Таблица всех эффектов: Adstudy

Сигма-ограниченная параметризация
Декомпозиция гипотезы

Эффект	SS	Степени свободы	MS	F	p
Св. член	3298.	1	3298.	497.4	0,000*
GENDER	9.	1	9.	1.3	,259
ADVERT	.	1	.	.0	,875
GENDER*ADVERT	.	1	.	.0	,984
RESPONSE	81.	2	40.	5.2	,007*
RESPONSE*GENDER	4.	2	2.	.3	,754
RESPONSE*ADVERT	10.	2	5.	.7	,517
RESPONSE*GENDER*ADVERT	9.	2	4.	.6	,572

Двойной щелчок мыши на эффекте вызовет график или таблицу средних.

Это диалоговое окно содержит сводную **Таблицу всех эффектов** (содержащую уже известную нам информацию). Эта таблица позволяет выбрать отдельные эффекты и построить графики соответствующих средних (или, по желанию, электронные таблицы соответствующих средних значений).

График средних для главного эффекта. Дважды щелкните на значимом главном эффекте **RESPONSE – Отклик** (отмеченном звездочкой в столбце **p**), чтобы посмотреть соответствующий график.

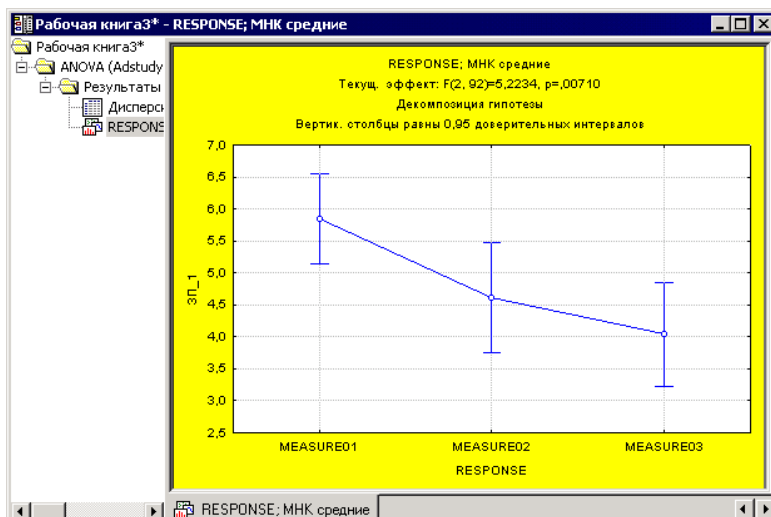


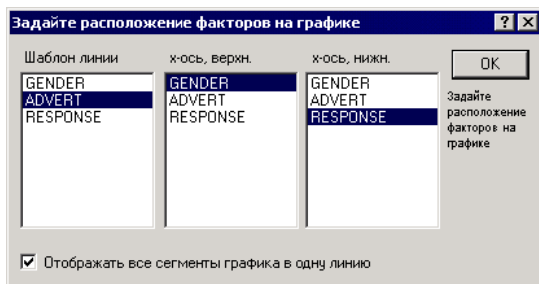
График имеет явный отрицательный тренд: среднее по трем последовательным измерениям (вопросам) постепенно убывает.

Несмотря на то, что наш план не содержит значимых взаимодействий (см. выше описание **Таблицы всех эффектов**), мы рассмотрим взаимодействия высшего порядка,

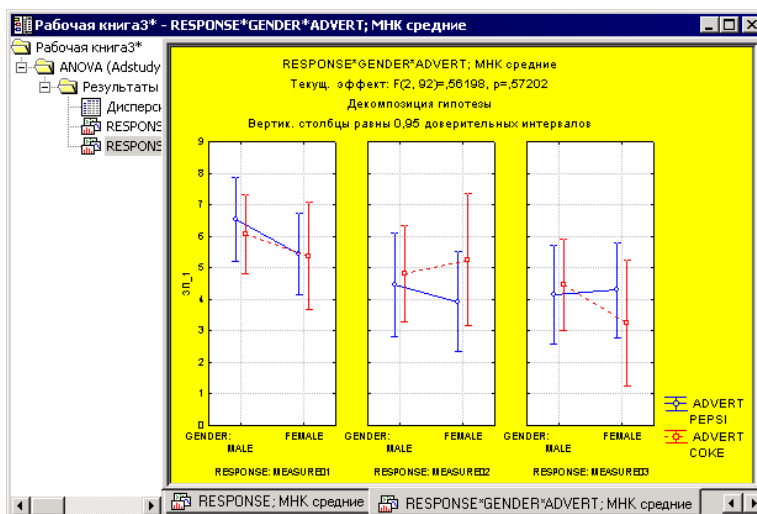
чтобы оценить достоверность такого заметного отрицательного тренда на разных уровнях межгрупповых факторов.

График средних для трехфакторного взаимодействия. Для того чтобы просмотреть график взаимодействий высшего порядка, дважды щелкните на строке, обозначенной как **RESPONSE*GENDER*ADVERT**, в диалоговом окне **Таблица всех эффектов**. Данная строка соответствует взаимодействию между факторами 1 (**Gender – Пол**), 2 (**Advert – Предпочтение**) и 3 (**Response – Отклик**).

При этом откроется промежуточное диалоговое окно **Задайте расположение факторов на графике**, позволяющее выбрать расположение факторов на графике (отметим, что в отличие от предыдущего графика, построенного для обычного фактора, в данном случае имеется несколько различных способов отображения результатов).



В нашем примере нажмите **OK**, чтобы принять расположение факторов, предложенное программой. Появится график средних.



Как видно из построенного графика, структура средних (рассматриваемых на различных уровнях межгрупповых факторов) не содержит каких-либо заметных отклонений от общей структуры с отрицательным трендом, обнаруженной на предыдущем графике (для главного эффекта **RESPONSE – Отклик**).

Теперь можно продолжить интерактивную проверку остальных эффектов и провести апостериорные, спланированные сравнения, просмотреть дополнительные описательные статистики и выполнить дальнейший разведочный анализ полученных результатов.

Интерактивный анализ данных в системе STATISTICA. Этот простой пример показывает основные принципы интерактивного анализа данных с помощью системы STATISTICA. Система не требует, чтобы пользователь еще до проведения анализа указал всю интересующую его информацию, которую следует вывести на экран. Ведь анализ даже достаточно простого плана может породить большое число электронных таблиц и просто необозримое количество графиков. Соответственно, при проведении реального анализа, до изучения основных результатов, мы с трудом можем представить, какие графики или таблицы заинтересуют нас в первую очередь. Именно поэтому система STATISTICA предоставляет пользователю возможность выбрать определенные типы вывода и интерактивно провести последовательные сравнения и моделирующий анализ уже *после того*, как данные обработаны и получены основные результаты.

Гибкие вычислительные процедуры системы STATISTICA и широкий выбор методов графического представления данных любого типа открывают перед пользователем безграничные возможности проведения разведочного анализа и проверки статистических гипотез.

Автоматизация анализов (макросы и язык SVB). Любые действия, осуществленные вами во время интерактивного анализа данных (включая задание планов и выбор опций вывода), автоматически записываются в виде кода промышленного стандарта Visual Basic.

Вы можете сохранить такие макросы для повторного использования (вы также можете поставить их в соответствие кнопкам панели инструментов, изменить или отредактировать их, объединить с другими программами и т. д.).

Для получения подробной информации см. Главу 7 - *STATISTICA Visual Basic* на стр. 155 или Главу 18 книги “Искусство анализа данных на компьютере”, [2].

3

ГЛАВА

ИНТЕРФЕЙС ПОЛЬЗОВАТЕЛЯ

Основные характеристики	79
Поддержка многозадачного режима	80
Три альтернативных интерфейса пользователя:	
1. Интерактивный пользовательский интерфейс	83
2. Язык SVB и работа с пакетом <i>STATISTICA</i> из других приложений	94
3. Интерфейс пользователя, ориентированный на Web	96





ИНТЕРФЕЙС ПОЛЬЗОВАТЕЛЯ

Основные характеристики

Индивидуальный режим работы

В системе *STATISTICA* предлагается четыре варианта пользовательского интерфейса, которые кратко описаны в следующих параграфах:

1. интерактивный интерфейс (см. стр. 82),
2. язык *SVB* (см. стр. 94),
3. Web – интерфейс (см. стр. 96).

При этом нужно иметь в виду, что:

- во многих отношениях эти варианты пользовательского интерфейса не исключают друг друга; поэтому можно использовать их в различных комбинациях в зависимости от конкретной ситуации и индивидуальных предпочтений;
- настраиваемые меню и панели инструментов могут использоваться для совмещения разных вариантов интерфейса, например, для быстрого доступа к макросам или наиболее часто используемым файлам;
- почти все характеристики интерфейсов определяются соответствующими настройками по умолчанию, которые можно изменять по желанию пользователя (изменяя тем самым внешний вид и поведение программы); как правило, рекомендуется настроить систему так, чтобы полностью использовать возможности пакета *STATISTICA* с учетом своих предпочтений и специфики решаемых задач (см. раздел *Настройка интерактивного пользовательского интерфейса*, стр. 148).

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 осуществляется любым из следующих способов:

- через обычные меню;
- через последовательности “горячих” клавиш;
- с помощью панелей инструментов и активных полей строки состояния;
- через пользовательские панели инструментов (всплывающие или зафиксированные панели инструментов с кнопками, которые определяет сам пользователь, и которые могут соответствовать макросам и командам);
- через контекстные меню, связанные с конкретными объектами (ячейками, значками Рабочих книг, графическими объектами), которые вызываются нажатием правой кнопки мыши на данном объекте.

Мы рекомендуем вам предварительно испытать все предлагаемые варианты пользовательского интерфейса системы *STATISTICA*, а затем выбрать наиболее подходящий для вас.

Поддержка многозадачного режима

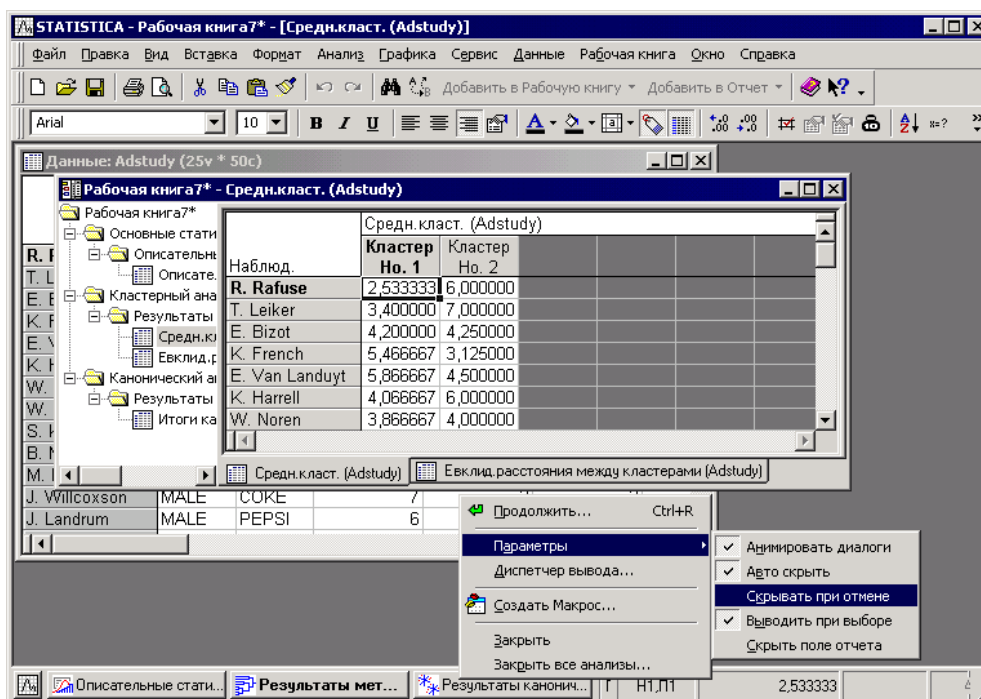
Как было замечено выше, вы можете запускать одновременно несколько копий системы *STATISTICA*. Они могут выполнять одинаковые или различные типы анализа (обычно называемые модулями), например, **Основные статистики**, **Множественная регрессия**, **Дисперсионный анализ** и т.д.

Более того, в одном приложении системы *STATISTICA* одновременно может быть открыто несколько анализов. Они могут быть одного и того же или различного типа

(например, пять модулей **Множественная регрессия** и два модуля **Дисперсионный анализ**). Анализы могут проводиться над одним и тем же или различными исходными файлами данных (несколько исходных файлов данных можно открыть одновременно).

Индивидуальные “анализы” – функциональные части вашей работы. Чтобы сделать более удобным “многозадачный” режим работы в системе *STATISTICA*, вся работа организуется как набор функциональных частей, называемых “анализами”. Они представлены с помощью кнопок на панели **Анализ** в нижней части окна приложения (выше строки состояния, см. иллюстрацию ниже, где **Основные статистики**, **Кластерный анализ** и **Канонический анализ** запущены одновременно).

Во время работы создается, как минимум, одна кнопка анализа, и последующие кнопки добавляются по мере открытия новых анализов. Настройка этой возможности системы *STATISTICA* осуществляется с помощью ряда опций.



По умолчанию, когда вы вызываете таблицу или график из диалогового окна результатов, происходит отображение заданного объекта, и диалоговое окно автоматически минимизируется. Оно превращается в соответствующую кнопку анализа в нижней части экрана.

Чтобы снова развернуть диалоговое окно и возобновить анализ, щелкните по этой кнопке (или нажмите сочетание клавиш CTRL+R).

В контекстном меню, соответствующем кнопке анализа, имеется набор опций для управления этим анализом (как показано выше). Данное меню отображается при щелчке правой кнопкой мыши по кнопке анализа на панели **Анализ**.

Полезный совет для пользователей, имеющих большие мониторы. Пользователи, имеющие большие экраны, могут отключить установленную по умолчанию минимизацию диалоговых окон. Большинство диалоговых окон имеет небольшой размер, поэтому их можно оставлять на экране и использовать как “полу-панели инструментов” для выбора объектов вывода.

Вы можете отменить минимизацию диалоговых окон как для отдельного анализа, так и для всей программы. В первом случае отмените опцию **Авто скрыть** в контекстном меню кнопки анализа, как показано выше. Во втором случае используйте вкладку **Анализ/Графики** в диалоговом окне **Параметры**, доступном через меню **Сервис - Параметры**.

Если вы запустили много анализов и Рабочая область *STATISTICA* загромождена, то вы можете скрыть все окна, относящиеся к отдельным анализам (или закрыть их с помощью команды **Закрыть все анализы** контекстного меню кнопки любого анализа). Можно также запустить новое приложение *STATISTICA*, что позволит по-другому организовать работу.

1. Интерактивный пользовательский интерфейс

Введение

Основные компоненты интерактивного пользовательского интерфейса в системе STATISTICA. Хотя интерактивный пользовательский интерфейс - не единственный в системе STATISTICA (см. Главу 6 – *Настройка системы STATISTICA*, стр. 147 и Главу 7 – *STATISTICA Visual Basic*, стр. 155), он является наиболее простым и чаще всего используется.

Многие компоненты этого пользовательского интерфейса можно увидеть на экране системы STATISTICA.

Во-первых, как и во многих популярных программах, строки меню и различные панели инструментов отображаются в верхней части экрана. Они могут быть настроены по вашему вкусу.

В нижней части экрана располагается панель **Анализ** (содержащая свернутые диалоговые окна анализов и графиков) и строка состояния. Также имеются контекстные меню, доступные с помощью правой кнопки мыши.

Файлы данных могут быть отображены в электронных таблицах, Рабочих книгах, Отчетах или отдельных окнах. Электронные таблицы результатов и графики могут быть вставлены в Рабочие книги, Отчеты и отдельные окна. Заметим, что дополнительные документы (такие как документы Microsoft Word или картинки Bitmap) могут быть также отображены в электронных таблицах, Рабочих книгах или Отчетах. Наконец, в окнах макросов отображается код на языке SVB.

Обычно все эти средства одновременно не используются. В зависимости от ваших потребностей или предпочтений вы можете упрощать или усложнять пользовательский интерфейс STATISTICA (см. стр. 147). Возможности настройки интерфейса подробно описаны в *Электронном руководстве*.

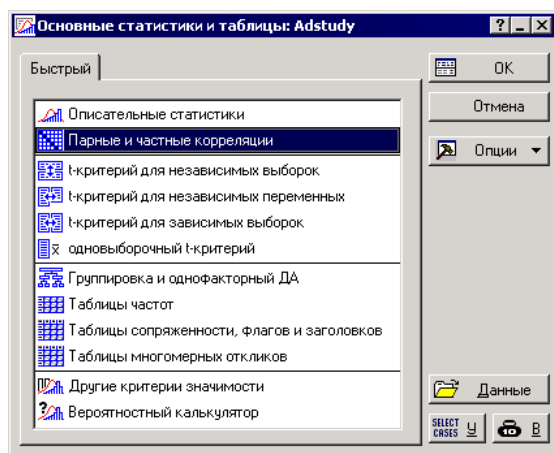
Модули. Так как система STATISTICA предоставляет набор статистических и графических процедур, каждая процедура может быть выполнена в одном и том же приложении STATISTICA. Это означает, что, например, можно вычислить статистики остатков, используя опции модуля **Множественная регрессия**, затем непосредственно

использовать полученные результаты в модуле **Факторный анализ** или другом исследовательском модуле, не запуская при этом приложение *STATISTICA* заново.

Для получения подробной информации об использовании результатов анализа в качестве входных данных см. раздел *Можно ли использовать результаты одного Анализа для выполнения другого?* в *Электронном руководстве*.

“Сеанс” интерактивного анализа

Стартовая панель. При запуске любого модуля из меню **Анализ** открывается его стартовая панель (как показано ниже, а также в примере *Пример 1: Корреляции* на стр. 45).



Каждая стартовая панель содержит список типов анализов, доступных в данном модуле.

Панель автоматически минимизируется в кнопку на панели **Анализ** при щелчке мышью где-либо вне стартовой панели. Пользователи, имеющие экраны с большим разрешением, могут отменить эту установку. Тогда следующие друг за другом диалоговые окна будут оставаться на экране (в каждой последовательности анализа). В этом случае они могут использоваться как “панели инструментов”, например, для вывода результатов (см. стр. 82).

Панели инструментов. Вы можете открыть любой модуль анализа с помощью кнопки на панели инструментов **Анализ**. Чтобы активизировать эту панель, щелкните правой кнопкой мыши на любой панели инструментов и выберите **Анализ** из контекстного меню, отображающего список доступных панелей инструментов.

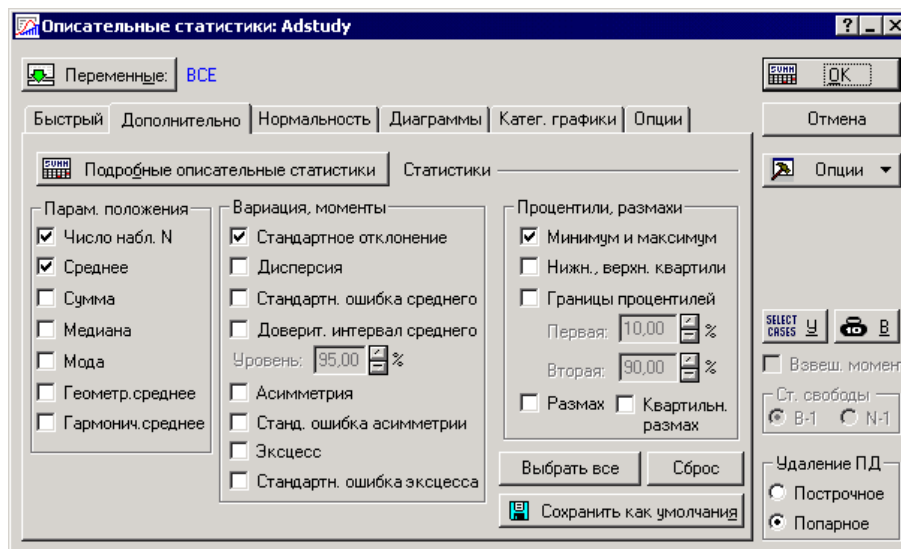
Кроме того, вы можете выбрать команду **Анализ** в меню **Вид - Панели инструментов**.

Вы можете также создать свою собственную панель инструментов, содержащую кнопки для наиболее часто используемых вами видов анализа (как показано на стр. 92).

См. также *Обзор панелей инструментов в Электронном руководстве*.

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

Часто в таких окнах имеется несколько вкладок, в которых сгруппированы опции, анализы и/или результаты по логическим категориям для облегчения поиска нужной опции.




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

Вывод. Как подробно описано в Главе 3 – *Три канала вывода для анализа* (стр. 101), и как показано в примерах *Пример 1: Корреляции* (стр. 45) и *Пример 2: Дисперсионный анализ* (стр. 68), последовательно выводимые электронные таблицы и графики появляются в Рабочих книгах. Эти Рабочие книги могут быть сохранены и затем заново открыты, что облегчает возврат к полученным результатам при необходимости.

Дополнительно вы можете направить весь вывод в Отчет анализа (см. стр. 104), который легко организуется (с помощью дерева Отчета), легко форматируется и удобен для вывода на печать.

Вы можете направить все результаты, вне зависимости от выполняемого анализа, в один Отчет, а можете осуществить вывод в различные окна.

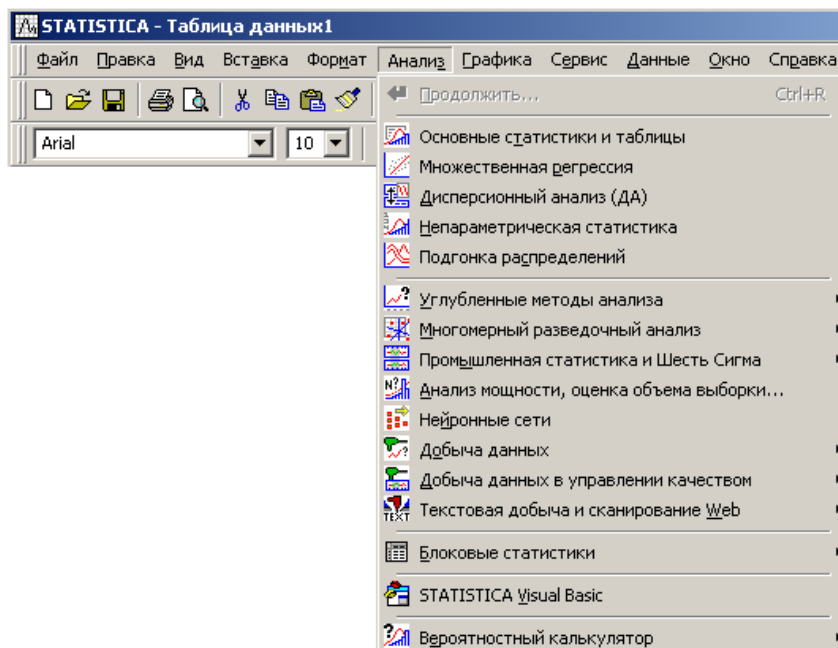
В любом случае параметры вывода для простого анализа или сессии можно задать, нажав на кнопку  **Опции** в диалоговом окне задания анализа или графика и выбрав команду **Диспетчер вывода**.

Общие параметры вывода доступны во вкладке **Диспетчера вывода** диалогового окна **Параметры**. Чтобы открыть это окно, выберите команду **Параметры** в меню **Сервис** или команду **Диспетчера вывода** в меню **Файл**.

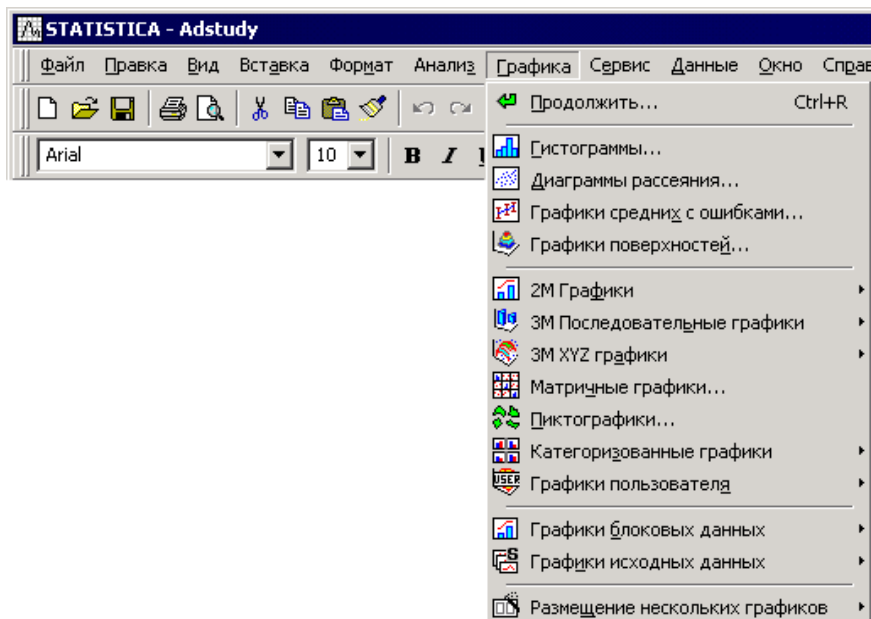
За подробной информацией обращайтесь к *Электронному руководству*.

Особенности процедуры анализа

Система *STATISTICA* позволяет получать прямой доступ ко всем статистическим и графическим окнам анализов с помощью меню **Анализ**



и Графика



(либо с помощью панелей инструментов **Анализ** и **Графика**).

Эти меню никогда не исчезают, т.е. доступны всегда, когда открыт какой-либо документ с исходными данными.

Меню **Анализ** предоставляет доступ ко всем типам анализа системы *STATISTICA*. Меню **Графика** предоставляет прямой доступ к набору наиболее используемых типов графиков (например, к диаграммам рассеяния, гистограммам, графикам средних с ошибками и т.п.).

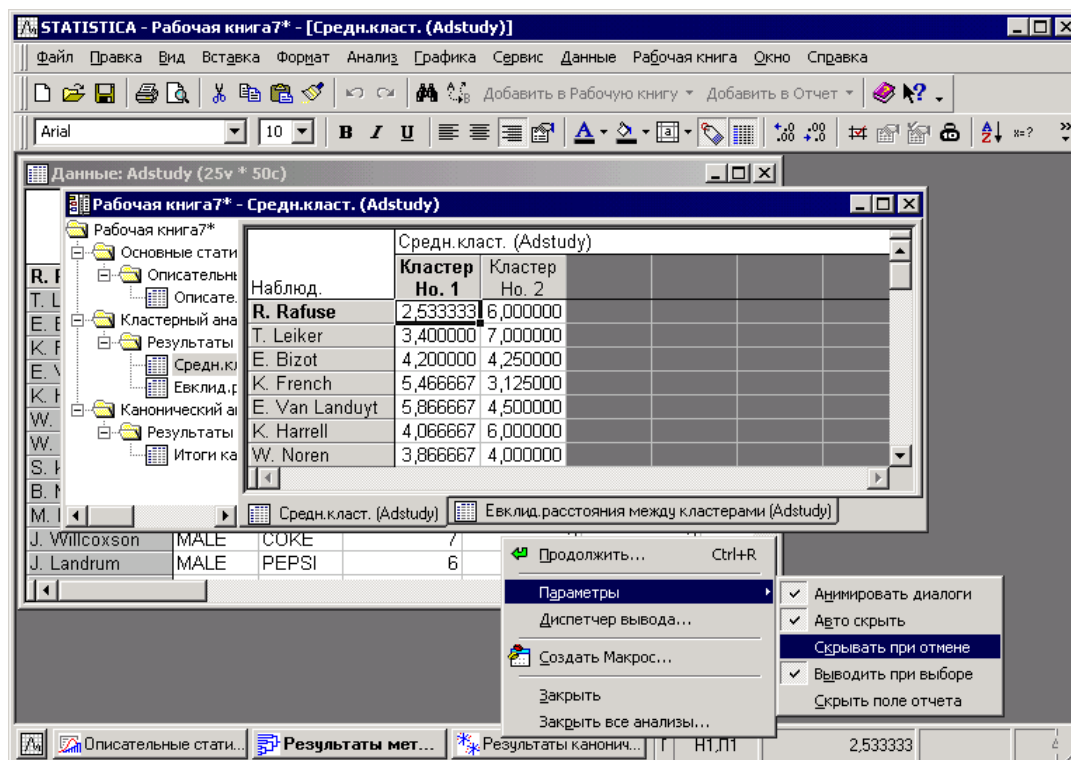
Также из меню **Графика** возможен иерархический доступ ко всем типам графиков в системе *STATISTICA*, включая **2М графики**, **3М последовательные** и **XYZ графики**, **Матричные графики**, **Пиктографики**, **Категоризованные графики**, **Графики пользователя**, **Графики блоковых данных** и **Графики исходных данных**.

Полное описание различных типов статистик и графиков, предлагаемых системой *STATISTICA*, приведено в словаре терминов *Электронного руководства*.

Также см. Приложение E: Семейство продуктов *STATISTICA* (стр. 209) для получения информации обо всех приложениях по анализу данных семейства продуктов *STATISTICA*.

Использование панели Анализ. Для поддержания удобного “многозадачного” режима системы *STATISTICA* (см. *Поддержка многозадачного режима*, стр. 80), процедуры анализа в ней организованы как функциональные единицы. Они представлены кнопками на панели **Анализ** в нижней части окна приложения (над строкой состояния, см. рисунок ниже, где одновременно запущены модули **Основные статистики**, **Кластерный анализ** и **Канонический анализ**).

Во время работы создается, как минимум, одна кнопка анализа, и последующие кнопки добавляются по мере открытия новых анализов.



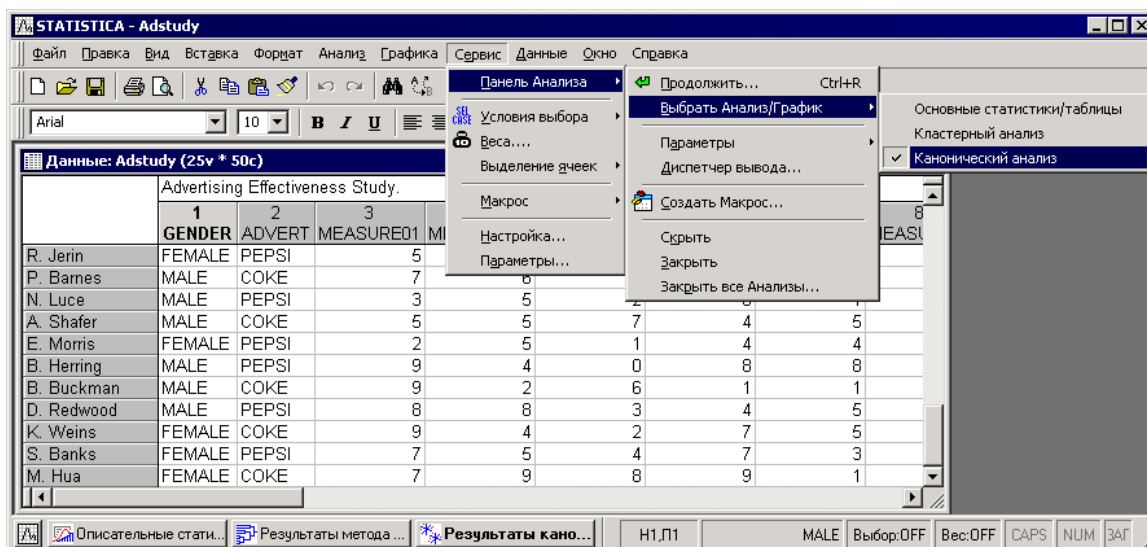
Минимизация диалоговых окон (и рекомендации для пользователей с большими экранами). Вы можете выбрать режим, когда все диалоговые окна системы минимизируются при переходе к другому окну системы *STATISTICA* или к другому приложению.

По умолчанию выбрана команда **Авто скрыть**; но если размеры вашего экрана позволяют свободно размещать несколько окон, рекомендуется отменить эту опцию. Это позволяет оставлять диалоговые окна на экране при выводе результатов из этих

диалоговых окон, и вы можете пользоваться этими диалогами как “панелями инструментов” для вывода. См. подробности на стр. 82.

Продолжение анализа/графика. Вы легко можете возобновить текущий анализ или график (то есть вызвать текущее окно заданного анализа). Выберите команду **Продолжить** в меню **Сервис - Панель Анализа** или же нажмите CTRL+R либо кнопку анализа/графика на панели **Анализ**.

Когда выполняется несколько анализов, вы можете выбрать один из них с помощью меню **Сервис - Панель Анализа - Выбрать Анализ/График** (как показано ниже).



Скрытие окон. Для облегчения работы с окнами вы можете скрывать все относящиеся к текущему анализу окна, когда этот анализ становится неактивным. Для этого установите опцию **Скрывать при выборе** в меню **Сервис - Панель Анализа - Параметры**. По умолчанию она отменена.

Заметим, что эта команда применима только тогда, когда результаты направляются в отдельные окна; см. описание **Диспетчера вывода** (стр. 101).


Кроме того, имеется команда закрытия всех окон документов, **Окно - Закрыть все** (или CTRL+L), а также команда закрытия всех анализов, **Сервис - Панель Анализа - Закрыть все Анализы**.

Вынесение окон на передний план. Установите опцию **Выводить при выборе** в меню **Сервис - Панель Анализа - Параметры**, чтобы вынести на передний план все окна, относящиеся к отдельному анализу, при выборе этого анализа.

Данная опция облегчает также организацию окон различных анализов. Она установлена по умолчанию.

Заметим, что эта команда применима только тогда, когда результаты направляются в отдельные окна; см. описание **Диспетчера вывода** (стр. 101).

Скрытие поля Отчета. По умолчанию поля Отчета располагаются в верхней части любых окон результатов (таких как **Результаты множественной регрессии**) и содержат краткую информацию об анализе.


Вы можете скрыть единичное поле Отчета с помощью кнопки  в нижнем правом углу этого поля. Вы можете также запретить появление всех полей Отчета в системе путем выбора опции **Скрыть поле Отчета** в меню **Сервис - Панель Анализа - Параметры**.

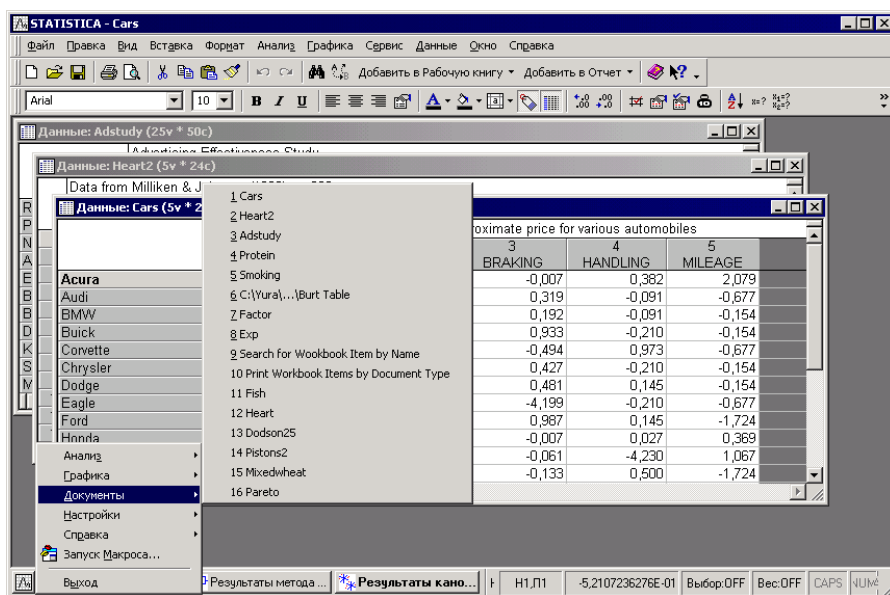
Типы документов

Система *STATISTICA* поддерживает работу с пятью основными типами документов:

- Рабочими книгами (см. стр. 102 и 109),
- электронными таблицами (мультимедийными таблицами) (см. стр. 113),
- Отчетами (см. стр. 104 и 119),
- графиками (см. стр. 122 и 127),
- макросами (*SVB*) (см. стр. 123 и 155).

Используя эти пять типов документов, вы можете управлять различного рода данными, осуществлять ввод данных и их обработку, создавать графики высокого качества, разрабатывать пользовательские приложения любой степени сложности, а также создавать Отчеты различного формата.

Вы можете быстро получать доступ к наиболее используемым документам. Нажмите стартовую кнопку *STATISTICA*  (в нижнем левом углу экрана) и выберите **Документы**.



Используя вкладку **Общие** окна **Параметры** (доступного через меню **Сервис - Параметры**), вы можете задать количество хранимых ссылок на документы, которые открывались последнее время (по умолчанию **9**).

Более подробную информацию о каждом типе документов см. в обзорах Рабочих книг, электронных таблиц, Отчетов, графиков и макросов (стр. 109). См. также *Электронное руководство*.

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. Команды меню и кнопки панелей инструментов для каждого типа документов подробно описаны в *Электронном руководстве*.

Заметим, что Рабочие книги не имеют собственной панели инструментов (хотя всегда отображается панель инструментов **Стандартная**), поскольку в Рабочих книгах могут отображаться различные документы. Таким образом, когда вы внутри Рабочей книги редактируете электронную таблицу, график, Отчет, макрос или внешний документ (например, электронную таблицу Microsoft Excel), отображаются те панели инструментов и меню, которые соответствуют данному типу документов.

Если вы выбрали “пустой узел” навигационного дерева Рабочей книги, то по умолчанию появляется панель инструментов **Анализ** (на месте предыдущей панели инструментов).

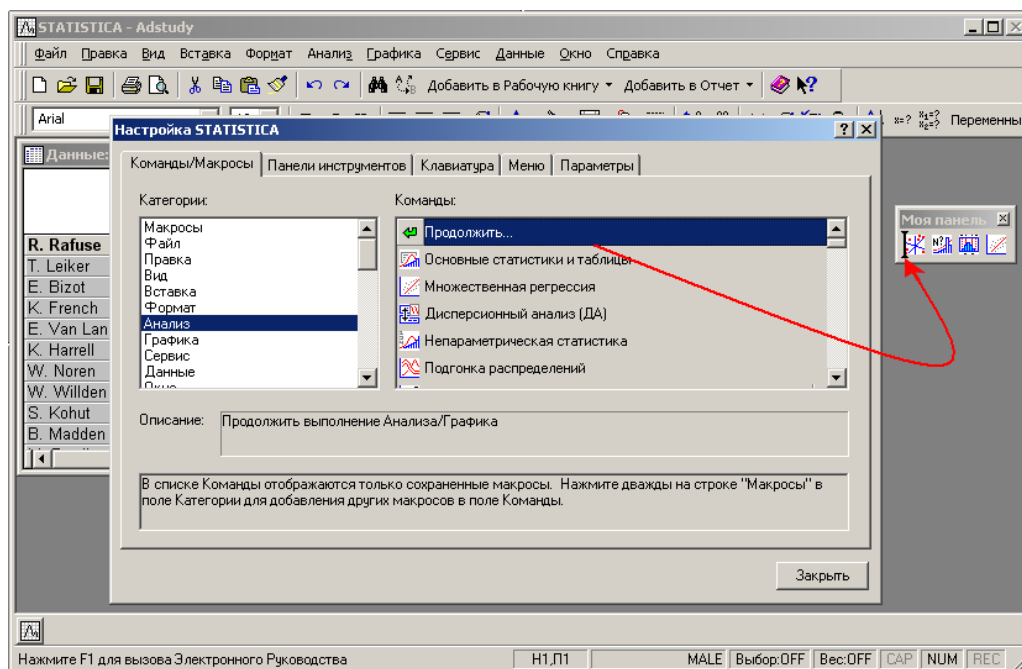
Пользовательские панели инструментов. В дополнение к набору панелей инструментов, который предоставляет система *STATISTICA*, вы можете также создавать собственные панели инструментов.

Эти панели инструментов могут содержать любые команды, доступные в системе *STATISTICA*, а также специальные элементы управления (например, имя шрифта, размер шрифта, стили графика и т.д.). Вы можете определить для каждой панели инструментов собственное имя и соответствующий тип документа.

Кроме того, вы можете настраивать все панели инструментов (включая существующие) путем добавления команд и элементов управления.

Чтобы создать панель инструментов (или изменить существующую), используйте вкладку **Панели инструментов** в окне **Настройка STATISTICA**, доступном из меню **Сервис - Настройка**.

Перетаскивая отдельные команды из этого диалога на панель инструментов, вы можете настроить ее соответствующим образом, как показано на рисунке ниже.



Вид и расположение панелей инструментов можно легко изменить (например, все панели инструментов могут быть либо прикрепленными, либо плавающими).

Все эти опции позволяют вам создавать уникальные панели инструментов и обустроить специализированный пользовательский интерфейс. В *Электронном руководстве* содержатся простые пошаговые инструкции по осуществлению таких настроек. В частности, см. раздел *Создание новой панели инструментов* в *Электронном руководстве*.

Пользовательские меню. Настройка меню так же проста, и производится с помощью вкладки **Меню** диалога **Настройка STATISTICA**, показанного выше (подробности см. в *Электронном руководстве*).

2. Язык SVB и работа с пакетом *STATISTICA* из других приложений

Стандартный язык программирования *STATISTICA* Visual Basic (интегрированный в систему *STATISTICA*) предоставляет намного больше возможностей, чем просто “вспомогательный язык программирования”, который используется для создания пользовательских приложений.

STATISTICA Visual Basic (SVB) использует огромные преимущества объектно-ориентированной структуры системы *STATISTICA* и позволяет получить доступ программными средствами практически ко всем функциональным деталям пакета *STATISTICA*. Даже самые сложные процедуры анализа и графический вывод результатов могут быть записаны в виде макросов на языке Visual Basic, которые затем можно воспроизвести или заново отредактировать.

Макросы представляют собой отдельные блоки, которые легко встраиваются в другие приложения.

SVB добавляет более 10 000 новых функций к стандартному синтаксису Microsoft Visual Basic, и является, таким образом, одним из самых функционально богатых и обширных интерфейсов прикладного программирования.

Дополнительную информацию о языке *SVB* см. в Главе 7 (стр. 155).

Работа с пакетом *STATISTICA* из других приложений. Одной из важнейших особенностей окружения *SVB* является способность управлять различными приложениями из одной программы.

Например, вы можете написать программу на языке *SVB*, вычисляющую прогнозы с помощью модуля *STATISTICA* **Временные ряды**, а затем выполнить ее в электронной таблице Excel или в документе Microsoft Word.

Обмен информацией между различными приложениями осуществляется путем представления этих приложений в программах Visual Basic в виде Объектов. Например, вы можете запустить статистический анализ в модуле *STATISTICA* **Основные статистики и таблицы** из программы Visual Basic в Excel, объявив внутри программы объект типа **Statistica.Application**.

После того как объект создан, программа Visual Basic получает доступ к свойствам и методам этого объекта. Свойства обычно представляются в виде переменных, методы - в

виде подпрограмм или функций, которые осуществляют определенные операции или вычисления внутри соответствующего объекта приложения.

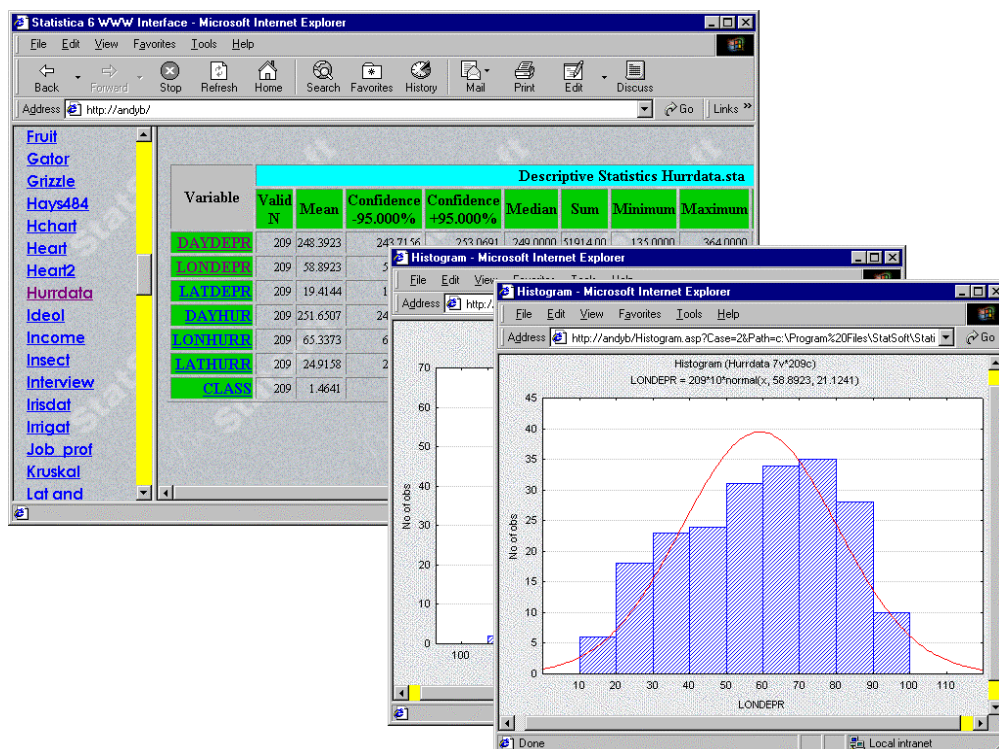
Вы можете вызывать процедуры *STATISTICA* напрямую из многих других приложений и языков программирования (например, C++, Java и др.).

3. Интерфейс пользователя, ориентированный на Web

Система *STATISTICA* поддерживает Web-технологии, и поэтому как “вывод”, так и “ввод” в *STATISTICA* можно осуществлять через Internet.

Существует возможность направления результатов анализа прямо на Web-сервер. Дополнительно *STATISTICA* предоставляет методы для построения весьма сложных автоматизированных систем, запускающих сценарии, которые могут, например, периодически получать данные из внешних источников, проводить их анализ и обновлять содержание HTML-страниц на заданных Web-серверах. (Заметим, что эти возможности доступны только для версий, лицензированных для размещения на Web-серверах.)

STATISTICA может быть запущена с пользовательским интерфейсом в стиле Web-браузера. В этом случае браузер Internet выступает в качестве внешнего интерфейса для *STATISTICA* (“тонкий клиент”).



Таким образом, аналитической мощностью системы *STATISTICA* можно воспользоваться с любого компьютера в мире, при условии, что он подключен к Internet. *Заметим, что такая возможность системы STATISTICA доступна только при соответствующем лицензировании и не может быть использована на локальной версии STATISTICA.*

STATISTICA Web-сервер (технические замечания). Эта последняя функциональная возможность поддерживается в системе *STATISTICA*, поскольку ее всесторонние средства автоматизации (основанные на объектной модели) могут быть запущены с помощью языка сценариев (например, VBScript, Jscript или C++).

Если *STATISTICA* установлена на сервере, вы можете получить доступ к модели автоматизации для обработки данных и получения результатов, которые вы можете отослать клиенту в виде HTML-таблиц и графических (JPG/PNG) файлов.

Более того, если *STATISTICA* установлена на стороне клиента, вы также можете использовать клиентские сценарии. Тот факт, что ваш Internet-браузер работает с документами ActiveX, позволяет вам открывать электронные таблицы и графики *STATISTICA* в браузере вместе с панелями инструментов и меню *STATISTICA*.

Web-ориентированный пользовательский интерфейс предлагается компанией StatSoft в виде набора упрощенных “диалоговых шаблонов”, вызываемых в Internet-браузере, что позволяет вам легко выполнять основные типы анализа и строить графики.

Заметим, что этот интерфейс может быть гибко настроен путем изменения сценариев анализов, выполняемых на стороне сервера. Web-ориентированный пользовательский интерфейс не только полностью настраивается, но также позволяет включать в себя другие приложения, которые взаимодействуют с системой *STATISTICA*.

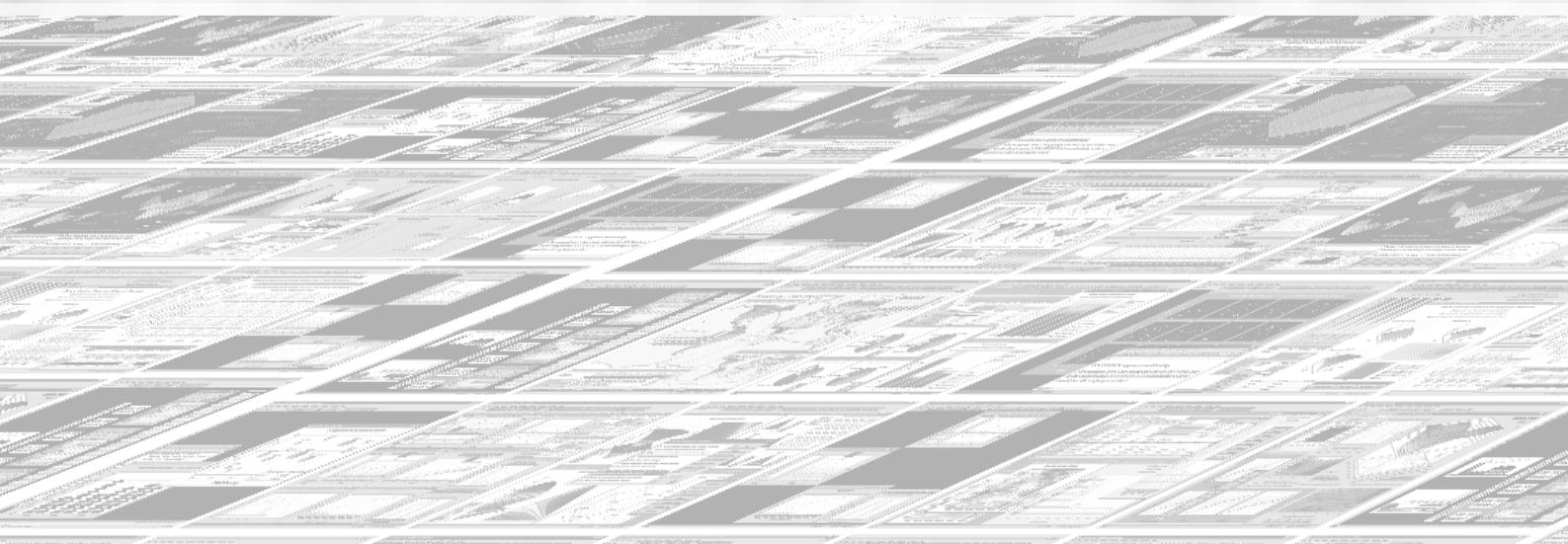
4

ГЛАВА

ТРИ КАНАЛА ВЫВОДА РЕЗУЛЬТАТОВ АНАЛИЗА

1. Рабочие книги.....	102
2. Отчеты	104
3. Автономные окна.....	106





ТРИ КАНАЛА ВЫВОДА РЕЗУЛЬТАТОВ АНАЛИЗА

Когда вы проводите анализ, *STATISTICA* выводит результаты в виде мультимедийных таблиц (электронных таблиц) и графиков.

Управление выводом осуществляется с помощью **Диспетчера вывода** (доступного из меню **Файл - Диспетчер вывода**, подробности см. на стр. 56).

Вы можете направлять результаты по трем основным каналам вывода. Это:

1. Рабочие книги,
2. Отчеты,
3. Автономные окна.

Каналы вывода результатов могут использоваться в различных сочетаниях (например, могут быть выбраны одновременно Рабочая книга и Отчет), и каждый канал вывода может быть настроен различным образом.

Кроме того, все объекты вывода (электронные таблицы и графики) могут содержать другие внедренные и связанные объекты и документы. Таким образом, *STATISTICA* предоставляет возможность иерархической организации вывода.

Каждый из трех каналов вывода результатов системы *STATISTICA* имеет свои преимущества, описанные в последующих трех параграфах. Всесторонние описания каждого типа документов, связанных с соответствующим каналом вывода результатов, даны в следующей главе (стр. 109).

1. Рабочие книги

Рабочие книги являются стандартным средством управления выводом (см. также стр. 109). Каждый итоговый документ (например, электронная таблица или график *STATISTICA*, так же как и документ Microsoft Word или Excel), представляется в виде вкладки в Рабочей книге.

Документы можно организовать в виде иерархии папок, или узлов документов, используя дерево просмотра, в котором отдельные документы, папки или целые ветви могут быть легко изменены.

	1	2	3	4	5	6
	GROUP	GENDER	TIME	PAID	STRESS R	CORRECT1
1	EXPERMTL	MALE	BEFORE	NOT_PAID	1,41	12
2	EXPERMTL	MALE	BEFORE	NOT_PAID	1,73	3
3	EXPERMTL	MALE	BEFORE	PAID	0,00	7
4	EXPERMTL	MALE	BEFORE	PAID	1,41	11
5	EXPERMTL	MALE	AFTER_1	NOT_PAID	12,83	8
6	EXPERMTL	MALE	AFTER_1	NOT_PAID	2,24	15
7	EXPERMTL	MALE	AFTER_1	PAID	2,24	15
8	EXPERMTL	MALE	AFTER_1	PAID	13,32	11
9	EXPERMTL	MALE	AFTER_2	NOT_PAID	13,74	19
10	EXPERMTL	MALE	AFTER_2	NOT_PAID	13,32	11
11	EXPERMTL	MALE	AFTER_2	PAID	13,00	9
12	EXPERMTL	MALE	AFTER_2	PAID	13,16	10
13	EXPERMTL	FEMALE	BEFORE	NOT_PAID	14,00	8
14	EXPERMTL	FEMALE	BEFORE	NOT_PAID	12,23	6
15	EXPERMTL	FEMALE	BEFORE	PAID	13,22	8
16	EXPERMTL	FEMALE	AFTER_1	PAID	12,23	7
17	EXPERMTL	FEMALE	AFTER_1	NOT_PAID	13,00	5

Например, набор документов можно перенести (с помощью метода “перетащить и отпустить”) в окно Отчета или в Рабочую область *STATISTICA* (где они отображаются в виде Автономных окон).

При настройке иерархии папок можно различными способами перенести в другие Рабочие книги целые ветви дерева.

Говоря техническим языком, Рабочие книги являются ActiveX-документами (см. стр. 163, а также *Электронное руководство*).

Рабочие книги совместимы с некоторыми другими форматами файлов (например, с документами Microsoft Office), которые можно добавлять в Рабочие книги и редактировать “на месте”.

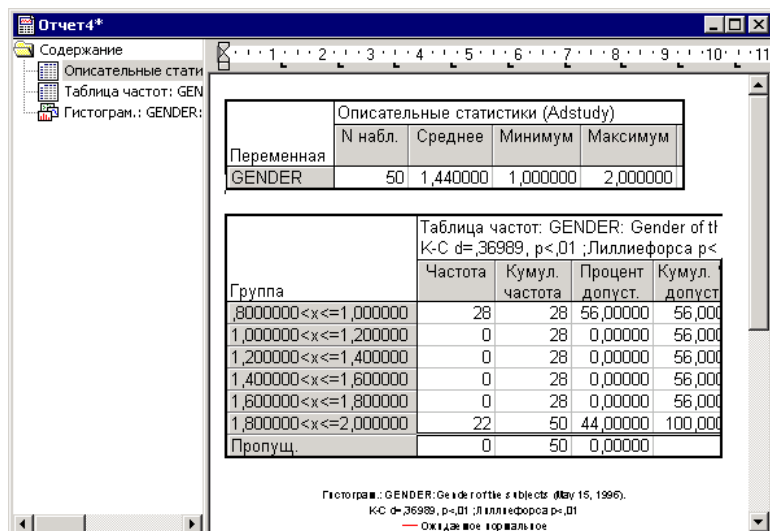
Пользовательские примечания и комментарии в Рабочих книгах. Рабочие книги предоставляют мощные возможности для эффективной работы с чрезвычайно большим количеством итоговых документов, являются, пожалуй, наилучшим способом управления выводом как для новичков, так и для продвинутых пользователей.

Может случиться, что одним из возможных недостатков будет то, что пользовательские комментарии (например, примечания) и дополнительная информация не могут быть явно вставлены в “поток” вывода Рабочей книги, как это возможно в традиционных текстовых документах, таких как Отчеты *STATISTICA* (см. следующий параграф). Однако заметим, что:

- Ко всем документам *STATISTICA* можно легко добавить комментарии
 - (a) напрямую, введя текст в графики, таблицы и Отчеты,
 - (b) косвенно, путем ввода заметок в поле **Комментарии** диалогового окна **Свойства документа** (доступного из меню **Файл - Свойства**).
- Форматированные документы с примечаниями и комментариями (в форме текстовых файлов, Отчетов *STATISTICA*, документов WordPad и т.п.) можно легко добавить в любое место иерархии итоговых документов в Рабочих книгах. Кроме того, такие документы могут быть сделаны узлами для групп подчиненных объектов, которым предназначены комментарии.

2. Отчеты

Отчеты в системе *STATISTICA* (см. также стр. 119) предоставляют более традиционный способ организации вывода, при котором каждый объект (например, электронная таблица *STATISTICA*, график *STATISTICA* или таблица Microsoft Excel) последовательно отображается в текстовом документе.



Несмотря на это, используемая в таких Отчетах технология предоставляет большие функциональные возможности.

Например, Отчет системы *STATISTICA* является также ActiveX-контейнером, в котором каждый из объектов (не только электронные таблицы и графики *STATISTICA*, но также и другие ActiveX-совместимые документы, например, таблицы Microsoft Excel) можно активизировать, настраивать и редактировать “на месте”. (Описание технологии ActiveX см. на стр. 163 или в *Электронном руководстве*.)

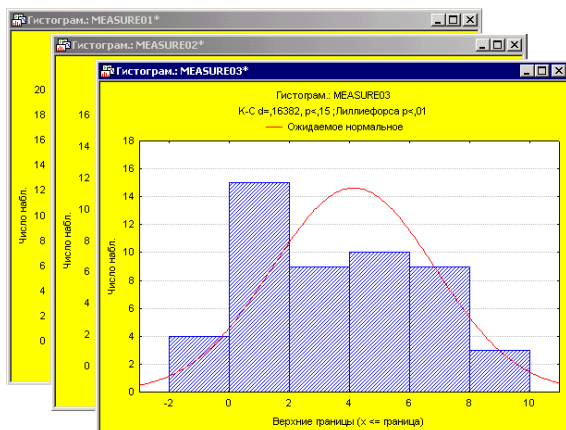
Очевидными преимуществами такой организации вывода (более традиционного, чем Рабочая книга) являются возможности добавления примечаний и комментариев к объектам, а также быстрой прокрутки и просмотра результатов вывода в таком виде, к какому привыкли пользователи. (Редактор поддерживает изменяющуюся скорость прокрутки окна и другие возможности IntelliMouse.)

Кроме того, только Отчеты содержат и сохраняют дополнительную информацию, включающую в себя подробный журнал заданных для анализа опций (например, выбранные переменные, их метки, длинные имена и т.п.). Объем дополнительной информации указывается во вкладке **Диспетчер вывода** диалогового окна **Параметры** (доступного из меню **Файл - Диспетчер вывода**), см. стр. 56.

Очевидным недостатком традиционных Отчетов является их одноуровневая структура, хотя для некоторых пользователей или приложений такая структура может оказаться более привлекательной.

3. Автономные окна

Наконец, вывод результатов в системе *STATISTICA* может быть также направлен в очередь Автономных окон. **Длина очереди** задается во вкладке **Диспетчер вывода** диалогового окна **Параметры** (доступного из меню **Файл - Диспетчер вывода**).



Существенным недостатком такого режима вывода является отсутствие организации и возможность загромождения Рабочей области приложения (заметим, что некоторые процедуры могут создавать сотни таблиц или графиков при нажатии одной кнопки).

Достоинством такого способа вывода является простота размещения этих объектов на Рабочей области *STATISTICA*.

Однако заметим, что вы можете заранее не настраивать вывод и не создавать большое число (зачастую ненужных) окон, которые могут загромоздить Рабочую область. Вместо этого отдельные объекты вывода, сохраненные в Рабочих книгах и Отчетах, можно при необходимости легко перенести из соответствующего дерева просмотра в Рабочую область приложения.

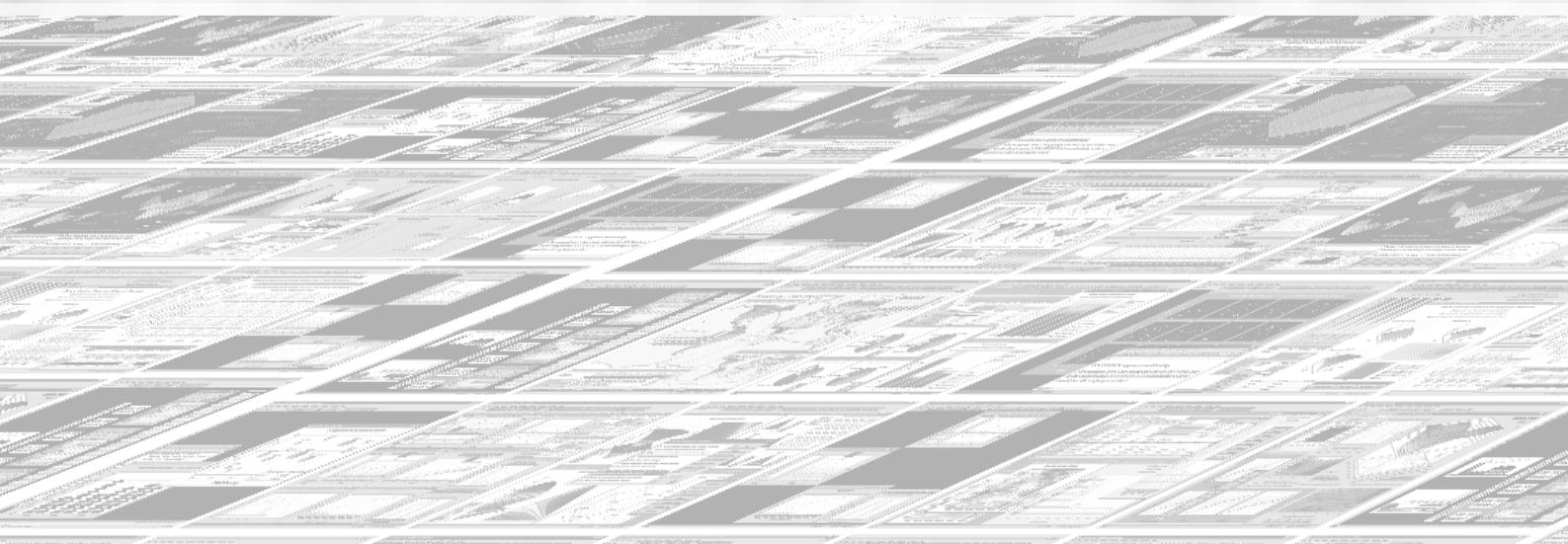
5

ГЛАВА

ДОКУМЕНТЫ СИСТЕМЫ *STATISTICA*

Рабочие книги.....	109
Электронные таблицы (мультимедийные таблицы).....	113
Отчеты	119
Графики	122
Макросы (программы на языке <i>SVB</i>).....	123

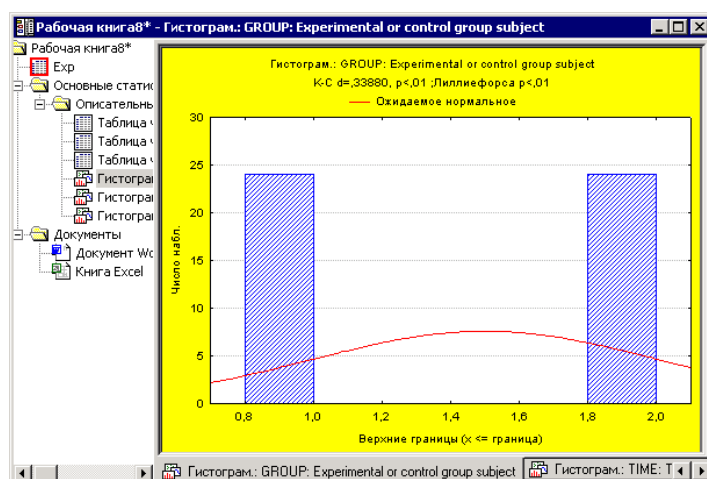




ДОКУМЕНТЫ СИСТЕМЫ *STATISTICA*

Рабочие книги

Рабочие книги (кратко описанные на стр. 102) являются стандартным средством управления выводом. В Рабочей книге каждый документ (например, электронная таблица или график *STATISTICA*, так же как и документ Microsoft Word или Excel) представлен в виде вкладки.



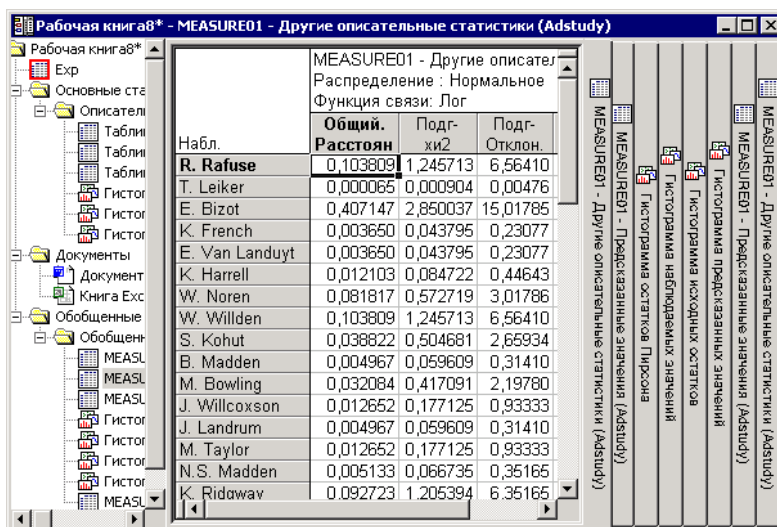
Говоря техническим языком, Рабочие книги *STATISTICA* являются оптимизированными ActiveX-контейнерами (см. стр. 163), которые позволяют эффективно обрабатывать большое число документов.

Документы можно организовать в виде иерархии папок, или узлов документов (по умолчанию, для каждого нового анализа создается своя папка). С использованием дерева просмотра можно легко управлять структурой, изменяя положение отдельных документов, папок или целых ветвей дерева.

Например, выбранные документы можно направить (скопировать или переместить с помощью метода “перетащить и отпустить”) в окно Отчета или в Рабочую область STATISTICA (где они отображаются в виде Автономных окон). При настройке иерархии папок можно различными способами перенести в другие Рабочие книги целые ветви дерева. Это позволяет легко настроить соответствующий порядок папок и документов.

Каждая Рабочая книга содержит два поля: навигационное дерево в стиле Explorer в левой части и поле просмотра файлов в правой части. В навигационном дереве (дереве Рабочей книги) может находиться несколько различных узлов, которые используются для создания логической структуры (например, все результаты анализа или все макросы, созданные в проекте, могут помещаться в отдельную папку).

Вкладки в нижней части поля просмотра документов (поля просмотра Рабочей книги) используются для удобства переходов между документами текущего узла. Вы можете легко разместить эти вкладки сверху, справа или слева от поля просмотра Рабочей книги, щелкнув правой кнопкой мыши на одной из вкладок и выбрав желаемое расположение в контекстном меню. Преимуществом вертикального расположения вкладок является то, что при этом они отображаются в виде нескольких строк, а не одной длинной строки (см. рисунок ниже). Это облегчает выбор нужной вкладки.



Эти вкладки также можно скрыть, увеличив тем самым поле просмотра Рабочей книги.

В отличие от многих приложений в стиле *Explorer*, которые позволяют иметь наследников только папкам, в Рабочих книгах *STATISTICA* любой элемент дерева может иметь своих наследников. Например, вы можете добавить электронную таблицу в вашу Рабочую книгу, а затем добавить все графики, созданные на основе этой таблицы, как дочерние элементы электронной таблицы.

При организации дерева Рабочей книги вы можете использовать стандартный метод “перетащить и отпустить” и любые допустимые операции с буфером обмена.

Рабочая книга может содержать все типы документов системы *STATISTICA*, включая электронные таблицы, графики, Отчеты и макросы. Также элементами Рабочих книг могут быть другие типы ActiveX-документов, такие как таблицы Excel, документы Word и др.

Если вы хотите отредактировать эти документы, вы можете использовать для этого поле просмотра Рабочей книги. Чтобы отредактировать документ Microsoft Word, дважды щелкните на объекте в дереве Рабочей книги. Документ Word откроется в поле просмотра, а меню Рабочей книги будет объединено с меню Microsoft Word.

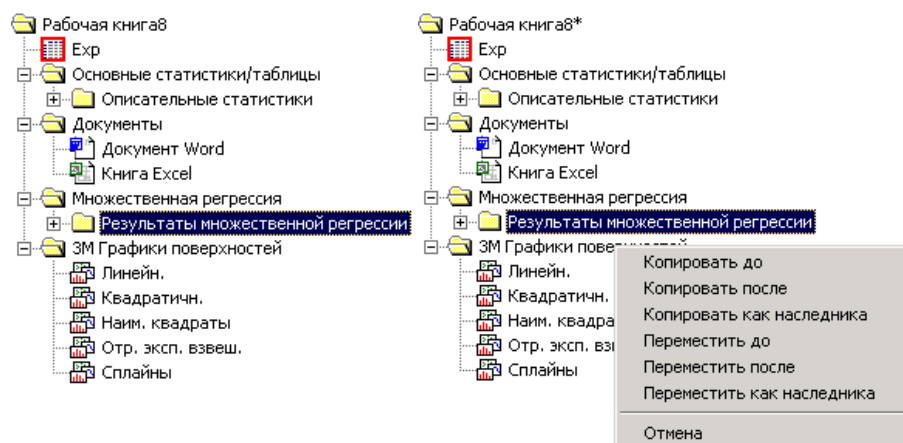
Рабочие книги могут также использоваться для сохранения всех результатов отдельного анализа.

Навигация по дереву Рабочей книги

Дерево Рабочей книги отображает структуру файлов и папок в Рабочей книге. Файлы и папки представляются в стиле *Explorer*. Элементы со знаком “плюс” обозначают папки или файлы, которые имеют дочерние элементы. Чтобы отобразить список элементов отдельной папки или файла, нажмите на соответствующий знак “плюс”.

Рабочая книга может поддерживать неограниченное число уровней.

Как отдельными элементами дерева просмотра, так и целыми ветвями можно гибко (интерактивно) управлять (например, копировать или перетаскивать мышью между Рабочими книгами или Отчетами, как показано далее на рисунке).










Чтобы выбрать элемент Рабочей книги для просмотра или редактирования, просто выберите нужный файл в дереве Рабочей книги и дважды щелкните на его значке. Документ откроется в поле просмотра Рабочей книги.

Вы можете перемещаться по дочерним элементам выделенного узла с помощью вкладок навигации, которые по умолчанию располагаются в нижней части поля просмотра Рабочей книги.

Как было отмечено ранее, вы можете разместить эти вкладки сверху, справа или слева от поля просмотра Рабочей книги. Для этого щелкните правой кнопкой мыши на одной из вкладок и выберите желаемое расположение вкладок в контекстном меню либо выберите соответствующую команду в меню **Рабочая книга - Управление вкладками**.

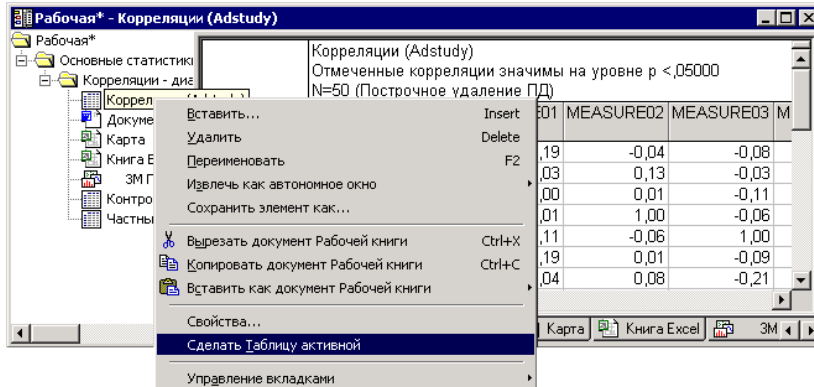
Заметим, что если вкладки находятся вверху или внизу поля просмотра, то они располагаются на одной строке, которую можно прокручивать в стороны. Если же вкладки находятся слева или справа поля просмотра, они располагаются на нескольких строках.

Каждому элементу дерева Рабочей книги соответствует определенный значок. Значок  обозначает папку, которая может содержать ряд документов и подпапок. Значки  (электронная таблица),  (Отчет),  (макрос) и  (график) представляют электронную таблицу, Отчет, макрос и график STATISTICA соответственно.

Документы, не являющиеся документами системы STATISTICA, обозначаются соответствующими их типу значками. Например, документы Word обозначаются значками , а файлы таблиц Excel – значками .

Порядок элементов в дереве Рабочей книги можно изменить, используя метод “перетащить и отпустить”, а также буфер обмена. См. разделы *Метод “перетащить и отпустить”* и *Использование Буфера обмена в Электронном руководстве*.

В контекстном меню дерева Рабочей книги (вызываемом щелчком правой кнопки мыши в любом месте дерева) доступны команды вставки, извлечения, переименования и удаления.



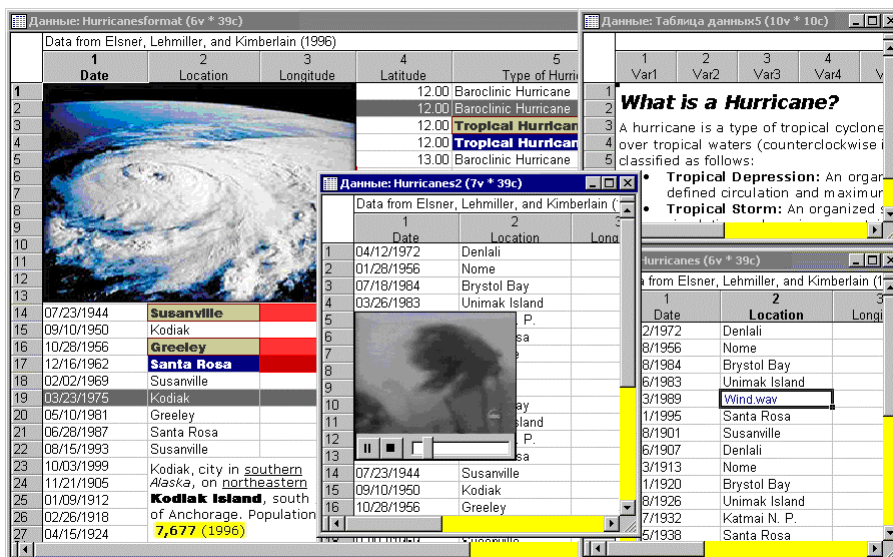
Эти команды также доступны в меню *Рабочая книга*.

Электронные таблицы (мультимедийные таблицы)

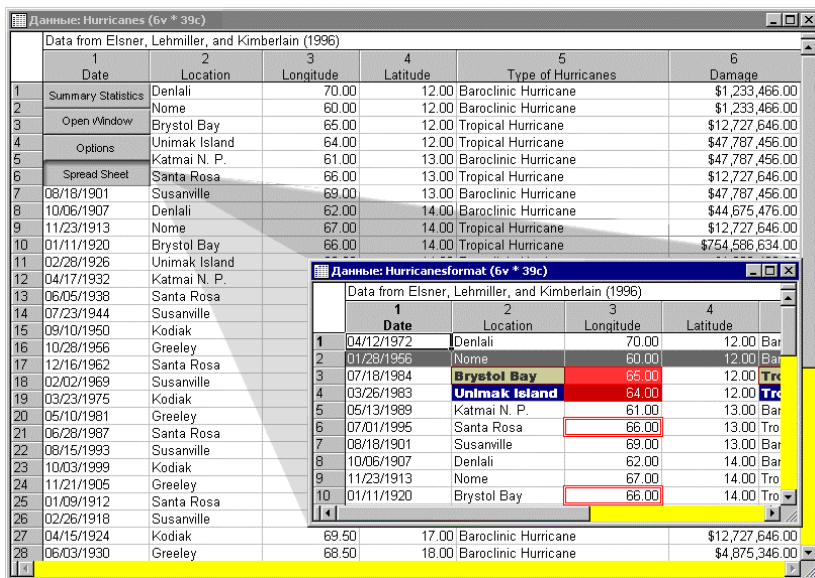
Электронные таблицы *STATISTICA* основаны на технологии мультимедийных таблиц компании StatSoft. Они используются для работы как с исходными данными, так и с численными и текстовыми результатами анализов.

Стандартная электронная таблица – это двумерная таблица, которая может содержать практически неограниченное число наблюдений (строк) и переменных (столбцов). При этом каждая ячейка может содержать практически неограниченное число символов.

В ячейки могут быть также вставлены звук, видео, графики, анимация, Отчеты с внедренными объектами или любые документы, использующие технологию ActiveX.



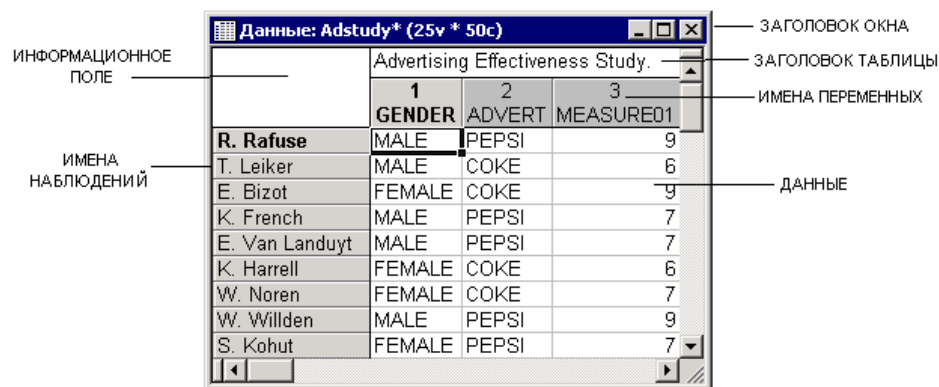
Так как электронные таблицы *STATISTICA* могут содержать макросы и поддерживают настраиваемый пользовательский интерфейс, то эти мультимедийные таблицы можно использовать в качестве оболочки для различных приложений, презентаций, анимаций и т.д. (Например, можно добавить окно списка опций или ряд кнопок в верхнем левом углу.)



Расположение данных в электронных таблицах. Данные в системе STATISTICA организованы в виде набора наблюдений и переменных. Если вы не знакомы с этими понятиями, то попробуйте представить наблюдения как записи программы управления базами данных (или строки электронной таблицы), а переменные – как поля (или столбцы электронной таблицы).


Каждое наблюдение состоит из набора значений переменных, и в первом столбце файла содержатся (необязательно) имена наблюдений.

Окно электронной таблицы содержит несколько основных частей.



Заголовок окна. В *Заголовке окна* отображается название электронной таблицы. Если электронная таблица является исходной таблицей данных, то *Заголовок окна* также содержит число переменных и наблюдений (например, **25v * 50c**). На приведенном выше рисунке в области *Заголовка окна* отображается текст **Данные: Adstudy (25v * 50c)**.

Информационное поле. Для того чтобы выделить всю электронную таблицу, один раз щелкните мышью в нижнем правом углу *Информационного поля*, которое располагается в верхнем левом углу окна электронной таблицы. (При этом указатель должен иметь вид обычной стрелки.)

Для выделения только *Информационного поля* (например, с целью форматирования) один раз щелкните мышью в верхнем левом углу *Информационного поля* (указатель мыши должен быть в виде знака .

Дважды щелкните мышью на *Информационном поле* для ввода в него текста или редактирования содержащегося в нем текста (например, дополнительных пояснений к электронной таблице).


Заголовок таблицы. Поле **Заголовок** расположено непосредственно над именами переменных в верхней части окна. Чтобы выделить это поле (например, для форматирования), щелкните по нему один раз.


Чтобы ввести или отредактировать текстовую информацию, дважды щелкните по этому полю. Нажмите CTRL+ENTER или ALT+ENTER для ввода новой строки (заметим, что иногда необходимо увеличивать высоту поля для того, чтобы видеть несколько строчек текста).

На приведенном выше рисунке поле **Заголовок** содержит текст **Оценка эффективности рекламы**.


Имена наблюдений. Эти ячейки расположены в левой части окна и содержат имена для каждого наблюдения.

Для ввода или редактирования текста дважды щелкните на любом **Имени наблюдения**.

Для выделения **Имени наблюдения** (например, с целью форматирования) один раз щелкните в его левой части (указатель мыши при этом должен иметь вид .

Для того чтобы выделить всю строку наблюдения (для редактирования), один раз щелкните по средней или правой части **Имени наблюдения** (указатель мыши в этом случае имеет вид .

Для выделения блока **Имен наблюдений** (без выделения соответствующих строк) используйте команду **Выбрать только имена наблюдений** в контекстном меню электронной таблицы.



Чтобы автоматически установить подходящую ширину ячеек **Имен наблюдений**, дважды щелкните по правой границе любой из них (указатель мыши в этом случае принимает вид двойной стрелки .


На приведенном выше рисунке ячейки **Имен наблюдений** содержат первые инициалы и фамилии респондентов.


Заметим, что **Имена наблюдений** необязательно присутствуют в окне таблицы данных. Вы можете их скрыть, отключив опцию **Показать имена наблюдений** в меню **Вид**. Если имен наблюдений нет, отображаются номера наблюдений.

Имена переменных. Эти ячейки расположены в верхней части каждого столбца и содержат названия переменных.

Для просмотра спецификаций отдельной переменной дважды щелкните по полю **Имени переменной**.

Для выделения поля **Имени переменной** (например, для форматирования) один раз щелкните по верхней части этого поля (указатель мыши должен при этом иметь вид  вместо .

Для выделения всего столбца переменной (например, с целью редактирования) один раз щелкните по нижней части поля **Имени переменной** (указатель мыши должен при этом иметь вид .

Чтобы автоматически установить подходящую ширину столбцов, дважды щелкните по правой границе **Имени переменной** (указатель мыши в этом случае принимает вид двойной стрелки .

На показанном выше рисунке первые две ячейки **Имен переменных** содержат текст **Gender – Пол** и **Advert – Упоминание**.

Вы можете также изменять способ отображения информации в ячейках **Имен переменных**. Например, можно показать номера столбцов, длинные имена переменных и/или названия типов переменных в электронной таблице. Все эти опции доступны в меню **Вид - Имена переменных**.

Данные (и редактирование внутри ячеек). Остальная область электронной таблицы содержит данные, представленные в виде наблюдений и переменных. Кроме того, в электронную таблицу можно включать любые другие внедренные или связанные объекты (например, мультимедийные объекты, макросы).

Текст в ячейках может быть практически неограниченной длины (обычно в настройках системы *STATISTICA* она ограничена 1000 символами для предотвращения случайной вставки большого объема нежелательной информации в одну ячейку). Имеются широкие возможности для форматирования текста в ячейках.

Исходные таблицы данных и таблицы результатов

STATISTICA предоставляет возможность открывать и использовать одновременно несколько электронных таблиц, что позволяет вам работать одновременно с несколькими входными файлами данных. Кроме хранения данных *STATISTICA* использует электронные таблицы для отображения численных результатов анализов.

Так как *STATISTICA* не делает различий по поддерживаемым возможностям между исходной таблицей данных (из которой *STATISTICA* извлекает данные) и таблицей

результатов (в которой отображаются результаты анализа), то можно использовать результаты одного анализа в качестве входных данных для дальнейших анализов.

Любая электронная таблица, открытая для отображения данных из имеющегося на диске файла, автоматически рассматривается как исходная таблица данных. Одновременно может быть открыто любое число исходных таблиц данных.

Во избежание путаницы таблицу результатов (содержащую результаты анализа) нельзя сразу же использовать для получения исходных данных для другого анализа. Перед тем как ее использовать для дальнейшего анализа, необходимо определить ее как исходную таблицу данных.

Кроме того, для исходных таблиц данных в заголовке окна отображаются количества переменных и наблюдений. Например, **Exp (8v * 48c)** является исходной таблицей данных, а **Exp** – нет.

Для того чтобы использовать значения таблицы результатов в качестве исходных данных, перейдите в соответствующую электронную таблицу и установите опцию **Исходная таблица данных** в меню **Данные**. После этого вы можете начать анализ, и *STATISTICA* будет использовать данные из этой электронной таблицы для анализа.

Заметим, что если вы перейдете к другой электронной таблице, которая ранее была определена как исходная таблица данных, то она также может быть использована для анализа.

В Рабочей книге в любой момент времени может быть выделена для анализа только одна электронная таблица, даже если в Рабочей книге имеется несколько исходных таблиц данных. Такая таблица называется активной исходной таблицей, и ее значок (в дереве Рабочей книги) обведен красной рамкой.

По умолчанию, когда таблица результатов определяется как исходная таблица данных, *STATISTICA* автоматически рассматривает ее как активную исходную таблицу. Для того чтобы выбрать другую исходную таблицу данных в качестве активной, установите опцию **Сделать Таблицу активной** в меню **Рабочая книга** или в контекстном меню дерева Рабочей книги.

Рабочая книга5* - Adstudy

ЭТО АКТИВНАЯ ИСХОДНАЯ ТАБЛИЦА ДАННЫХ

Рабочая книга5*
 Основные стати
 Adstudy
 Описательн
 Описате
 2М Гистограмм
 Гистограмм
 Отчет

	1	2	3	4	5
	GENDER	ADVERT	MEASURE01	MEASURE02	MEASU
K. Record	FEMALE	COKE		4	4
R. McKinney	MALE	COKE		7	0
C. Barrett	MALE	COKE		6	8
J. Fedrick	MALE	PEPSI		5	1
O. Vizquel	FEMALE	PEPSI		5	1
V. Rameriz	FEMALE	PEPSI		7	5
M. Kmiecik	MALE	COKE		3	6
D. McBee	MALE	PEPSI		2	3
J. Tang	FEMALE	COKE		1	7
P. Catron	FEMALE	PEPSI		9	7
M. Howard	MALE	COKE		9	2
L. Fine	MALE	COKE		7	9
C. Howard	MALE	COKE		5	3
D. Slicker	FEMALE	PEPSI		7	6
F. Porvo	FEMALE	PEPSI		6	7
R. Jerin	FEMALE	PEPSI		5	1

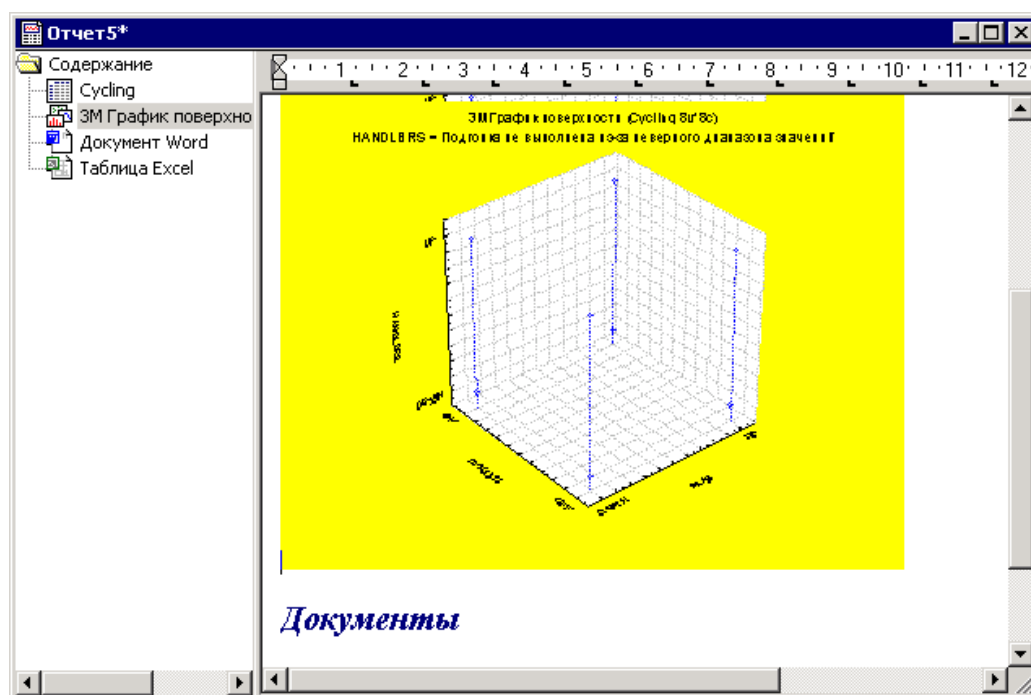
Adstudy Описательные статистики - диалог

Существует также возможность сделать недоступной для анализа уже открытую электронную таблицу. Для этого выберите электронную таблицу и отключите опцию **Исходная таблица данных** в меню **Данные**. Теперь *STATISTICA* автоматически активизирует последнюю таблицу исходных данных, для которой проводился анализ.

Отчеты

Отчеты в системе *STATISTICA* (кратко описанные на стр. 104) предлагают более традиционный способ представления результатов, чем Рабочие книги. Каждый объект (например, электронная таблица или график *STATISTICA* или же таблица Microsoft Excel) отображается последовательно в документе текстового редактора.

Тем не менее, технология создания Отчетов обладает широкими функциональными возможностями. Например, как и Рабочая книга, каждый Отчет системы *STATISTICA* является также ActiveX-контейнером (см. стр. 163). Каждый объект в нем (не только электронные таблицы и графики *STATISTICA*, но также и любые другие документы, поддерживающие технологию ActiveX, например, документы Microsoft Word) является настраиваемым и допускает редактирование непосредственно в окне Отчета.



Отчеты хранятся в формате STR, который является некоторой модификацией известного формата Microsoft RTF (Rich Text Format, *.rtf). Помимо информации в формате RTF файлы STR также содержат описание дерева объектов (которое нельзя сохранить в стандартных файлах RTF). Поэтому файлы Отчетов сохраняются по умолчанию с расширением *.str, но могут быть сохранены также как стандартные файлы RTF (в этом случае информация о дереве объектов не сохраняется).

Очевидными преимуществами такой организации вывода (более традиционного, чем Рабочая книга) являются возможности добавления примечаний и комментариев к объектам, а также быстрой прокрутки и просмотра результатов вывода в таком виде, к какому привыкли пользователи. (Редактор поддерживает изменяющуюся скорость прокрутки окна и другие возможности IntelliMouse.)

Кроме того, только Отчеты содержат и сохраняют дополнительную информацию, включающую в себя подробный журнал заданных для анализа опций (например, выбранные переменные, их метки, длинные имена и т.п.). Объем дополнительной информации указывается во вкладке **Диспетчер вывода** диалогового окна **Параметры** (доступного из меню **Файл - Диспетчер вывода**), см. стр. 56.

Очевидным недостатком традиционных Отчетов является их одноуровневая структура, хотя для некоторых пользователей или приложений такая структура может оказаться предпочтительной.




Навигация по дереву Отчета



Дерево Отчета показывает организацию файлов, содержащихся в данном Отчете. Файлы отображаются в формате *Проводника*. Но в отличие от Рабочих книг, которые поддерживают любое число уровней вложенности, Отчеты поддерживают только один уровень вложенности файлов.

В Отчетах можно выделять большое число элементов для печати, удаления и т.д. При этом используются стандартные для Windows кнопки CTRL и SHIFT.

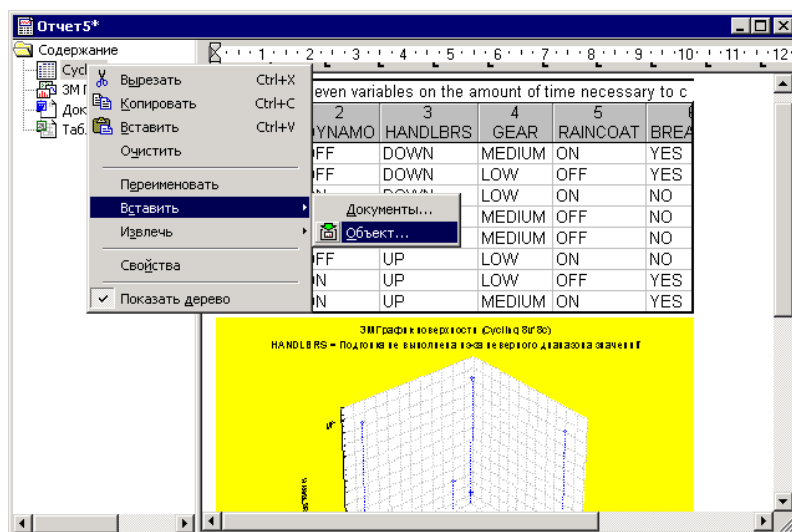
В Отчет можно поместить любой документ *STATISTICA*, включая электронные таблицы и графики. Кроме документов *STATISTICA* вы можете встраивать в Отчет другие типы объектов ActiveX/OLE, в том числе таблицы Excel, документы Word, растровые изображения и др.

Для редактирования документа дважды щелкните по нему мышью. При этом соответствующий файл откроется в окне просмотра, и панель инструментов **Отчет** объединится с панелью инструментов того приложения, которое связано с данным типом файла. Таким образом, вы получите доступ ко всем средствам редактирования документа.

Каждому элементу дерева соответствует определенный значок. Значки  (электронная таблица),  (макрос) и  (график) обозначают электронную таблицу, макрос и график *STATISTICA* соответственно.

Документы, не являющиеся документами системы *STATISTICA*, обозначаются соответствующими их типу значками. Например, документы Word обозначаются значками , а файлы таблиц Excel – значками .

При организации дерева Отчета можно воспользоваться методом “перетащить и отпустить”, а также буфером обмена.



В контекстном меню дерева Отчета (вызываемом щелчком правой кнопкой мыши в любом месте дерева, см. рисунок выше) доступны команды вставки, извлечения, переименования и удаления элементов дерева Отчета.

Графики

Система *STATISTICA* предоставляет широкий спектр возможностей по созданию графических документов, а также их настройке.

Как и другие документы *STATISTICA*, графики являются ActiveX-контейнерами (см. стр. 163). Это означает, что они могут содержать в себе ряд совместимых документов (например, рисунки Visio, иллюстрации Adobe, таблицы Excel и т.п.).

Графики *STATISTICA* являются также ActiveX-объектами. Следовательно, они могут быть связаны или внедрены в другие совместимые документы (например, документы Word), где их можно редактировать, дважды щелкнув по графику мышью.

Графики более подробно описаны в Главе 5 – *Графики*.

Макросы (программы на языке SVB)

Язык программирования SVB (интегрированный в систему STATISTICA) предоставляет намного больше возможностей, чем просто “вспомогательный язык программирования”, который используется для создания пользовательских приложений.

SVB использует огромные преимущества объектно-ориентированной структуры системы STATISTICA и позволяет получить доступ программными средствами практически ко всем функциональным деталям пакета STATISTICA.

Даже самые сложные процедуры анализа и графический вывод результатов могут быть записаны в виде макросов на языке Visual Basic, которые затем можно воспроизвести или заново отредактировать.

```

[interest.svb|SimpleInterest#148] InterestEarned# = (Return@ - Principal@)
[interest.svb|RunInterest# 51] SimpleInterest
[interest.svb|Main# 38] Dialog dlg

Object: [General] Proc: SimpleInterest

s.SetSize(PeriodCount,1)
s.CaseNameLength = 18
s.InfoBox.Value = "Simple interest over" + Str(PeriodCount) + " rollo
s.InfoBox.AutoFit

s.Variables(1).NumberFormat = "$0.00"
s.DisplayAttribute(scDisplayVariableNumbers) = False

Dim YearValue As Double
Dim InterestEarned As Double
YearValue = 1 / PeriodsPerYear
'A = P(1 + rt)
Return = Principal * (1 + (Rate * YearValue))
'because the interest is simple, then the interest will remain consta
InterestEarned = (Return - Principal)

For i = 1 To PeriodCount
  If i = 1 Then
    'already calculated the first return
    s.Value(1,1) = Return
    'write to report
    TextString = TextString & "Total after rollover 1: " & Forma
  Else
    'A = P(1 + rt)
    s.Value(1,1) = (Return + (InterestEarned * (i - 1)))
    'write to report
    TextString = TextString & "Total after rollover " & i & ": "
  s.Case(i).RowName = "Rollover " + Str(i)
End If
    
```

Макросы представляют собой самостоятельные блоки, которые легко встраиваются в другие приложения.

SVB добавляет более 10 000 новых функций к стандартному синтаксису Microsoft Visual Basic и является, таким образом, одним из самых функционально богатых и обширных интерфейсов прикладного программирования.

Макросы *STATISTICA* могут быть сохранены в различных форматах, в зависимости от того, как вы намереваетесь их использовать (см. *Руководство по STATISTICA Visual Basic* и *Электронное руководство* для получения дополнительной информации).

Вы можете также скопировать макросы в буфер обмена и вставить их в другие программы или документы.

Более подробно язык *SVB* описан в Главе 7 (стр. 155).

6

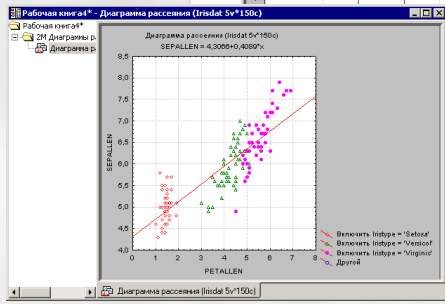
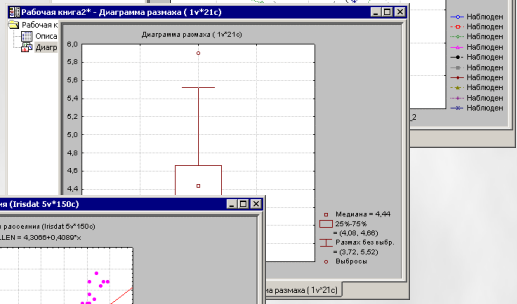
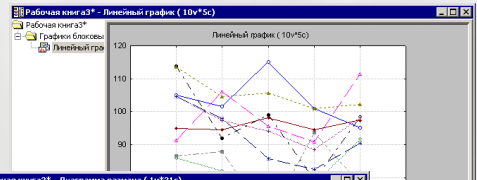
ГЛАВА

ГРАФИКИ

Обзор.....	127
Настройка параметров графиков.....	128
Создание графиков с помощью <i>SVB</i>	133
Благодарности, ссылки, стандарты	135
Основные типы графиков.....	136
Графики исходных данных.....	137
Графики блоковых данных	140
Графики из меню Графика	142
Другие специализированные графики	144

- Гистограммы...
- Диаграммы рассеяния...
- Графики средних с ошибками...
- Диаграммы размаха...
- Диаграммы диапазонов...
- Пиктограммы рассеяния...
- Диаграммы рассеяния с образами...
- Диаграммы рассеяния с гистограммами...

- Гистограммы...
- Диаграммы рассеяния...
- Графики средних с ошибками...
- Диаграммы размаха...
- Нормальные вероятностные графики...
- Графики квантиль-квантиль...
- Графики веродности-вероятность...
- Линейные графы
- Линейные графы
- Последовательные
- Круговые диаграммы
- Графики пропускания
- Графики функций
- 3D XYZ графики
- 3D Тернарные графики
- Категоризованные графики
- Категоризованные графики
- Графики функций
- Диаграммы исходных данных...
- Гистограммы двух переменных...
- Диаграммы диапазонов...
- Диаграммы размаха...



Графики

STATISTICA

3M Графики функций пользователя

Пользовательская функция Вид | Параметры 1 | Параметры 2

Имя: Новая функция

Тип: Функция

Диапазон изменения X: Мин: 0, Макс: 10

Функция: Y =

3M Гистограммы

Быстрый | Дополнительно | Вид | Категории | Параметры 1 | Параметры 2

Тип графика: Вертикальная

Распределение: Нормальное

Интервалы: Переменные

Ось Y: N

- Продолжить... Ctrl+R
- Гистограммы...
 - Диаграммы рассеяния...
 - Графики средних с ошибками...
 - Графики поверхностей...
 - 3D Графики
 - 3D Последовательные графики
 - 3D XYZ графики
 - Матричные графики...
 - Пиктограммы...
 - Категоризованные графики
 - Графики пользователья
 - Графики блоковых данных
 - Графики исходных данных
 - Размещение нескольких графиков

3M Гистограммы двух переменных

Быстрый | Дополнительно | Вид | Параметры 1 | Параметры 2

Переменные: X: нет, Y: нет

Интервалы для X: Переменные: нет

Интервалы для Y: Переменные: нет

3M Графики поверхностей

Быстрый | Вид | Категории | Параметры 1 | Параметры 2

Подгонка: Линейная, Квадратная, Полиномиальная, Логарифмическая, Экспоненциальная, Ничья квадрат, Отр. экстр. точка, Сплайны, Робастная регресс.

3M Матричные графики

Быстрый | Дополнительно | Вид | Параметры 1 | Параметры 2

Тип графика: Квадр. матрица рассеяния, Квадр. матрица линей, Прямоуг. матрица рассеяния, Прямоуг. матрица линей

Подгонка: Линейная, Полиномиальная, Логарифмическая, Экспоненциальная, Ничья квадрат, Отр. экстр. точка, Сплайны, Робастная регресс.

ГРАФИКИ

Обзор

Во многих случаях использование графического представления информации позволяет более эффективно работать с числовыми данными. Система *STATISTICA* предлагает большой выбор графических методов, содержащих программные средства для обработки данных и проверки гипотез.

Широкий выбор графических опций. Система *STATISTICA* включает широкий спектр графических методов для визуального представления результатов исследований. Все графические средства системы *STATISTICA* обеспечивают возможность выбора встроенного аналитического интерактивного метода анализа и содержат большой набор программ настройки, позволяющих пользователю интерактивно управлять отображением информации на экране.

Гибкие средства управления одновременно несколькими графиками дают возможность пользователю объединять различные графические изображения и осуществлять динамический обмен между приложениями (используя средства OLE)

Комплексная поддержка языка Visual Basic и других языков программирования. Параметры графиков *STATISTICA* можно изменить с помощью встроенного языка *SVB* или любого другого совместимого языка программирования. Таким образом, пользователю предоставляются практически неограниченные возможности настройки графических изображений.

Пользовательские графики впоследствии могут быть внедрены в пользовательский интерфейс системы *STATISTICA* (например, они могут быть помещены на кнопки панелей инструментов или добавлены в меню).

Основные типы графиков. В программном пакете *STATISTICA* предлагаются разнообразные графические методы, с помощью которых пользователь может запрашивать или самостоятельно организовывать построение графиков.

Эти методы предоставляют большой набор графиков различных типов, таких как графики исходных данных, графики блоковых данных и специализированные графики. Они рассматриваются в разделе *Основные типы графиков* на стр. 136.

Графические методы дополняют друг друга, обеспечивая высокий уровень взаимосвязи между числовыми данными (начальными, промежуточными и конечными) и их графическим представлением.

Например, в качестве элемента автоматического вывода в программе статистической обработки может быть введен запрос на построение специализированных графиков. Кроме того, фактически любые данные (и/или метки), которые создаются в процессе работы программы, могут быть выведены в графическом виде с помощью встроенных сервисных программ.

Настройка параметров графиков

Интерактивная настройка графиков. Опции настройки графиков *STATISTICA* охватывают сотни свойств и инструментов для корректировки любой детали отображения графиков и обработки данных.

Эти опции организованы в виде иерархической структуры. Наиболее часто используемые опции доступны из контекстных меню, вызываемых с помощью двойного щелчка мыши или щелчка правой кнопкой мыши на соответствующем элементе графика.

Постоянные установки и опции автоматизации. Начальные (принятые по умолчанию) установки всех свойств графиков можно легко отрегулировать. Таким образом, даже вид и свойства графиков по умолчанию будут удовлетворять вашим запросам либо будут требовать минимального вмешательства с вашей стороны.

Существует как минимум четыре способа осуществления таких настроек:

- 1. Диалоговое окно Параметры.** Самым простым способом, пожалуй, является задание соответствующих параметров по умолчанию во вкладках *Графика 1* и *Графика 2* диалогового окна *Параметры* (доступного из меню *Сервис - Параметры*).

Сделанные здесь изменения отразятся на стилях по умолчанию (см. пункт 2 ниже), используемых системой STATISTICA, и будут автоматически сохранены в конфигурационном файле системы (например, в разных проектах можно задавать различные установки).

См. также описание вкладки **Диспетчер конфигураций** диалогового окна **Параметры** в *Электронном руководстве*.

2. Набор графических стилей. Все многочисленные параметры, определяющие вид графика (от элементарных, таких как цвет текста в сносках, до общих, таких как глобальные свойства графического документа), могут быть сохранены как отдельные “стили”.

Этим стилям можно присваивать свои имена и затем снова применять их (нажав заданную комбинацию клавиш или кнопку на панели инструментов).

Система может управлять тысячами стилей и их комбинаций, тем самым давая вам возможность осуществлять желаемые настройки с минимальными усилиями.

Все определенные пользователем или измененные стили автоматически сохраняются в конфигурационном файле (например, различные наборы или системы стилей могут использоваться для различных проектов).

См. также описание вкладки **Диспетчер конфигураций** диалогового окна **Параметры** в *Электронном руководстве*.

3. Графики пользователя. Новые типы графиков могут быть созданы различными способами и затем добавлены в меню, диалоговые окна или на панели инструментов.

Если пользовательский график, который вы хотите повторно использовать, не является построенным “с нуля” (например, с помощью графических примитивов), а основан на одном из графиков меню **Графика**, то можно легко добавить этот пользовательский график в данное меню как новый тип графиков.

Для этого просто нажмите кнопку **Добавить в меню как пользовательский график** во вкладке **Параметры 2** диалогового окна создания графика.

Все параметры пользовательского графика автоматически записываются в конфигурационный файл системы STATISTICA (например, в разных проектах можно использовать различные наборы пользовательских графиков).

См. также описание вкладки **Диспетчер конфигураций** диалогового окна **Параметры** в *Электронном руководстве*.

4. **Язык SVB.** Заметим, что язык *SVB* (со всеми его инструментами рисования и библиотекой графических процедур, ориентированных на систему *STATISTICA*) можно использовать для вывода практически любых графических или мультимедийных результатов.

Полученные изображения или мультимедиа могут быть внедрены на панели инструментов *STATISTICA*, в меню или диалоговые окна.

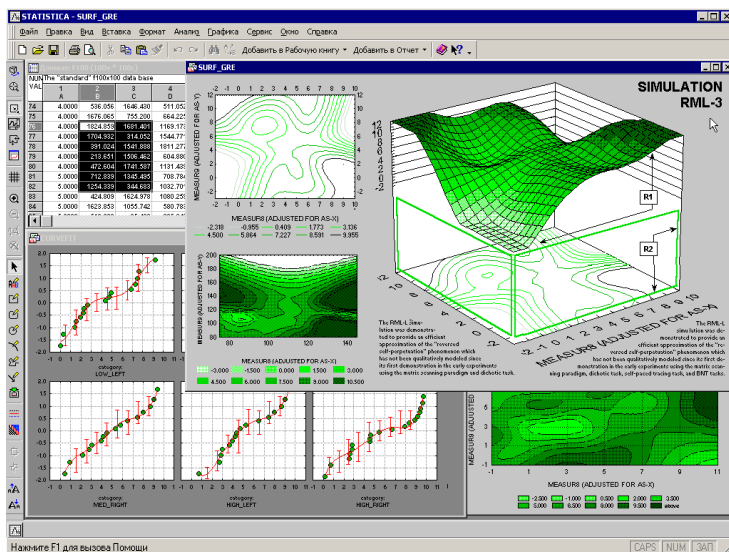
Дополнительную информацию о методах настройки графиков см. в *Электронном руководстве*.

Как было отмечено ранее, существуют различные методы создания графиков в системе *STATISTICA*. Можно сказать, что эти методы представляют собой различные типы “интерфейсов” (связующих элементов) между числовыми данными и их графическим представлением.

На круговой диаграмме числа могут просто отображать значения из определенной колонки таблицы (например, переменной **Продажи**) для последовательных наблюдений этой таблицы (например, наблюдений **Год 2002**, **Год 2003**, **Год 2004** и т.д.).

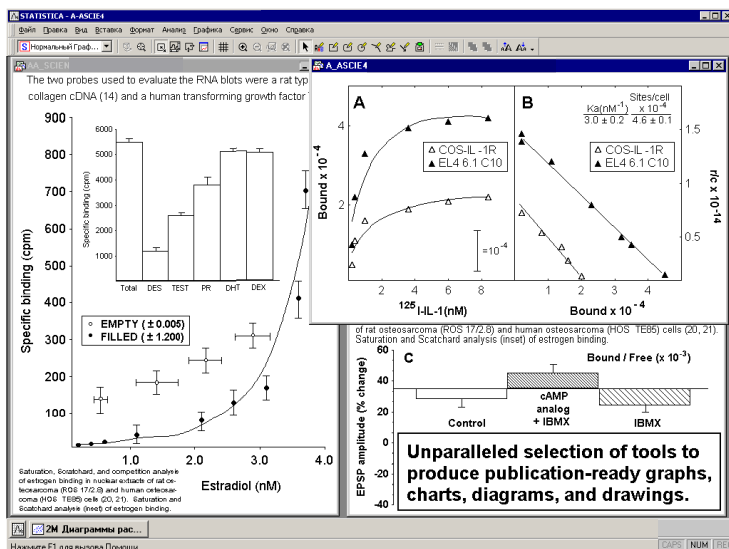
Также числа на круговой диаграмме могут представлять собой результаты некоторых расчетов. К примеру, слои диаграммы могут отражать относительные частоты наблюдений по группам (например, число лет, когда величина переменной **Продажи** была меньше 10 млн. долларов, между 10 и 20 млн. долларов и выше 20 млн. долларов).

Независимо от того, какой метод был использован для создания графика (т.е. откуда были взяты или каким образом вычислены значения, представленные на графике), для изменения его параметров или объединения его с другими документами могут быть использованы все графические настройки и средства управления несколькими графиками системы *STATISTICA*.

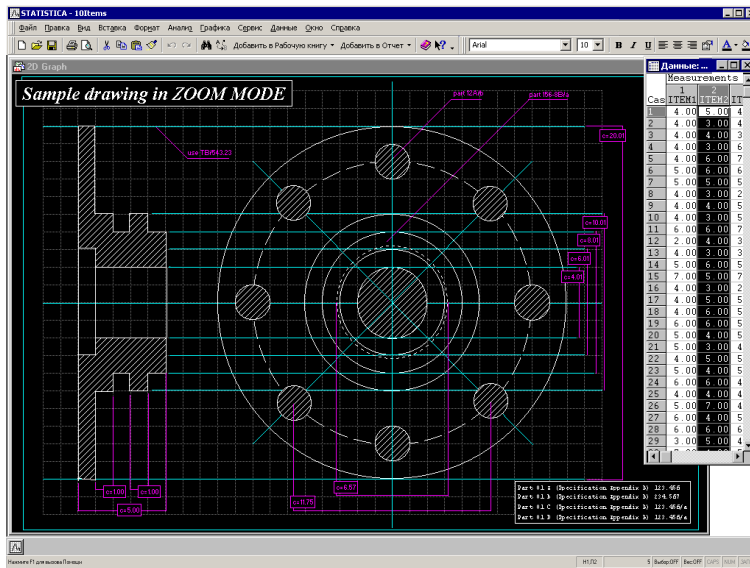


Все встроенные аналитические средства, доступные из графического окна системы *STATISTICA* (такие как подгонка функций, сглаживание, сжатие, расслоение, вращение, закрашивание и др.), могут применяться независимо от источника данных и способа построения графика.

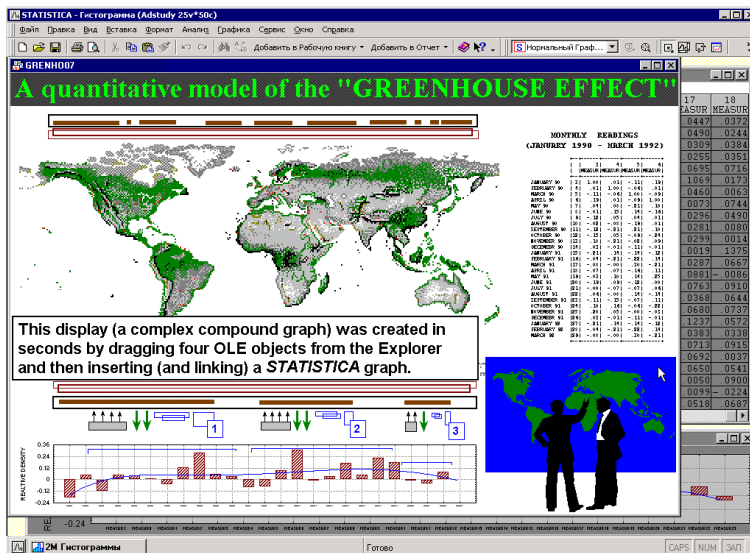
Средства редактирования графиков, предлагаемые системой *STATISTICA*, дают возможность пользователю создавать не только специализированные научные и технические готовые к публикации рисунки



и точные чертежи,



но также диаграммы, плакаты, карты и другие материалы,



которые обеспечивают эффективное и наглядное представление информации.

Графики, записанные в файл или извлеченные из приложения STATISTICA каким-то другим способом (например, скопированные в буфер обмена или присоединенные к

другому приложению), остаются полноценными объектами ActiveX (см. стр. 163). Они содержат не только все пользовательские функции и другие вложенные объекты, но также и все данные, необходимые для редактирования изображения или изучения его частей (подгонки, сглаживания и т.д.).

Так как графики системы *STATISTICA* являются объектами ActiveX, они могут быть связаны или внедрены в другие совместимые документы (например, документы Excel или Word), где их можно редактировать, дважды щелкнув по графику мышью.

Графики *STATISTICA* являются также контейнерами ActiveX и, следовательно, могут содержать в себе широкий набор внедренных и связанных документов, таких как рисунки Visio, иллюстрации Adobe, таблицы Excel или документы Word.

Более того, *STATISTICA* поддерживает иерархическую структуру внедренных объектов до четырех уровней вложения, что означает возможность создания “документов, содержащих документы, содержащих документы, которые содержат документы”.

Создание графиков с помощью *SVB*

Ко всем графическим возможностям системы *STATISTICA* можно получить доступ с помощью встроенного языка *SVB* или других совместимых языков.

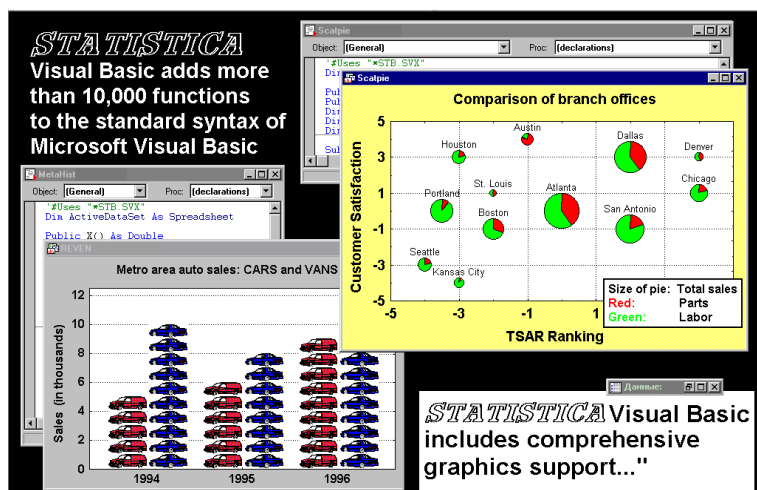
Язык *SVB* (со всеми его инструментами рисования и библиотекой графических процедур, ориентированных на систему *STATISTICA*) можно использовать для вывода практически любых графических или мультимедийных результатов.

Полученные изображения или мультимедиа могут быть внедрены на панели инструментов *STATISTICA*, в меню или диалоговые окна.

В языке *SVB* предусмотрено три способа работы с графиками:

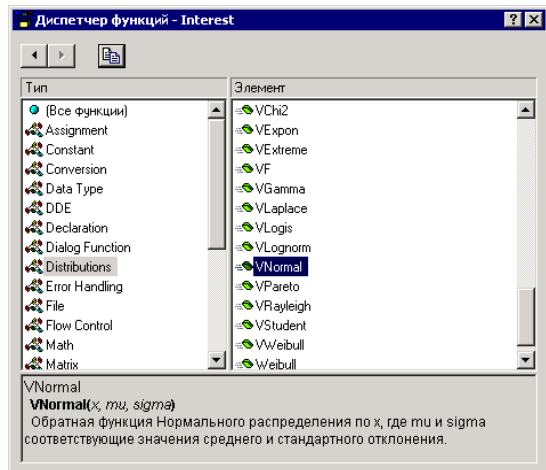
- создать новый график и затем изменить его, распечатать или сохранить и т.д.;
- получить доступ к уже существующему графическому окну для его изменения;
- открыть существующий графический файл и затем изменить его, вывести на печать, сохранить измененный файл и т.д.

Все графики, доступные в системе *STATISTICA*, могут быть созданы с помощью языка *SVB* и затем настроены с использованием процедур системы *STATISTICA* или общих возможностей, предлагаемых этим комплексным языком.



Как и все остальные функции языка *SVB*, функции доступа к графической библиотеке можно легко включить в программу на языке *SVB* с помощью иерархически организованного **Мастера функций**.

Он содержит краткое описание всех функций и средства для вставки нужной функции непосредственно в исходный код вашей программы (т.е. в редактор *SVB*, см. стр. 160).



Для получения дополнительной информации о работе с графическими библиотеками системы *STATISTICA* из программ на языке *SVB* см. *Электронное руководство*.

Благодарности, ссылки, стандарты

Электронное руководство содержит разделы, посвященные отдельным типам графиков. Они включают обзоры основных понятий и примеры типичного применения графиков. В них также дано описание отличительных свойств соответствующих типов графиков.

Более подробное введение в науку (и искусство) графического представления данных можно найти в многочисленных работах, посвященных этому вопросу. Среди источников, наиболее полно охватывающих спектр проблем, связанных с построением графиков, можно рекомендовать следующие: Вуја и Тукей (1991), Chambers, Cleveland, Kleiner и Tukey (1983), Cleveland (1984, 1985), Kolata (1984), Tufte (1983, 1990), Tukey (1977), а также Тукей и Тукей (1981).

Большое число статей по специальным вопросам, связанным с выбором графического представления статистических данных, можно найти в сборнике докладов *Ежегодного семинара отделения статистической графики Американской статистической ассоциации*.

Параметры по умолчанию для большинства статистических графиков, предлагаемых в системе *STATISTICA*, соответствуют установленным в настоящее время соглашениям. Эти соглашения описаны в литературе по построению статистических и технических графиков или же отвечают общепринятым стандартам, обычно используемым в большинстве научных журналов (например, в журнале *SCIENCE*).

В то же время практически любые устанавливаемые по умолчанию параметры системы *STATISTICA* могут быть настроены, если необходимо, чтобы они соответствовали каким-то специальным требованиям (см. стр. 128).

Средства построения графиков в системе *STATISTICA* являются гибкими “инструментами”, позволяющими выйти за рамки готовых шаблонов.

Более того, эти инструменты могут быть настроены, к ним могут быть добавлены новые, и все они могут быть размещены на панелях инструментов или в меню для многократного использования (см. стр. 128).

Помимо широкого набора стандартных статистических и технических графиков система *STATISTICA* предлагает большое число уникальных типов графиков, а также поддерживает возможности их настройки.

Нужно сказать, что во время разработки большинства этих графиков сотрудники компании StatSoft вели постоянный диалог с тысячами пользователей. Последняя версия

графических опций системы *STATISTICA* представляет собой результат этого взаимодействия. Были учтены замечания и “перечни пожеланий”, подготовленные пользователями в ответ на запросы сотрудников компании StatSoft.


Многие уникальные графические средства, предлагаемые системой *STATISTICA* (например, многократные вложенные выборки, интерактивная категоризация и др.), были введены в нее по предложениям пользователей.

Компания StatSoft выражает им благодарность за этот вклад в разработку программы.

Основные типы графиков

Помимо специализированных статистических графиков, которые можно построить с помощью опций итоговых диалоговых окон любой статистической процедуры (см. стр. 144), существуют два основных типа, или класса, графиков:

- Графики исходных данных (**Графики исходных данных**, см. стр. 137, и графики из меню **Графика**, см. стр. 142);
- **Графики блоковых данных** (см. стр. 140).

Они могут быть вызваны из меню **Графика**, панели инструментов **Графика**, контекстных меню и меню стартовой кнопки системы *STATISTICA* .

Основное различие между этими типами графиков заключается в представлении данных, используемых для их построения.

Графики исходных данных. **Графики исходных данных** и их расширенная версия в меню **Графика** являются представлением исходных данных из текущей таблицы данных. Обычно на них представляются переменные или подмножества их значений, если используются условия выбора наблюдений.

Заметим, что если графики этого типа получены с помощью контекстного меню таблицы результатов, которая не содержит фактических данных (например, корреляционной матрицы), то *STATISTICA* все равно будет использовать исходные данные для построения графика. Например, если вы вызовете контекстное меню из ячейки корреляционной матрицы, то будет построен график для переменных, определяемых выделенной ячейкой.

Графики блоковых данных. **Графики блоковых данных**, в свою очередь, совершенно не зависят от “исходных данных” или “файла данных”. Они предоставляют возможность

визуализации числовых значений выделенного в данный момент блока любой электронной таблицы.

Выделенный блок может содержать значения как определенных пользователем подмножеств численных результатов, так и значения произвольно выбранных подмножеств исходных данных.

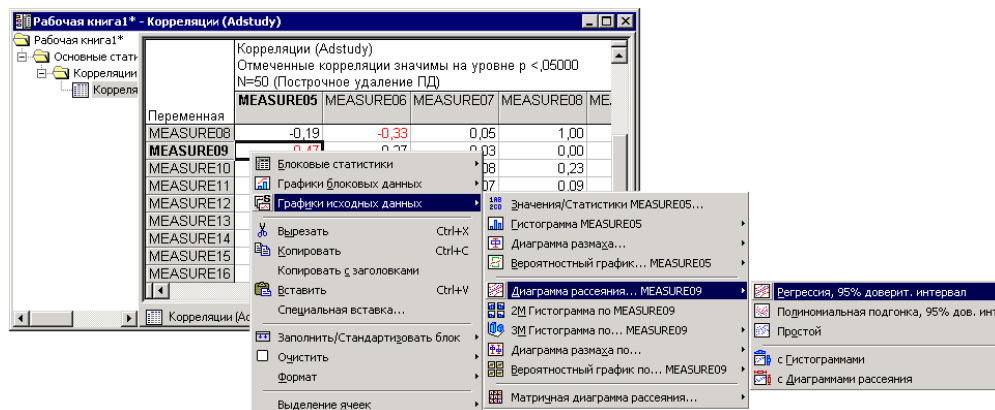
Общие черты двух типов графиков. Заметим, что основные типы графиков используют одинаковые опции для настроек и одинаковые виды графиков.



Например, создать специализированный категоризованный тернарный график можно и на основе исходного набора данных, и на основе заданного блока значений из таблицы результатов.

Эти два основных типа графиков подробно обсуждаются в последующих двух параграфах, за которыми следует параграф о меню **Графика**. Данное меню содержит полный набор всех графиков исходных данных (на которые обычно ссылаются как на графики из меню **Графика**), но также предоставляет доступ к **Графикам блоковых данных** и другим опциям.

Графики исходных данных

Команда **Графики исходных данных** доступна из контекстных меню всех электронных таблиц и предоставляет быстрый доступ к наиболее часто используемым типам графиков, основанных на текущем наборе исходных данных.




Заметим, что все эти графики доступны также из меню **Графика**, из меню стартовой кнопки системы *STATISTICA*  на панели состояния или с помощью кнопки **Галерея графиков**  в любом диалоговом окне задания графика.

Графики исходных данных не предоставляют такой широкий спектр возможностей, как соответствующие графики из меню **Графика**. Однако их проще и легче вызвать, чем графики из меню **Графика**:

- **Графики исходных данных** могут вызываться прямо из контекстных меню электронных таблиц,
- **Графики исходных данных** не требуют от пользователя выбора переменных (этот выбор определяется текущим положением курсора в электронной таблице),
- **Графики исходных данных** не требуют промежуточной настройки параметров (формат соответствующих графиков определяется по умолчанию).

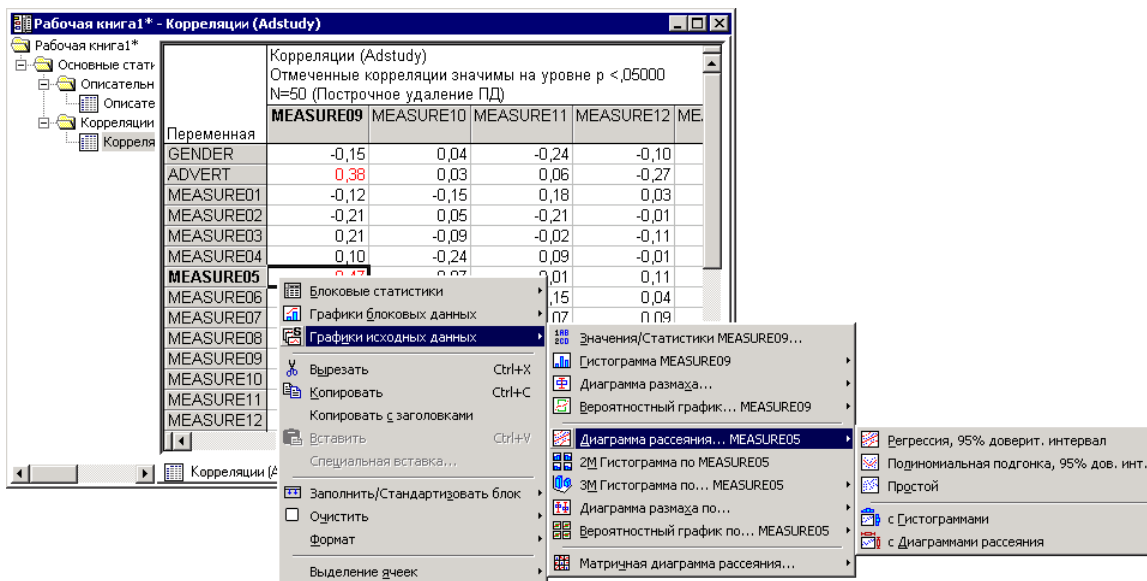
Графики исходных данных используют данные непосредственно из текущего исходного файла данных, и могут только попросить вас выбрать переменные, соответствующие текущей позиции курсора (в электронной таблице любого типа).

Например, если вы щелкнете правой кнопкой мыши на значении коэффициента корреляции в электронной таблице результатов и воспользуетесь командой **Диаграмма рассеяния...**, то *STATISTICA* создаст 2М диаграмму рассеяния, используя исходные значения двух переменных, к которым относится выбранный коэффициент корреляции (см. также *Пример 1* на стр. 45).

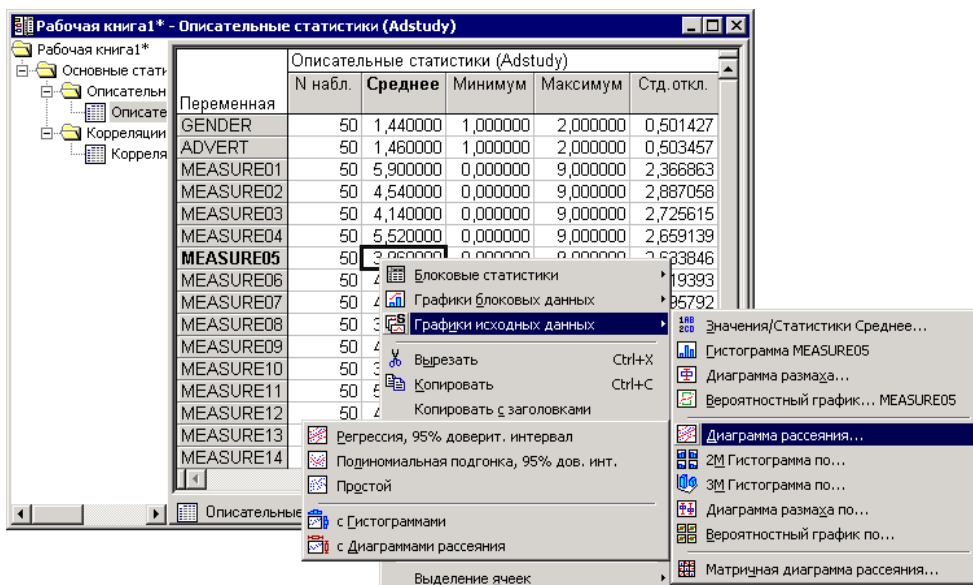
Наиболее простой способ вызова **Графиков исходных данных** предоставляет контекстное меню электронной таблицы. Но вы также можете воспользоваться командами меню **Графика** или командами меню стартовой кнопки системы *STATISTICA* .

В любом случае на экране появится подменю, из которого можно выбрать один из статистических графиков, применимых к текущей переменной (т.е. переменной, на которую в данный момент указывает курсор в электронной таблице).

Если электронная таблица представляет собой матрицу или имеет такой формат, при котором курсор может указывать не только на одну, но и на две переменные (как на рисунке ниже), то в меню **Графики исходных данных** появляется возможность построения двумерного графика для выбранной пары переменных.





В противном случае (когда курсор указывает только на одну переменную, как в таблице описательных статистик, показанной ниже), при построении любого из двумерных графиков система *STATISTICA* попросит вас указать вторую переменную. Например, в случае, который иллюстрирует показанный ниже рисунок, если вы выберете команду **Диаграмма рассеяния...**, то *STATISTICA* запросит у вас имя второй переменной.



Если выделенная область соответствует более чем одной переменной (т.е. выделен блок данных), то **Графики исходных данных** будут построены для первой выделенной переменной.

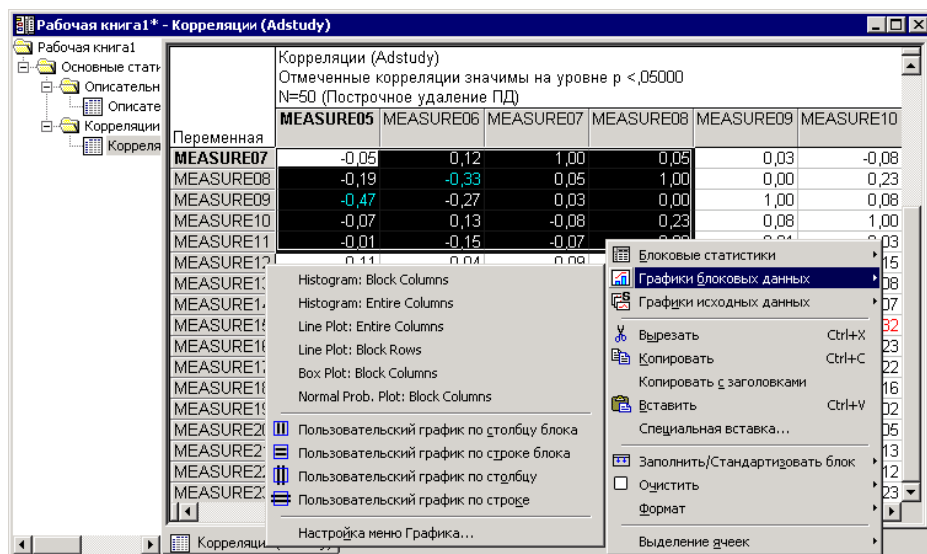
При создании **Графиков исходных данных** система *STATISTICA* учитывает текущие условия выбора и веса наблюдений.

Однако заметим, что условия выбора наблюдений или весов должны быть заданы для текущей электронной таблицы (с помощью команд меню **Сервис - Условия выбора - Правка** и **Сервис - Веса**), а не только для отдельного анализа (в соответствующих диалоговых окнах задания анализа/графика с помощью кнопок  и ). Последние условия не будут учитываться при построении **Графиков исходных данных**.

Для получения дополнительной информации о различных **Графиках исходных данных** см. *Электронное руководство*.


Графики блоковых данных

В отличие от **Графиков исходных данных**, **Графики блоковых данных** используют в качестве исходных данных выделенный (непрерывный) блок данных в активной таблице.



Заметим, что эти графики абсолютно не зависят от типа “исходных данных”. Они обрабатывают значения (числа) любого выделенного блока и не учитывают “смысл” этих чисел (например, числа могут представлять собой исходные данные или значения коэффициентов корреляции).

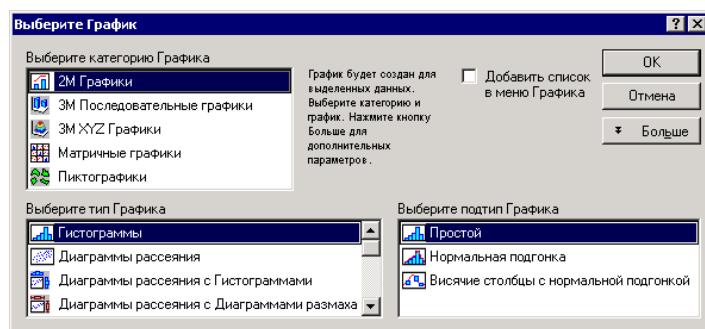
Эти графики предоставляют эффективные средства визуализации, изучения и обобщения численного вывода анализа, отображаемого в таблицах результатов (например, гистограммы результатов расчета по методу Монте-Карло в модуле **Моделирование структурными уравнениями** или диаграмма размаха агрегированных средних из многомерной таблицы множественной классификации в модуле **Дисперсионный анализ**).

Самый простой (и, можно сказать, самый логичный) способ вызова **Графиков блоковых данных** предоставляет контекстное меню, связанное с выделенным блоком в электронной таблице. Но вы также можете воспользоваться меню **Графика** или меню стартовой кнопки системы **STATISTICA** .

При создании **Графиков блоковых данных** вы можете выбрать предлагаемые графики (например, **Гистограмма: по столбцам блока** или **Линейный график: по строкам блока**). Кроме того, вы можете создать свои собственные пользовательские графики как для выделенных ячеек в строках и столбцах, так и для всех ячеек в строках или столбцах, в которых выделен блок.

Графики по умолчанию. Вы можете создать графики по умолчанию (первые шесть команд на показанном выше рисунке) одним щелчком мыши. Подробное описание этих графиков см. в *Электронном руководстве*.

Пользовательские графики. При выборе одного из четырех **Пользовательских графиков** появится диалоговое окно **Выборите График**, которое предоставляет набор опций для создания настраиваемых графиков.




Более детальную информацию о пользовательских графиках см. в *Электронном руководстве*.

Настройка графиков. Как и большинство параметров системы *STATISTICA*, **Графики блоковых данных** полностью настраиваются.

В меню **Графики блоковых данных** выберите команду **Настройка меню Графика** для отображения одноименного диалогового окна. В этом окне содержатся опции для удаления, переименования и редактирования графиков из текущего списка, а также для добавления новых (заданных пользователем) графиков в меню **Графики блоковых данных**.

Например, если вы хотите включить нормальную подгонку в гистограммы, создаваемые с помощью команды **Гистограмма: по столбцам блока**, выберите соответствующую строчку в диалоговом окне **Настройка меню Графика**, нажмите кнопку **Изменить** и измените **Подтип Графика** на **Нормальную подгонку**. Все последующие графики, создаваемые с помощью команды **Гистограмма: по столбцам блока**, будут включать в себя нормальную подгонку к данным.

Графики из меню Графика

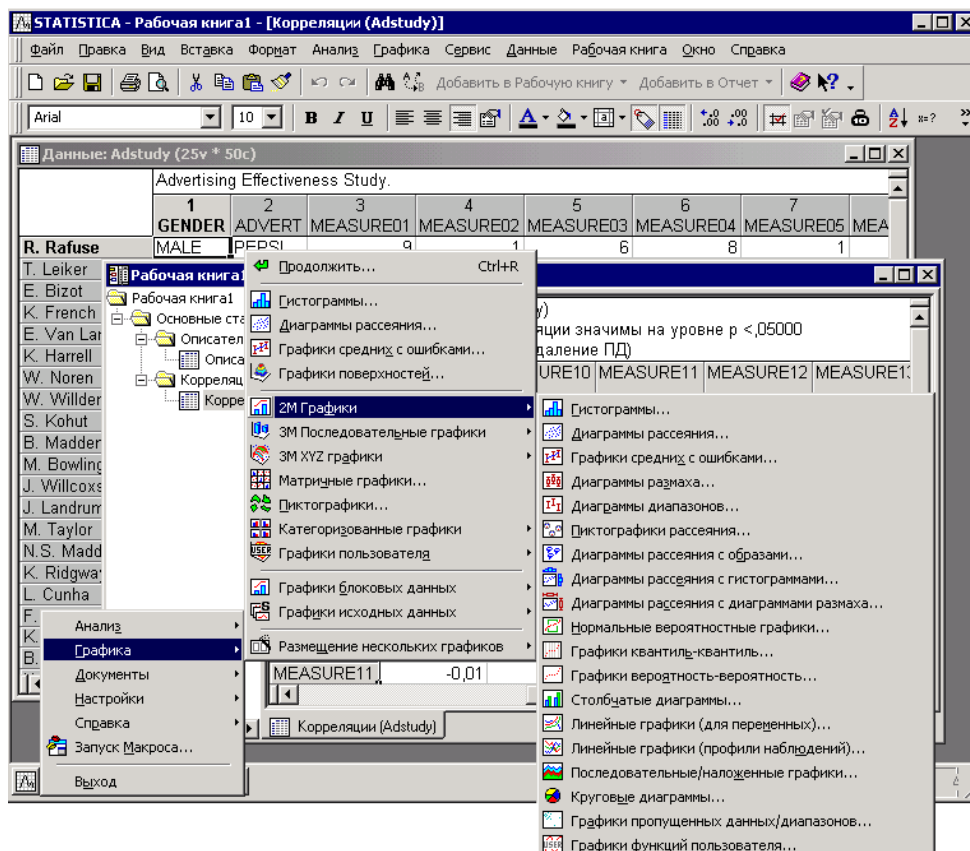
Команды в меню **Графика** предоставляют полный набор статистических графиков и все вспомогательные настройки системы *STATISTICA*. Эти команды присутствуют не только в меню **Графика**, но также в меню стартовой кнопки системы *STATISTICA* .

Заметим, что за исключением **Графиков блоковых данных** (которые также включены в это меню), все остальные типы графиков в меню **Графика** не ограничены значениями текущей таблицы результатов.

Данные для построения этих графиков берутся прямо из текущей таблицы исходных данных, так же как и в случае построения **Графиков исходных данных** (которые были описаны выше).

При этом предлагаются как стандартные методы для графического представления исходных данных (например, различные диаграммы рассеяния, гистограммы, графики средних значений), так и стандартные графические аналитические методы исследований (например, категоризованные нормальные вероятностные графики, вероятностные графики с исключенным трендом или графики доверительных интервалов линий регрессии).

При создании таких графиков система *STATISTICA* учитывает текущие условия выбора и веса наблюдений для выбранных переменных.



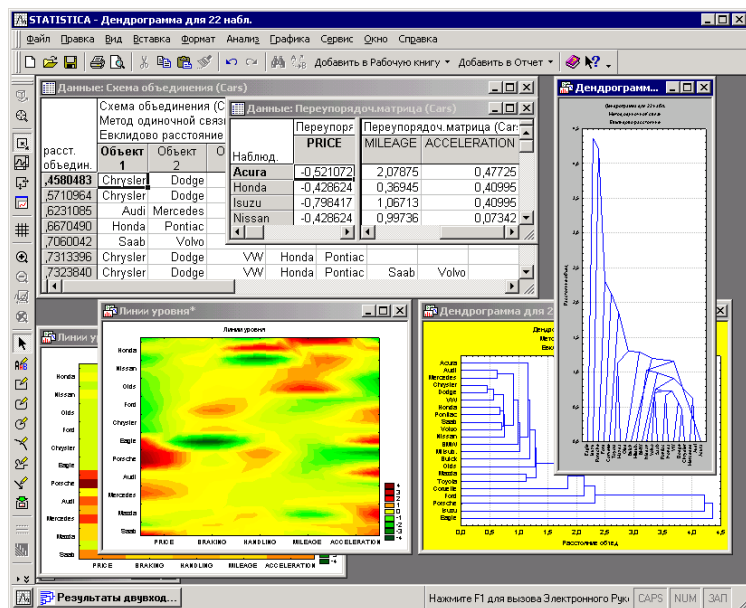
Меню **Графика** включает в себя **2М Графики**, **3М Последовательные графики**, **3М XYZ графики**, **Матричные графики**, **Пиктографики**, **Категоризованные графики** и **Графики пользователя**.

Заметим, что верхняя часть этого меню содержит четыре или пять наиболее часто используемых типов графиков (**Гистограммы**, **Диаграммы рассеяния**, **Графики средних с ошибками** и т.д.), а нижняя часть содержит полный список всех типов графиков.

Как и все остальные меню системы *STATISTICA*, меню **Графика** может легко настраиваться (с помощью вкладки **Меню** диалогового окна **Настройка**, вызываемого командой **Сервис - Настройка**). См. раздел **Графики в меню Графика в Электронном руководстве**.

Другие специализированные графики

Помимо стандартного набора **Графиков исходных данных**, **Графиков блоковых данных** и графиков из меню **Графика** (см. выше) система *STATISTICA* предлагает другие специализированные статистические графики. Они вызываются непосредственно из диалоговых окон результатов (т.е. окон, содержащих выходные параметры текущего анализа) и связаны с каким-либо типом анализа.



Параметры специализированных графиков определяются типом анализа, который их использует. Поэтому для получения более подробной информации см. соответствующие разделы *Электронного руководства*.

7

ГЛАВА

НАСТРОЙКА СИСТЕМЫ *STATISTICA*

Настройка интерактивного пользовательского интерфейса.....	148
Настройка документов.....	149
Временные и постоянные настройки.....	150
Общие стандартные настройки	151
Настройка параметров графиков.....	152
Поддержка различных конфигураций <i>STATISTICA</i>	153
Настройка конфигураций для нескольких пользователей	153





НАСТРОЙКА СИСТЕМЫ *STATISTICA*

Система *STATISTICA* предоставляет широкие возможности настройки пользовательского интерфейса, позволяя приспособивать стандартный пользовательский интерфейс для конкретных решаемых вами задач.

Фактически, *STATISTICA* “предвидит” ваши действия в процессе работы, так как запоминает ваши предпочтения. Например, если вы запустите анализ из вкладки **Дополнительно** диалогового окна задания анализа, то при следующем открытии этого окна вкладка **Дополнительно** будет выбрана автоматически (вместо вкладки **Быстрый**).

Настраиваемыми являются практически все параметры пользовательского интерфейса: команды меню, панели инструментов, клавиатурные команды и т.д.

Сам процесс настройки осуществляется достаточно быстро и просто (см., например, иллюстрацию процесса настройки панели инструментов на стр. 92).

Вы можете установить локальные и общие настройки графиков, электронных таблиц, Рабочих книг, Отчетов и т.д., а также использовать различные конфигурации системы *STATISTICA* (как для одного пользователя, так и для нескольких пользователей, работающих в сети).

Кроме того, вы можете определить совершенно новый пользовательский интерфейс (см. стр. 92 и 94).

Настройка интерактивного пользовательского интерфейса

Система *STATISTICA* предлагает различные средства для создания новых пользовательских интерфейсов (см. стр. 92), включая Internet-ориентированные пользовательские интерфейсы (см. стр. 96). С другой стороны, вы можете легко изменить практически все аспекты используемого по умолчанию интерактивного пользовательского интерфейса.

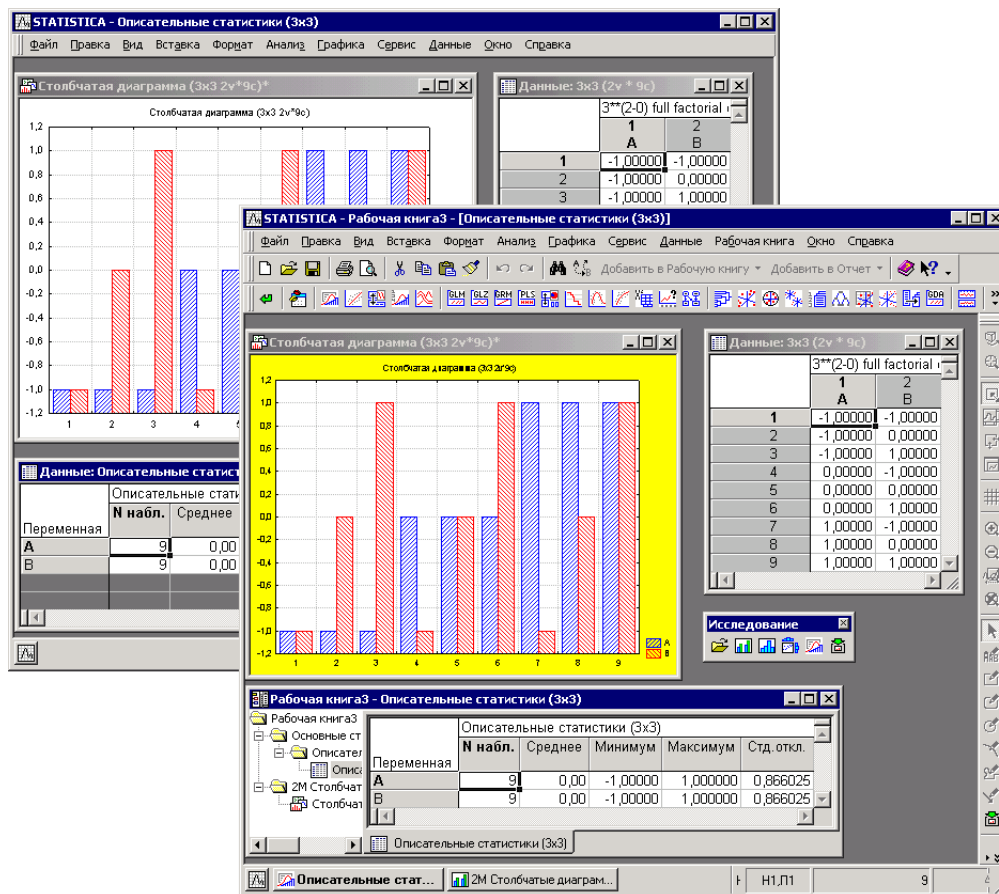
Например, вы можете просто добавлять обычные опции, а можете изменять их.

В зависимости от требований задачи и личных предпочтений (а также эстетических соображений) можно использовать разнообразные “режимы” и условия работы программы.

В частности, можно убрать с экрана все иконки, панели инструментов, строки состояния, длинные меню, Рабочие книги. Можно отключить метод “перетащить и отпустить”, динамические (автоматические) связи между графиками и данными, трехмерные эффекты в таблицах и диалоговых окнах и затребовать “голый” последовательный вывод результатов с простыми черно-белыми электронными таблицами и графиками. Можно настроить систему таким образом, чтобы она одновременно поддерживала только один простой отчет. См. экран, расположенный слева на рисунке ниже.

Можно поступить и другим образом:

создать сложные глобальные и локальные панели инструментов, использовать все преимущества специальных инструментов и средств управления, иконки, панели инструментов, макросы, усовершенствованные мультимедийные таблицы, Рабочие книги, метод “перетащить и отпустить”. Можно установить сложные динамические (автоматические) связи между графиками и данными, а также внутренние связи между графическими объектами. Вы можете настроить цвета и специальные шрифты в окнах результатов, отрегулировать стандартные графические стили и режимы вывода графиков. Все результаты можно направлять в иерархически организованные Рабочие книги, которые упрощают анализ сложных совокупностей данных и позволяют сравнивать различные аспекты выводимых результатов. См. экран, расположенный справа на рисунке ниже.



Настройка документов

Имеется большое количество различных средств для настройки расположения и работы документов системы STATISTICA (см. Главу 4 – *Документы системы STATISTICA*, стр. 109).

Например, STATISTICA имеет мощную систему управления стандартными параметрами каждого графического объекта и объединения настроек в многоуровневые “стили”.

Также вы можете создавать пользовательские шаблоны и форматы электронных таблиц (мультимедийных таблиц) и даже настраивать для них различные события как отклики на различные действия (например, что должно произойти при двойном нажатии мышью на таблице).

Для получения дополнительной информации см. *Электронное руководство*.

Временные и постоянные настройки

Многие параметры, определяющие общий вид системы STATISTICA, можно настраивать как через меню **Вид**, так и через меню **Сервис**. Каждый из этих методов выполняет свою специальную функцию. А именно:

Меню Вид. Изменения, сделанные в выпадающем меню **Вид**, влияют только на текущий вид системы STATISTICA или на активное окно документа (например, позволяют скрыть панель инструментов или изменить используемый в таблице шрифт).

Меню Сервис. Возможности изменения параметров, доступные из меню **Сервис - Параметры** (см. следующий параграф) позволяют настраивать постоянные параметры, используемые программой по умолчанию.

При этом общие характеристики, которые относятся документам определенного типа (например, графику или электронной таблице), не применяются к текущему документу. Вместо этого они сохраняются как программные параметры по умолчанию и вступают в силу только при создании следующего (то есть нового) документа этого типа.

Так, например, если вы измените **Шаблоны Таблиц** во вкладке **Таблицы** диалогового окна **Параметры** (доступного через меню **Сервис - Параметры**), то новые настройки будут приведены в действие только при создании новой электронной таблицы.

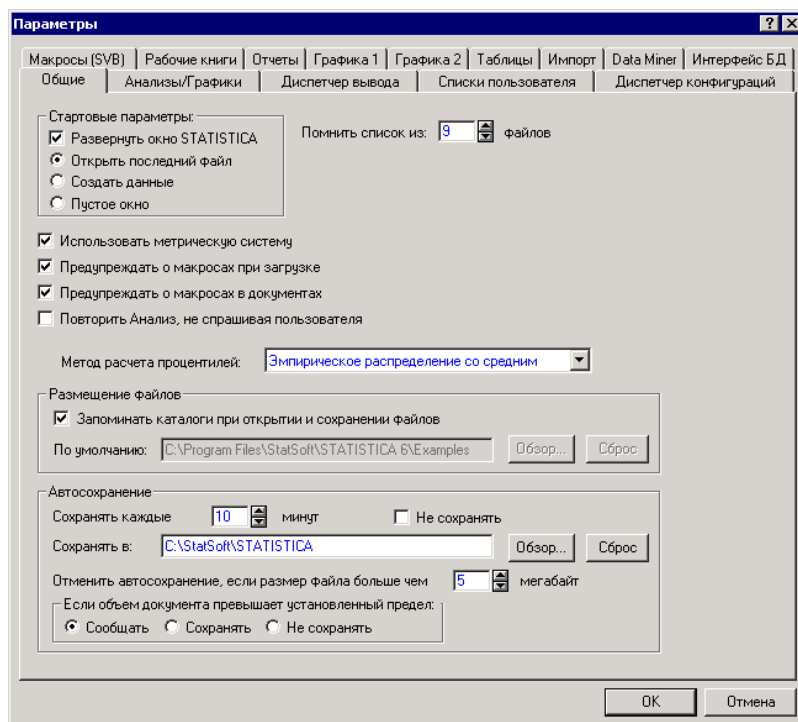
Однако эти настройки не окажут никакого влияния на файлы, открываемые с диска, так как соответствующие электронные таблицы будут появляться в том виде, в каком они были сохранены (настроить уже существующие объекты можно с помощью команд выпадающего меню **Вид**).

Общие стандартные настройки

Настройка общих параметров системы. Настройку общих параметров системы STATISTICA можно изменить в любой момент с помощью соответствующих вкладок диалогового окна **Параметры** (доступного через меню **Сервис - Параметры**). Эти параметры определяют:

- общие аспекты работы системы STATISTICA (разворачивание окна STATISTICA на весь экран при запуске, использование Рабочих книг и Отчетов, расположение файлов, пользовательские списки и т.д.),
- режим вывода результатов (например, в Рабочие книги анализа, в Отчеты анализа и т.д.),
- общий вид окна приложения (значки, панели инструментов и т.д.),
- вид окон документов.

На следующем рисунке показана вкладка **Общие** диалогового окна **Параметры**.



Все эти и другие общие параметры могут быть настроены независимо от типа активного окна документа (например, таблица или график).

Дополнительную информацию о каждой вкладке окна **Параметры** см. в *Электронном руководстве*.

Переключение между различными наборами установок по умолчанию (конфигурациями). Команды для работы с различными конфигурациями системы находятся во вкладке **Диспетчер конфигураций** диалогового окна **Параметры**. Эти команды позволяют поддерживать несколько конфигураций, а также переключать их для различных проектов (или пользователей).

Дополнительную информацию см. в описании вкладки **Диспетчер конфигураций** на стр. 153 и в *Электронном руководстве*.

Настройка параметров графиков

Интерактивная настройка графиков. Опции настройки графиков *STATISTICA* охватывают сотни свойств и инструментов для корректировки любой детали отображения графиков и обработки данных.

Эти опции организованы в виде иерархической структуры. Наиболее часто используемые опции доступны из контекстных меню, вызываемых с помощью двойного щелчка мыши или щелчка правой кнопкой мыши на соответствующем элементе графика.

Постоянные настройки и опции автоматизации. Начальные (принятые по умолчанию) установки всех свойств графиков можно легко отрегулировать. Таким образом, даже вид и свойства графиков по умолчанию будут удовлетворять вашим запросам либо будут требовать минимального вмешательства с вашей стороны.

Различные свойства графиков системы *STATISTICA* могут быть настроены с помощью:

1. диалогового окна **Параметры** (доступного через меню **Сервис - Параметры**),
2. набора графических стилей,
3. графиков пользователя,
4. языка *SVB*.

Способы настройки кратко обсуждаются в Главе 5 – *Графики* (стр. 128). Для получения дополнительной информации см. *Электронное руководство*.

Не существует ограничений по сложности настройки пользовательских графиков системы *STATISTICA*. Это обусловлено тем, что для вывода практически любых графических или мультимедийных результатов можно использовать язык *SVB* (со всеми его инструментами рисования и библиотекой графических процедур, ориентированных на систему *STATISTICA*).

Полученные изображения или мультимедиа могут быть внедрены на панели инструментов *STATISTICA*, в меню или диалоговые окна.

Поддержка различных конфигураций *STATISTICA*

Система *STATISTICA* сохраняет все свои параметры, когда вы выходите из программы, и восстанавливает их при следующем запуске приложения.

Вы можете создать несколько различных конфигураций этих параметров, используя вкладку **Диспетчер конфигураций** диалогового окна **Параметры** (доступного через меню **Сервис - Параметры**).

С помощью диспетчера конфигураций вы можете сохранить текущее состояние программы в новой или существующей конфигурации. Кроме того, вы можете перезагрузить систему *STATISTICA* с другой конфигурацией. Имеются также средства для импорта и экспорта конфигурации в отдельные файлы, которые могут храниться отдельно от файлов программы.

Настройка конфигураций для нескольких пользователей

Правила, описанные в предыдущем разделе, применимы и для сетевой настройки системы *STATISTICA*.

В сети *STATISTICA* устанавливается только на одну машину (на сервер), но каждый пользователь может конфигурировать систему по-своему, потому что параметры конфигурации хранятся локально.

Отметим, что для правильной установки программы на удаленном диске (на сервере в сети) необходимо выбрать режим **Сетевая установка** в программе **Установка STATISTICA**.

Сетевая версия *STATISTICA* необходима для обеспечения надежной работы одновременно нескольких пользователей с системой или же одного пользователя, в случае если система не установлена на локальном компьютере.

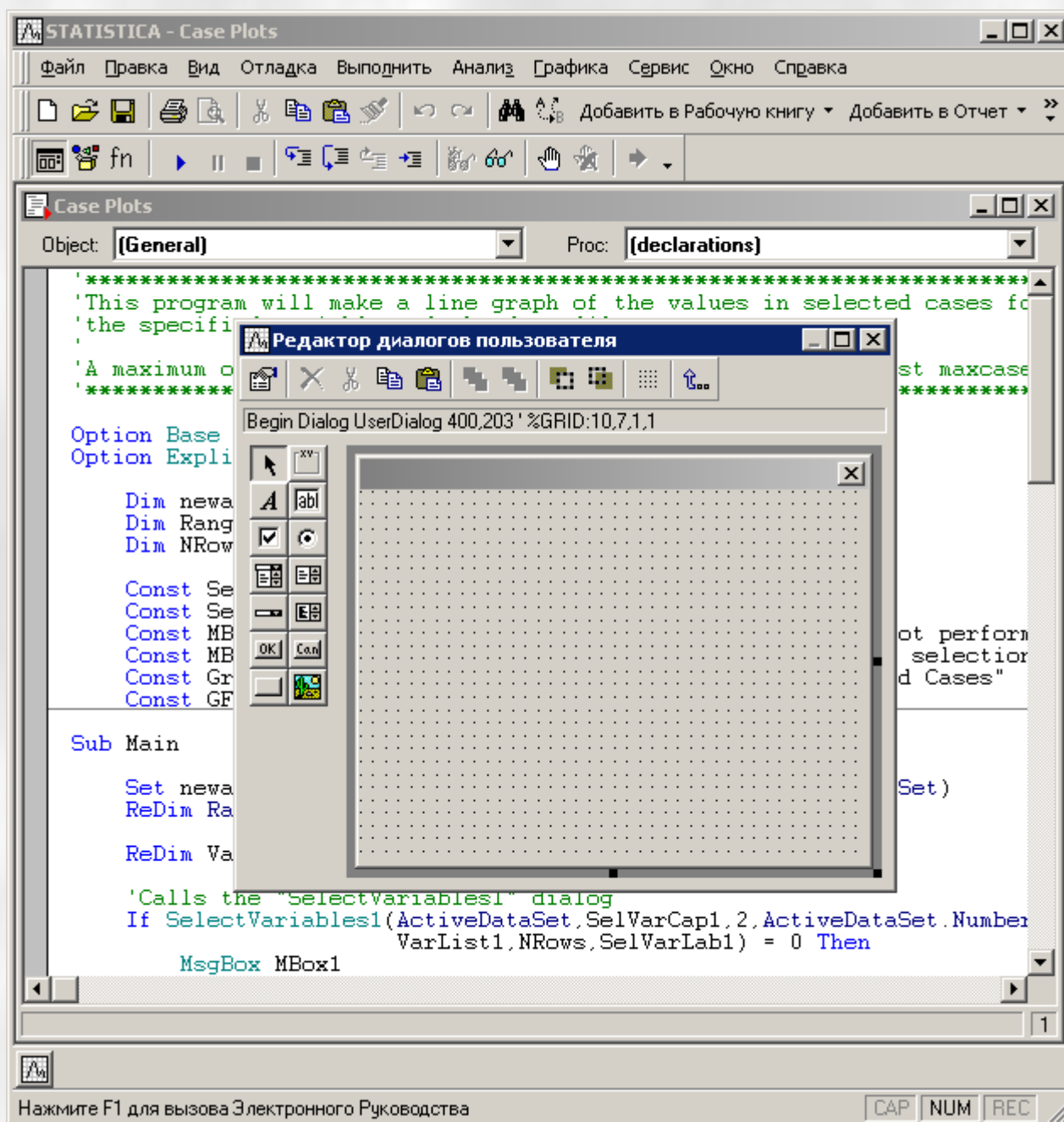
8

ГЛАВА

STATISTICA VISUAL BASIC

Запись макросов.....	158
Среда разработки <i>SVB</i>	160
Выполнение макросов <i>SVB</i>	162
Объекты и документы ActiveX (технические замечания)	163
Поддержка OLE.....	164
Поддержка DDE.....	164





STATISTICA **VISUAL BASIC**

Стандартный язык программирования *SVB* (интегрированный в систему *STATISTICA*) предоставляет намного больше возможностей, чем просто “вспомогательный язык программирования”, который используется для создания пользовательских приложений.

SVB использует огромные преимущества объектно-ориентированной структуры системы *STATISTICA* и позволяет получить доступ программными средствами практически ко всем функциональным деталям пакета *STATISTICA*. Даже самые сложные процедуры анализа и графический вывод результатов могут быть записаны в виде макросов на языке Visual Basic, которые затем можно воспроизвести или заново отредактировать.

Макросы представляют собой отдельные блоки, которые легко встраиваются в другие приложения.

SVB добавляет более 10 000 новых функций к стандартному синтаксису Microsoft Visual Basic, и является, таким образом, одним из самых функционально богатых и обширных интерфейсов прикладного программирования.

Применение программ *SVB*. Программы *SVB* имеют широкий круг применения: от простой автоматизации часто используемой последовательности действий до разработки сложных аналитических систем, сочетающих в себе мощные оптимизированные процедуры системы *STATISTICA* с возможностями расширения пользовательского интерфейса.



Сценарии для анализов, созданные с помощью *SVB*, могут быть интегрированы в различные среды программирования, выполнены из корпоративных программных систем, а также на порталах Internet и Intranet.

Создание программ SVB. Имеется два основных способа создания программ на языке SVB: запись макросов и использование среды разработки SVB. См. следующие два раздела.

Заметим, что эти два метода могут сочетаться друг с другом. Например, записанный макрос можно затем отредактировать в среде разработки.

Запись макросов

Простейший способ создания программ на языке SVB – это запись макросов. Когда вы запускаете процедуру анализа (из меню **Анализ**) или вызываете график (из меню **Графика**), автоматически создается код на языке Visual Basic, в котором записываются все выбранные вами параметры анализа и опции вывода.

Для отображения этого кода выберите команду **Создать Макрос** в выпадающем меню кнопки  Опции (в окне задания любого анализа или графика). Вы также можете выбрать команду **Создать макрос Анализа/Графика** в меню **Сервис - Макрос** или нажать кнопку  на панели инструментов **Анализ**.

Этот код впоследствии можно запускать на выполнение и редактировать путем изменения опций, переменных, файлов данных, добавления пользовательских интерфейсов и т.д.

В системе *STATISTICA* существуют три вида макросов, которые могут автоматически создаваться.

1. Макрос анализа

По умолчанию каждый раз, когда вы выполняете анализ из меню **Анализ** или строите график из меню **Графика**, в фоновом режиме записываются макросы SVB.

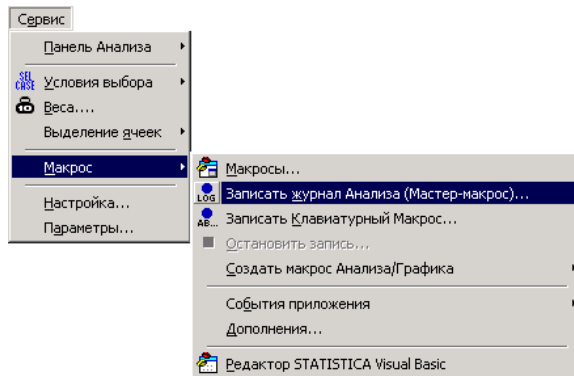
Впоследствии вы можете использовать эти “строительные блоки” для создания различных приложений.

2. Мастер-макрос (журнал)

Если же вы намереваетесь создать программу SVB, которая включает последовательность нескольких выполняемых в интерактивном режиме процедур

анализа, то перед запуском анализа выберите команду **Записать журнал Анализа (Мастер-макрос)** в меню **Сервис - Макрос**.

Журнал будет также включать операции управления данными, содержащиеся в меню **Данные**.



Для получения дополнительной информации о записи журнала анализа нажмите клавишу F1, когда выделена данная команда, либо обратитесь к *Руководству по STATISTICA Visual Basic*.

3. Клавиатурный макрос

Это третий (и, возможно, самый простой) вид макроса, который записывает последовательность нажатий клавиш клавиатуры, не учитывая при этом контекст, в котором происходило нажатие клавиш.

Такие макросы полезны для автоматизации ввода текста, например, комментариев, заголовков, условий выбора и т.п.

Среда разработки SVB

Программы также могут быть написаны с нуля в профессиональной среде разработки SVB. Она предоставляет удобный редактор программного кода с мощным отладчиком (позволяющим устанавливать точки останова и т.п.),

```

[interest.svb|SimpleInterest#148] InterestEarned# = (Return@ - Principal@)
[interest.svb|RunInterest# 51] SimpleInterest
[interest.svb|Main# 38] Dialog dlg

Object: [General] Proc: SimpleInterest

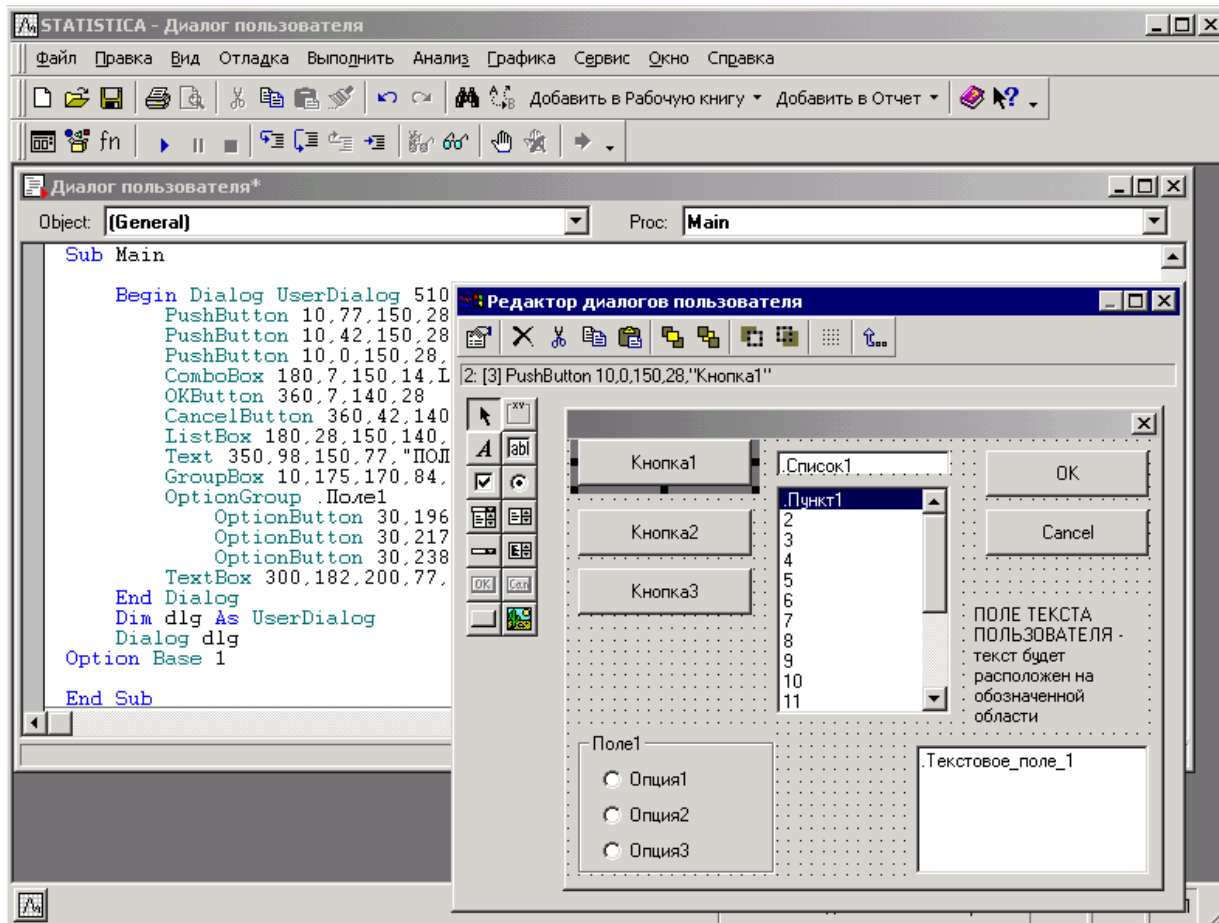
s.SetSize(PeriodCount,1)
s.CaseNameLength = 18
s.InfoBox.Value = "Simple interest over" + Str(PeriodCount) + " rollo
s.InfoBox.AutoFit

s.Variables(1).NumberFormat = "$0.00"
s.DisplayAttribute(scDisplayVariableNumbers) = False

Dim YearValue As Double
Dim InterestEarned As Double
YearValue = 1 / PeriodsPerYear
'A = P(1 + rt)
Return = Principal * (1 + (Rate * YearValue))
'because the interest is simple, then the interest will remain consta
InterestEarned = (Return - Principal)

For i = 1 To PeriodCount
  If i = 1 Then
    'already calculated the first return
    s.Value(1,1) = Return
    'write to report
    TextString = TextString & "Total after rollover 1: " & Forme
  Else
    'Simple Interest Report"
    'A = P(1 + rt)
    s.Value(i,1) = (Return + (InterestEarned * (i - 1)))
    'write to report
    TextString = TextString & "Total after rollover " & i & ":
  s.Case(i).RowName = "Rollover " + Str(i)
End If
    
```

интуитивный редактор диалоговых окон



и другие средства для эффективного написания кода.

Чтобы запустить редактор SVB, выберите команду **Создать** в меню **Файл** и используйте вкладку **Макрос (SVB)** в появившемся окне **Создать Новый Документ**.

Выполнение макросов *SVB*

Макросы *SVB* можно запустить на выполнение как внутри системы *STATISTICA*, так и из других совместимых с Visual Basic приложений (например, Microsoft Excel, Microsoft Word, отдельная среда программирования Visual Basic).

Однако заметим, что когда вы запустите программу *SVB* или попытаетесь вызвать функции системы *STATISTICA* из другого приложения, то все вызовы к функциям *STATISTICA* (в отличие от функций MS Visual Basic) будут выполнены только в том случае, если соответствующие библиотеки *STATISTICA* присутствуют на вашем компьютере.

Таким образом, *пользователь программы должен иметь лицензию на использование соответствующих библиотек системы STATISTICA.*

Библиотека функций системы *STATISTICA* (более 10 000 процедур) открыта для использования не только из Visual Basic (как встроенного, так и стандартного), но также из любого другого совместимого языка программирования, например, из C/C++, Java или Delphi.

Производительность макросов *SVB*. Преимуществами языка Visual Basic (по сравнению с другими языками программирования) являются простота использования и широкое распространение. Возможный недостаток программ, написанных на Visual Basic, состоит в том, что они не обеспечивают такую же высокую производительность, как приложения, написанные на низкоуровневых языках программирования (например, Си).

Тем не менее, эта потенциальная проблема обычно не проявляется во время работы приложений *SVB*, особенно тех, которые главным образом используют вызов аналитических, графических функций *STATISTICA* и функций управления данными. Эти функции оптимизированы для обеспечения производительности, сравнимой с работой функций в самой системе *STATISTICA*.

Объекты и документы ActiveX (технические замечания)

Термин ActiveX используется в нескольких смыслах, и его определение зависит от контекста. Однако мы можем выделить ActiveX-объекты и ActiveX-документы.

ActiveX-объекты. Другое название для ActiveX-объектов – OLE-объекты (Object Linking and Embedding – объектное связывание и внедрение). Это понятие является центральным в технологии Microsoft COM (Component Object Model – модель компонентных объектов), позволяющей получать доступ к объектам единообразным способом.

Благодаря использованию стандартных протоколов, объекты, созданные в одном приложении, можно сохранять и редактировать в другом приложении. Для поддержки такой возможности последнее приложение должно быть ActiveX-клиентом, а приложение, в котором создается объект, должно быть ActiveX-сервером.

STATISTICA является и тем, и другим. В качестве ActiveX-клиента *STATISTICA* позволяет внедрять объекты из других приложений в окна электронных таблиц, графиков и Отчетов, а также связывать объекты с этими окнами. В качестве ActiveX-сервера *STATISTICA* позволяет внедрять электронные таблицы и графики в другие приложения, а также связывать эти объекты с другими приложениями.

ActiveX-документы. ActiveX-документы используют больше возможностей управляющих элементов ActiveX. Благодаря этому возможно внедрение целых документов в другие приложения.

ActiveX-контейнер позволяет использовать документы из других приложений, а ActiveX-сервер позволяет использовать свои документы в любых ActiveX-контейнерах. Опять же, *STATISTICA* является и тем и другим.

Рабочие книги *STATISTICA* являются ActiveX-контейнерами. Поэтому документы из других ActiveX-серверов можно отображать в Рабочей книге. Примерами являются документы Microsoft Word и Excel. С этими документами можно работать непосредственно в Рабочей книге *STATISTICA*.

Электронные таблицы, графики и Отчеты *STATISTICA* являются ActiveX-серверами. Они могут быть размещены внутри любого ActiveX-контейнера, например, Microsoft Internet Explorer или Microsoft Binder.

Поддержка OLE

STATISTICA поддерживает стандарт OLE (Object Linking and Embedding – объектное связывание и внедрение), который позволяет связывать числовые значения, текст, графики или другие объекты в документах *STATISTICA* (например, электронных таблицах, Отчетах) с документами других приложений Windows.

Говоря техническим языком, вы можете установить OLE-связь между файлом-“источником” (или сервером) и документом системы *STATISTICA* (файлом-“клиентом”) таким образом, что при внесении изменений в файл-источник автоматически будут обновляться данные в соответствующей части документа *STATISTICA* (файле-клиенте).

Кроме того, система *STATISTICA* сама может выступать “источником” (или сервером) для других документов, поддерживающих OLE. Например, вы можете связать значения в электронной таблице *STATISTICA* с таблицей в документе Microsoft Word таким образом, что документ Word будет обновляться при изменении данных в электронной таблице.

Поддержка DDE

STATISTICA также поддерживает соглашения по динамическому обмену данными (Dynamic Data Exchange, DDE). DDE-связи устанавливаются между файлом-“источником” (или сервером), например, электронной таблицей Microsoft Excel, и файлом данных *STATISTICA* (файлом-“клиентом”) таким образом, что при внесении изменений в файл-источник автоматически обновляются данные в соответствующей части электронной таблицы *STATISTICA* (файле-клиенте).

Таким образом, вы можете динамически связывать области данных в электронной таблице с подмножествами данных в других приложениях Windows.

Динамическое связывание двух файлов обычно находит применение в промышленных системах, где файл данных *STATISTICA* динамически связывается с измерительным прибором, подсоединенным к последовательному порту (например, для постоянного автоматического обновления измерений).

Как и в случае OLE, осуществление подобной процедуры значительно проще, чем может показаться, и не требует специальных технических знаний о механизме DDE.

9

ГЛАВА

STATISTICA QUERY – ЗАПРОС *STATISTICA*

Обзор.....	167
Краткие пошаговые инструкции.....	168
Обработка данных с удаленных серверов “на месте”	169



STATISTICA QUERY - ЗАПРОС STATISTICA

Замечание: Объяснение всех технических терминов, используемых в данном обзоре (например, ODBC, SQL и т.д.), можно найти в словаре *Электронного руководства*.

Обзор

Запрос *STATISTICA* используется для облегчения доступа к данным целого ряда баз данных (включая многие большие системные базы данных, такие как Oracle, MS SQL Server, Sybase и т.д.) с помощью технологии Microsoft OLE DB.

Технология баз данных OLE DB обеспечивает универсальную интеграцию данных через локальную сеть предприятия, от сервера к рабочим станциям независимо от типа данных.

OLE DB предлагает более обобщенную и более эффективную стратегию доступа к данным, чем более старая технология ODBC, поскольку OLE DB предоставляет доступ к большему числу типов данных и базируется на модели компонентных объектов (COM).

Запрос *STATISTICA* поддерживает составные таблицы баз данных. Определенные записи (строки таблиц) могут быть выбраны путем ввода SQL выражений, которые Запрос *STATISTICA* строит автоматически, как только вы выбрали компоненты запроса с помощью графического интерфейса и/или опций меню и диалогов. Следовательно, для создания даже сложных запросов глубокие знания синтаксиса SQL не нужны.

Также можно создавать составные запросы к одной или сразу нескольким базам данных, в результате которых данные будут помещены в отдельную электронную таблицу.

В этом окне вы можете составлять ваши собственные SQL выражения, указав необходимые для включения в ваш запрос таблицы, поля, связи, критерии и т.д. (с помощью меню **Таблица, Связь и Критерии**).

3. После создания запроса выберите команду **Вернуть данные в STATISTICA** в меню **Файл**.

Появится диалоговое окно **Экспорт внешних данных в таблицу**, в котором можно ввести имя запроса, указать, куда вы хотите поместить результаты запроса, а также задать дополнительные параметры.

Более подробную информацию см. в *Электронном руководстве*.

Обработка данных с удаленных серверов “на месте”

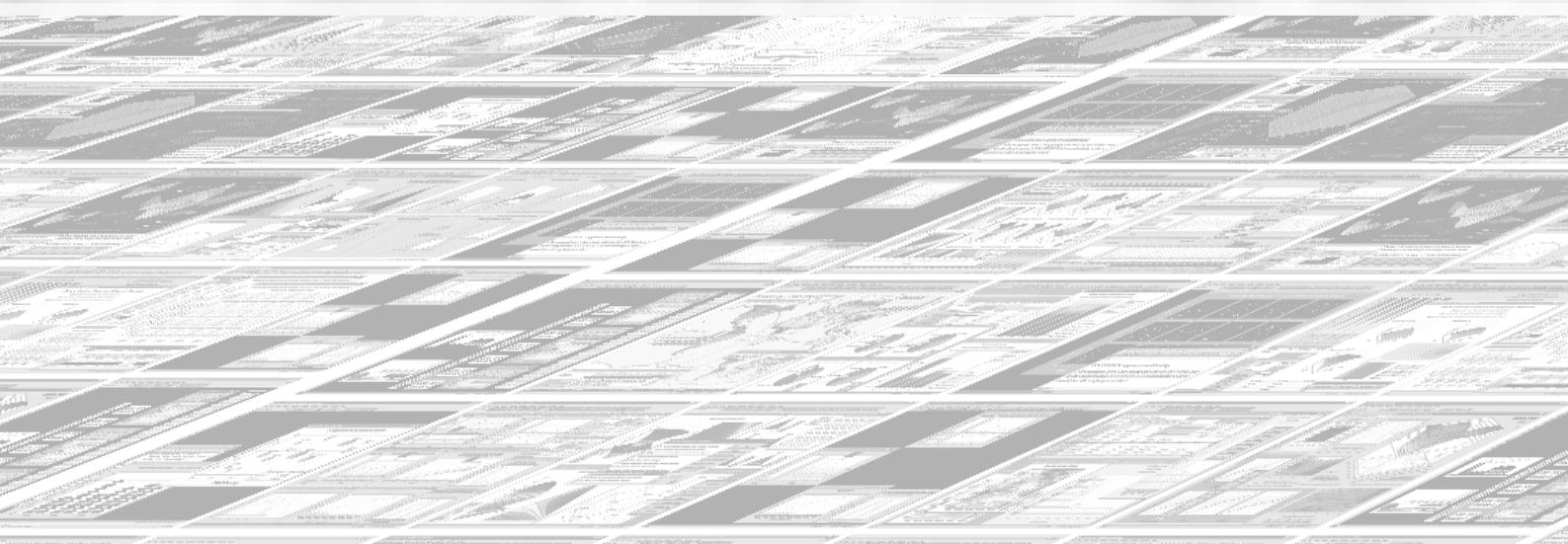
Возможности построения запросов (описанные в предыдущем разделе), предлагаемые в корпоративных версиях *STATISTICA* (см. *STATISTICA Enterprise Systems – Корпоративные системы STATISTICA*, стр. 213), дополнительно расширены средствами обработки данных с удаленных серверов “на месте”. Таким образом, снимается необходимость импорта данных и создания локального файла, содержащего эти данные.

Эта технология полезна для обработки очень больших файлов данных, так как дает значительный выигрыш в производительности, а также позволяет обрабатывать файлы, превосходящие по размерам емкость локального запоминающего устройства.

ОБЗОР ФУНКЦИОНАЛЬНЫХ ВОЗМОЖНОСТЕЙ СИСТЕМЫ *STATISTICA*

Аналитические средства.....	173
Уникальные возможности	173
Основная “философия” работы в <i>STATISTICA</i>	174
Используемые технологии программного обеспечения (технические замечания)	175
Web-интеграция	176
Отзывы и награды.....	177





ОБЗОР ФУНКЦИОНАЛЬНЫХ ВОЗМОЖНОСТЕЙ СИСТЕМЫ *STATISTICA*

Система *STATISTICA* содержит исчерпывающий набор аналитических процедур в области изучения бизнеса, добычи данных, науки и промышленного производства. Она позволяет строить различные графики, эффективно управлять данными и разрабатывать собственные приложения.

Аналитические средства

STATISTICA не только включает в себя универсальные статистические, графические процедуры и средства управления данными, но также реализует специализированные методы анализа данных (например, для добычи данных, бизнеса, социальных наук, медицинских исследований, инженерных приложений).

Все аналитические инструменты *STATISTICA* доступны как отдельные компоненты единого интегрированного пакета. Управлять ими можно различными способами: через пользовательский интерфейс или с помощью стандартного языка программирования *SVB*.

Интерфейс пользователя легко настраивается, а язык *SVB* позволяет автоматизировать задачи любой сложности. Применяя язык программирования, вы можете как избавиться себя от рутинной работы с данными, так и упростить работу сложных масштабных проектов (например, интегрировать систему *STATISTICA* с другими приложениями или даже с большой корпоративной системой).

Уникальные возможности

Вот лишь некоторые уникальные возможности системы *STATISTICA*:

- широкий выбор аналитических процедур и их исчерпывающая реализация;
- набор качественных, настраиваемых по желанию пользователя графиков, дополняющих каждую вычислительную процедуру;
- эффективный и дружелюбный пользовательский интерфейс;
- интегрированный стандартный язык программирования *SVB*, добавляющий более 10 000 новых функций к стандартному синтаксису Microsoft Visual Basic;
- большой набор технологий программного обеспечения (см. *Используемые технологии программного обеспечения* ниже), которые позволяют системе *STATISTICA* обрабатывать практически неограниченные объемы данных, достигать очень высокой производительности и поддерживать разнообразные опции настройки.

Одна из наиболее важных особенностей системы *STATISTICA* заключается в том, что подобные технологии позволяют даже новичку приспособить работу системы в соответствии со своими предпочтениями. Вы можете настроить практически каждый элемент *STATISTICA*, включая детали интерфейса.

С одной и той же версией *STATISTICA* могут работать:

- Начинающие пользователи, не искушенные в анализе данных. Они имеют возможность решать типовые задачи, принимая стартовые установки по умолчанию во всех диалоговых окнах анализов (используя, например, вкладку **Быстрый**) или даже работая со *STATISTICA* через обычный Web-браузер.
- Опытные аналитики, ученые и разработчики программ, которые могут интегрировать любую оптимизированную процедуру системы *STATISTICA* (более 10 000 функций) в пользовательское приложение или вычислительную среду, используя объектно-ориентированные технологии и Web-технологии.

Общая “философия” работы в *STATISTICA*

Стандартная конфигурация *STATISTICA* (пользовательский интерфейс и параметры системы) является результатом многолетнего опыта StatSoft в области анализа данных. Мы поддерживаем связь с несколькими сотнями тысяч пользователей со всех континентов. Главным образом, на основе их отзывов мы выяснили, насколько различаются их потребности и предпочтения.

В итоге, система *STATISTICA* была разработана таким образом, чтобы предоставить пользователям максимально гибкий и удобный интерфейс для анализа и визуализации данных.

Хотя *STATISTICA* позволяет использовать мощный арсенал современных технологий программного обеспечения (см. *Используемые технологии программного обеспечения* ниже), вам совсем не обязательно знать их тонкости, так как они работают полностью автоматически.

Новички при работе с системой могут ограничиться использованием нескольких кнопок, которые не требуют объяснений. При этом дополнительные опции доступны, как правило, в соседней вкладке.

Практически каждый аспект системы *STATISTICA* (начиная со стартовой конфигурации и заканчивая способом вывода результатов) может быть изменен несколькими щелчками мыши. Кроме того, *STATISTICA* запоминает последние сделанные вами изменения.

Вы можете легко заменить практически каждое диалоговое окно, используемое при работе (например, упростить его, расширить или связать с пользовательскими настройками).

Таким образом, *STATISTICA* всегда будет работать так, как вы хотите.

Используемые технологии программного обеспечения (технические замечания)

Вышеуказанные свойства системы *STATISTICA* – производительность, гибкость настройки и широкий набор опций – были бы невозможны, если бы в *STATISTICA* не использовались современные технологии, управляющие функциями приложения. В *STATISTICA* поддерживаются практически все современные передовые технологии программного обеспечения.

Любая из тысячи функций *STATISTICA* доступна с помощью объектной модели, и к любой из них может быть получен прозрачный доступ при интеграции с другими приложениями (например, с помощью Visual Basic, C++ или Java).

В *STATISTICA* внедрена поддержка технологии OLE/ActiveX в режиме клиента и сервера для всех типов документов. Не накладывается практически никаких ограничений на объем и сложность данных при сохранении документов.

Также *STATISTICA* поддерживает Web-технологии и технологии мультимедиа.

Вычислительные и графические процедуры могут использовать различные методы оптимизации (например, точность вычислений “четвертого порядка” позволяет превысить предел, установленный стандартами IEEE, при вычислениях с плавающей точкой). В результате, работа *STATISTICA* характеризуется непревзойденной скоростью, высокой точностью вычислений и оперативностью, достигаемой с помощью многопоточности.

Доступ к данным основан на гибкой технологии работы с потоками. Она позволяет системе *STATISTICA* легко работать как с исходными файлами данных, расположенными на локальном диске, так и с запросами к многомерным базам данных, содержащим терабайты информации в удаленных хранилищах данных. Кроме того, в последнем случае существует возможность обрабатывать данные “на месте” (т.е. без импортирования на локальный диск; эта возможность реализована в корпоративных версиях системы *STATISTICA*).

Вы можете одновременно запустить несколько копий системы *STATISTICA*. Каждая копия при этом может проводить анализы на основе нескольких одновременно открытых исходных файлов данных и запросов. Результаты могут быть организованы в виде отдельных проектов.

Исходные файлы данных, таблицы результатов и графики *STATISTICA* могут быть практически неограниченного размера и состоять из иерархии документов различных типов.

Результаты могут быть направлены по нескольким каналам вывода, например, в Рабочие книги, Отчеты, Internet и т.п.

Web-интеграция

Поскольку *STATISTICA* поддерживает Web-технологии, то как “вывод”, так и “ввод” в *STATISTICA* можно осуществлять через Internet.

Существует возможность направления результатов анализа прямо на Web-сервер. Дополнительно *STATISTICA* предоставляет методы для построения весьма сложных автоматизированных систем, запускающих сценарии, которые могут, например, периодически получать данные из внешних источников, проводить их анализ и обновлять содержание HTML-страниц на заданных Web-серверах.

STATISTICA поддерживает ввод через Internet двумя способами (заметим, что эта возможность распространяется только на версии, предназначенные для инсталляции на Web-сервер):

- Обращение к “данным Internet” с помощью технологии XML. Эта технология позволяет считывать данные, хранящиеся в различных форматах, если инструкции по их чтению и интерпретации хранятся вместе с самими данными и, таким образом, доступны XML-совместимым приложениям.
- Использование Internet-браузера в качестве интерфейсной части *STATISTICA* (“тонкий клиент”), что позволяет запускать систему *STATISTICA* с абсолютно любого компьютера в мире, если он подключен к Internet. *Необходимо отметить, что такая возможность требует наличия соответствующей лицензии и не может быть использована на локальной версии STATISTICA.*

STATISTICA Web-сервер (технические замечания). Эта последняя функциональная возможность поддерживается в системе *STATISTICA*, поскольку ее всесторонние средства автоматизации (основанные на объектной модели) могут быть запущены с помощью языка сценариев (например, VBScript, Jscript или C++).

Если *STATISTICA* установлена на сервере, вы можете получить доступ к модели автоматизации для обработки данных и получения результатов, которые вы можете отослать клиенту в виде HTML-таблиц и графических (JPG/PNG) файлов.

Более того, если *STATISTICA* установлена на стороне клиента, вы также можете использовать клиентские сценарии. Тот факт, что ваш Internet-браузер работает с документами ActiveX, позволяет вам открывать электронные таблицы и графики *STATISTICA* в браузере вместе с панелями инструментов и меню *STATISTICA*.

Web-ориентированный пользовательский интерфейс предлагается компанией StatSoft в виде набора упрощенных “диалоговых шаблонов”, вызываемых в Internet-браузере, что позволяет вам легко выполнять основные типы анализа и строить графики.

Заметим, что этот интерфейс может быть гибко настроен путем изменения сценариев анализов, выполняемых на стороне сервера. Web-ориентированный пользовательский интерфейс не только полностью настраивается, но также позволяет включать в себя другие приложения, которые взаимодействуют с системой *STATISTICA*.

Отзывы и награды

Мы рады сообщить, что система *STATISTICA* получала высшую оценку во всех независимых сравнительных обзорах статистического программного обеспечения, в которых она участвовала. За все время развития компьютерной индустрии лишь немногие программные продукты добивались подобного успеха.

Более подробную информацию о компании StatSoft, а также отзывы о системе *STATISTICA* вы можете найти на нашем Web-сайте по адресу <http://www.statsoft.ru>.

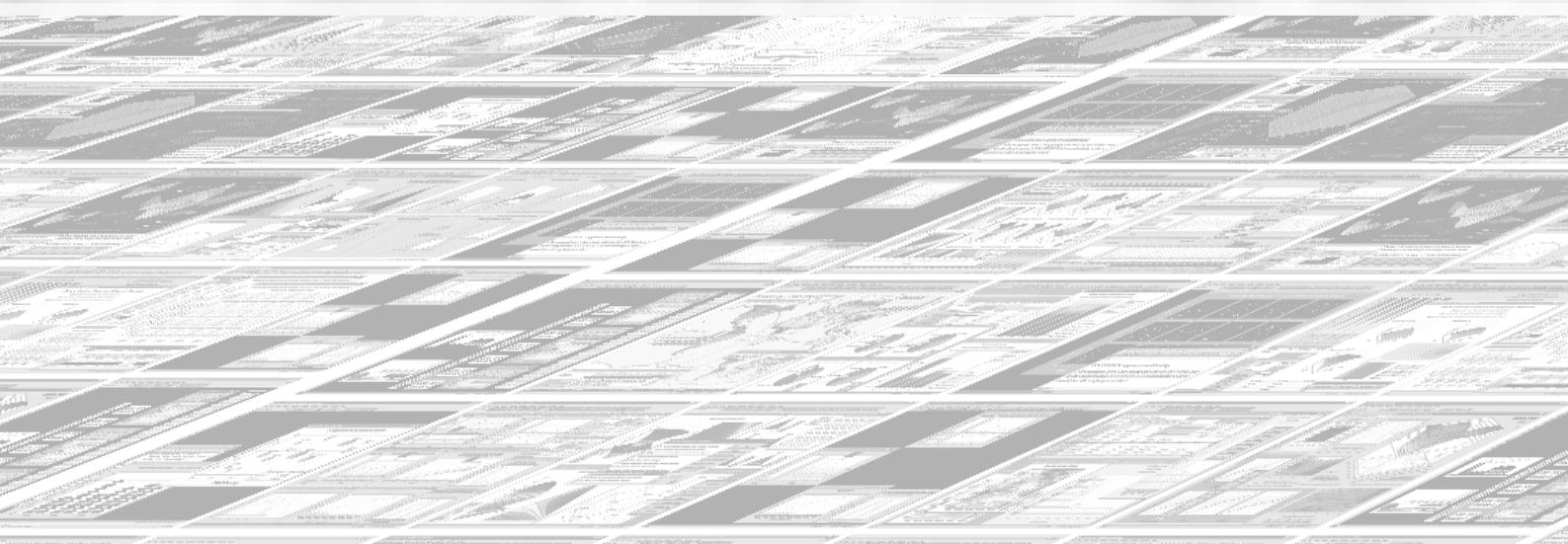
В

ПРИЛОЖЕНИЕ

ПОЛУЧЕНИЕ ПОДРОБНОЙ СПРАВКИ

Электронное руководство – более 100 мегабайт справочной информации, иллюстраций и примеров	181
Другие ресурсы и средства технической поддержки	183





ПОЛУЧЕНИЕ ПОДРОБНОЙ СПРАВКИ

Электронное руководство – более 100 мегабайт справочной информации, иллюстраций и примеров

Самым удобным способом получения справки о системе *STATISTICA* является обращение к *Электронному руководству*.

Электронное Руководство

Скрыть Назад Вперед Печать

Содержание | Указатель | Поиск | Избранное

Вопросы и ответы

- Общие
 - Какие типы графиков пре...
 - Зависят ли настройки от...
 - Можно ли автоматически...
 - В каких форматах можно...
 - Что такое формат WMF?
 - Что такое формат BMP?
 - Что такое формат JPEG?
 - Что такое формат PNG?
 - Что такое собственный ф...
 - Как я могу экспортироват...
 - Чем отличается график о...
 - Что такое категоризованн...
 - Как определять группы д...**
 - Как построить тернарнус...
 - Как построить график в п...
 - Что такое составные оси
 - Как используется мышь г...
 - Как выбрать объект на пр...
 - Что такое 100(1-alpha)% д...
- Настройка графика
- Подгонка, Построение функц...
- Интерактивный анализ
- Составные графики, внедрен...
- Печать графиков
- Примеры
- Замечания и техническая инфор...
- Работа с Рабочими книгами
- Работа с Отчетами
- Работа с Макросами (STATISTICA Vis...

Как определять группы для категоризованных графиков?

Если категоризованные графики вызваны из окон вывода специальных процедур, анализирующих подгруппы данных, то они автоматически отображат уже проанализированные подгруппы (т.е. подгруппы определенные в процессе анализа). С другой стороны, категоризованные графики, вызванные из меню *Графика*, обладают достаточно обширным набором методов разделения на подгруппы с использованием одной или двух группирующих переменных. Кроме того, пользовательские определения подгрупп могут использовать все переменные входного файла.

В частности, группы могут быть определены при помощи:

- Целых значений группирующих переменных (*Целье число*),
- Разделения группирующих переменных на необходимое число интервалов одинаковой длины (*Категории*),
- Пользовательских интервалов (интервалы) группирующих переменных, определенных специальными границами интервалов (*Границы*),
- Специальными значениями (т.е., кодами) группирующих переменных (*Коды*) и
- Заданных пользователем "сложных подгрупп" (*Сложные подгруппы*), которые могут быть введены при помощи условий логического выбора неограниченной сложности (данный метод категоризации может использовать значения всех переменных данного файла, как показано ниже, см. [Методы категоризации](#)).

Scatterplot (Extended Factor.sta 20v*100c)

Scatterplot (Extended Factor.sta 20v*100c)

HOME 2: <= 104.624

K_2

Электронное руководство представляет собой гипертекстовый документ, в котором содержатся объяснения каждого элемента системы *STATISTICA*. Кроме этого, оно включает в себя бесчисленные примеры, обзоры и иллюстрации, а также тысячи подсказок, позволяющих оптимизировать работу.

В *Электронное руководство* встроены **Статистический советник** (см. стр. 67) и **Словарь терминов**. Оно также позволяет обратиться к *Электронному учебнику* StatSoft.

The screenshot shows a web browser window titled "Электронный учебник - Начальная страница - Microsoft Internet Explorer". The address bar shows the URL <http://www.statsoft.ru/home/textbook/default.htm>. The main content area displays a lesson titled "Определение корреляции" (Definition of correlation). The text explains that correlation is a measure of dependence between variables, with the most common being Pearson's correlation. It notes that Pearson's correlation assumes variables are measured on an interval scale. Other coefficients can be used for non-informative scales. The correlation coefficient ranges from -1.00 to +1.00. A value of -1.00 indicates a strict negative correlation, and a value of +1.00 indicates a strict positive correlation. A value of 0.00 indicates no correlation.

Below the text are four scatter plots showing the relationship between MEASURE1 and MEASURE2 for four different groups:

- GROUP 1: $r = -0.90$ (strong negative correlation)
- GROUP 2: $r = -0.50$ (moderate negative correlation)
- GROUP 3: $r = 0.00$ (no correlation)
- GROUP 4: $r = 0.40$ (moderate positive correlation)

On the right side of the page is a "СОДЕРЖАНИЕ" (CONTENTS) sidebar with the following items:

- Элементарные понятия
- Основные статистики
- Анализ выживаемости
- Анализ мощности
- Анализ надежности
- Анализ процессов
- Анализ соответствий
- Временные ряды
- Графические методы
- Деревья классификации
- Дискриминантный анализ
- Дисперсионный анализ
- Канонический анализ
- Кластерный анализ
- Компоненты дисперсии
- Контроль качества

At the bottom of the page, there is a footer with the StatSoft logo, copyright information (© StatSoft, Inc. 1984-2001), and a "GLOSSARY StatSoft" link.

Электронный учебник, разработанный компанией StatSoft, рекомендован для использования информационным порталом *Encyclopedia Britannica*



за свое “высокое качество, точность, удобное преподнесение информации и практичность”. Этот уникальный учебник используется на протяжении многих лет в различных университетах и научно-исследовательских организациях во всем мире.

Другие ресурсы и средства технической поддержки

Web-сайт. Web-сайт компании StatSoft является одним из самых популярных ресурсов в Internet, связанных с анализом данных. На этом сайте вы получите доступ не только к различным аналитическим ресурсам, но и сможете воспользоваться:

- регулярно обновляемым разделом *Часто задаваемые вопросы*,
- разделом, в котором пользователи текущих версий продуктов *STATISTICA* могут получить обновления программного обеспечения.

Мы постоянно работаем над улучшением совместимости различных версий *STATISTICA*. Поэтому, загрузив соответствующее обновление, вы сможете решить некоторые проблемы, связанные с нестандартной конфигурацией системы или конфликтами с другими приложениями.

Техническая поддержка по e-mail. Если вы не нашли ответ на свой вопрос в *Электронном руководстве* или на Web-сайте StatSoft, то можете обратиться к нам по электронной почте.

Пожалуйста, укажите в письме ваш серийный номер (его можно узнать, выбрав команду **О программе** в меню **Справка**), информацию о вашем компьютере (тип процессора, объем оперативной памяти и свободное место на диске), а также версию используемой операционной системы.

Если вы находитесь в России, отправляйте электронные письма по адресу info@statsoft.ru. Иначе обратитесь в региональный офис компании StatSoft (см. ниже).

Техническая поддержка по телефону. Вы также можете позвонить в региональный офис StatSoft и поговорить с техническим специалистом.

Если вы живете в России, то звоните по телефону (095) 787-77-33.

Координаты всех региональных офисов StatSoft вы можете узнать, выбрав команду **О программе** в меню **Справка**, а затем вкладку **Офисы StatSoft** в появившемся окне **О программе**.



Перед тем, как обращаться в службу технической поддержки, уточните, пожалуйста, ваш серийный номер, информацию о вашем компьютере (тип процессора, размер оперативной памяти и свободное место на диске), а также версию используемой операционной системы.

C

ПРИЛОЖЕНИЕ

НОВЫЕ ВОЗМОЖНОСТИ
STATISTICA 6







НОВЫЕ ВОЗМОЖНОСТИ *STATISTICA 6*

Пользователи, знакомые с предыдущей версией *STATISTICA*, обнаружат следующие новшества в 6-й версии системы:

- Усовершенствованный интерфейс пользователя, который полностью настраивается и может быть легко (а в некоторых случаях автоматически) приспособлен для работы различных пользователей: от новичков до профессиональных аналитиков. В новой версии *STATISTICA* более удобна работа с диалоговыми окнами, которые теперь разделены на вкладки. Новый интерфейс предоставляет полностью настраиваемые меню и панели инструментов. Вы можете организовать собственные способы доступа к любому элементу системы *STATISTICA*, используя средства автоматизации, средства управления обработкой событий и т.п.
- Язык программирования *SVB*, сочетающий все возможности языков *STATISTICA* BASIC и SCL (реализованных в версии 5.5) с большим количеством новшеств. Все функциональные возможности *STATISTICA 6* доступны из среды Visual Basic в виде стандартной библиотеки объектов.
- Отсутствие практически всех системных ограничений, которые имелись в версии 5.5. Например, файлы данных теперь могут содержать практически неограниченное число переменных. Также сняты ограничения на количество символов в названиях переменных, наблюдений, в текстовых метках. Документы в новой версии системы могут содержать любое число внедренных и связанных объектов.
- Документы нового формата, поддерживающие технологию ActiveX, а также новые оптимизированные средства управления выводом, в том числе мультимедийные таблицы, иерархические Рабочие книги и мощный редактор Отчетов.

- Возможность одновременного открытия нескольких файлов данных в одном или нескольких приложениях *STATISTICA*; средства для запуска нескольких анализов в одной рабочей области.
- Дальнейшее усовершенствование графических возможностей *STATISTICA*. Появились новые типы графиков и сотни новых параметров их настройки. Имеется мощная система управления графическими стилями. Все параметры настройки графиков доступны с помощью *SVB*.
- Новая технология доступа к базам данных, основанная на запросах, которая позволяет работать с очень большими файлами данных.
- Полная интеграция с Internet. Возможности объединения системы *STATISTICA* с другими программными окружениями включают в себя средства для запуска системы *STATISTICA* на сервере в сети и средства доступа к процедурам *STATISTICA* с помощью Internet-браузера. Имеются гибкие средства ввода/вывода через Internet.
- Большое число специализированных усовершенствований. Например, параметры настройки безопасности, новые мощные средства импорта/экспорта, новые опции построения условий выбора наблюдений, интеллектуальные редакторы Отчетов, средства построения графиков и оформления мультимедийных таблиц, возможность отмены и повтора нескольких последних действий и многое другое.

D

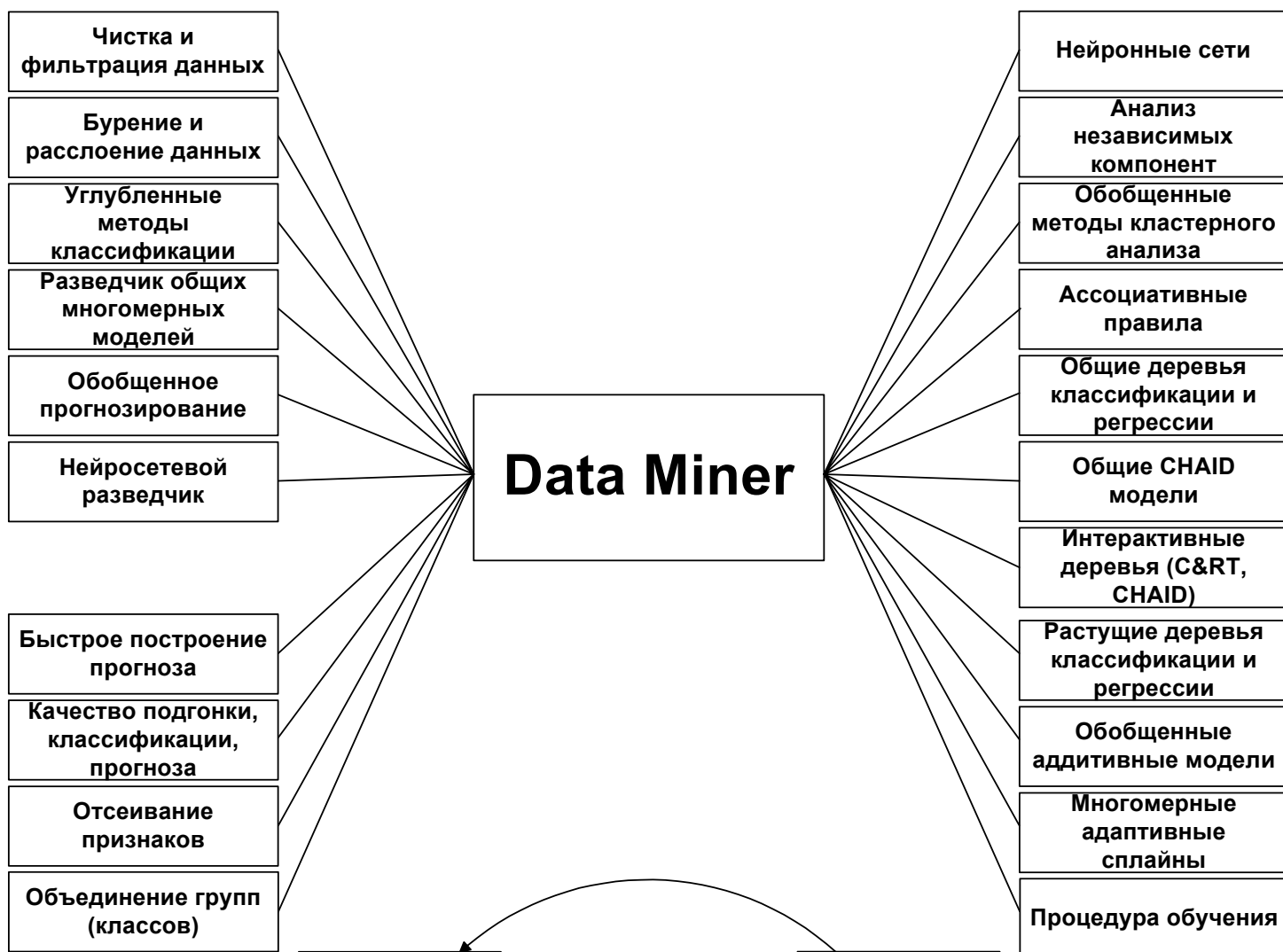
ПРИЛОЖЕНИЕ

НОВЕЙШИЕ МЕТОДЫ АНАЛИЗА ДАННЫХ

Методология “Шесть Сигма” и <i>STATISTICA</i>	191
<i>STATISTICA</i> Data Miner.....	197



НОВЕЙШИЕ МЕТОДЫ АНАЛИЗА ДАННЫХ



НОВЕЙШИЕ МЕТОДЫ АНАЛИЗА ДАННЫХ

Методология “Шесть Сигма” и *STATISTICA*

“Шесть Сигма” - это хорошо структурированная методология управления данными, позволяющая устранить дефекты, снизить потери, организовать контроль качества в различных областях производства, сфере услуг, управлении и других направлениях деятельности. Методология “Шесть Сигма” базируется на хорошо изученных и апробированных статистических методах контроля качества, анализа данных и систематическом тренинге всего персонала организации, вовлеченного в “процесс Шесть Сигма”.

Почему “Шесть Сигма” так популярна? В последние годы интерес к “Шесть Сигма” значительно возрос, поскольку появилось множество примеров успешного внедрения данной методологии как на предприятиях США, так и в Европе.

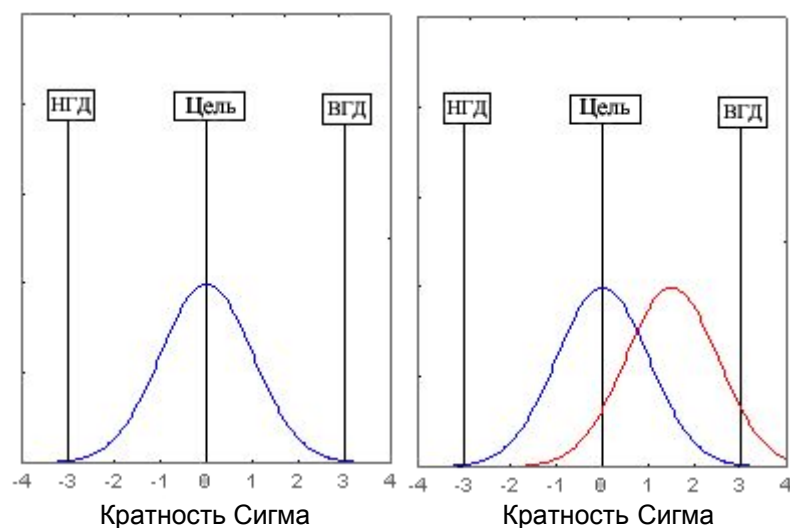
“Шесть Сигма” удачно зарекомендовала себя как стратегия управления качеством. Кроме того, она обеспечивает существенную экономию средств. К примеру, компания Motorola (лидирующий член консорциума по разработке методологии “Шесть Сигма”) сообщила, что за последние 12 лет им удалось сохранить свыше 11 миллиардов долларов. Allied Signals сообщают об 1 миллиарде сбережений за несколько лет благодаря “Шесть Сигма”.

Техническая справка

Термин “Шесть Сигма” (торговый знак принадлежит компании Motorola, в которой была разработана концепция) отражает статистическую цель данного подхода, а именно - стремление достичь пренебрежимо малого числа дефектов.

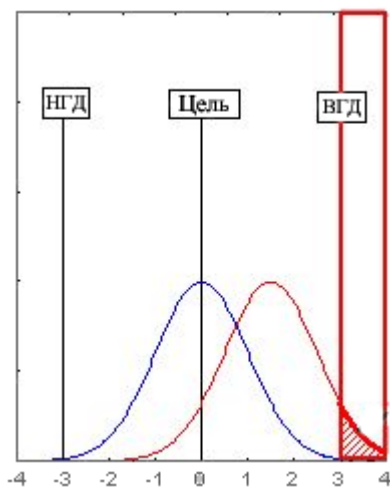
Для достижения данной цели процесс не должен выдавать более, чем 3.4 дефекта на миллион возможностей воспроизвести дефект (где под термином “дефект” понимают наличие любого неприемлемого параметра итоговой продукции, произведенной в условиях тщательного наблюдения за процессом).

Основная идея концепции может быть проиллюстрирована следующим образом. Рассмотрим процесс, в котором распределение отклонений от среднего значения предполагается нормальным и контрольные пределы установлены на интервалах $3 \cdot \text{Сигма}$, где Сигма – оценка стандартного отклонения. Такой процесс носит название “процесс Три Сигма”. Предположим, он отклонился от прежнего среднего на $1.5 \cdot \text{Сигма}$.



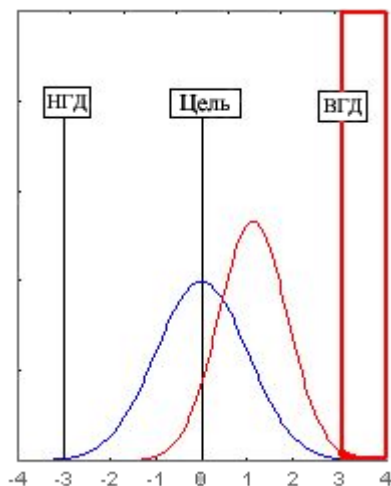
НГД – нижняя граница допуска, **ВГД** – верхняя граница допуска.

Далее, предположим, что мы произвели один миллион деталей и хотим подсчитать число деталей, оказавшееся за верхним контрольным пределом. На рисунке ниже это число соответствует заштрихованной области под графиком распределения процесса.

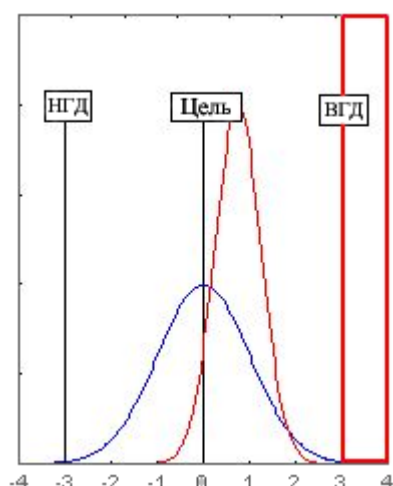


Для процесса “Три Сигма” число деталей вне контрольного предела будет 66.807 (на практике, конечно, число деталей является целым числом, но порядок величины сохраняется).

Число бракованных деталей для процесса “Четыре Сигма”, контрольные пределы для которого отстоят на $4 * \text{Сигма}$ от средней линии, будет равно 6.210.



Для процесса “Шесть Сигма” мы обнаружим всего 3.4 детали, которые оказались вне верхнего контрольного предела.



Термин “Шесть Сигма” возникает при попытке достигнуть такой изменчивости процесса, чтобы значения $\pm 6 * \text{Сигма}$ находились между верхней и нижней границами спецификации (допуска) для процесса. В этом случае, даже если среднее процесса сместится на $1.5 * \text{Сигма}$, число дефектов останется очень малым.

Предположим, что область выше верхней границы спецификации выражена в количестве дефектов на один миллион возможностей произвести дефект. Процесс $6 * \text{Сигма}$, сдвинутый вверх на $1.5 * \text{Сигма}$, будет производить только 3.4 дефекта на миллион возможностей.

Смещение. Процесс может со временем смещаться. Компания Motorola при применении стратегии Шесть Сигма определила, что приемлемое смещение процесса со временем составляет примерно $1.5 * \text{Сигма}$.

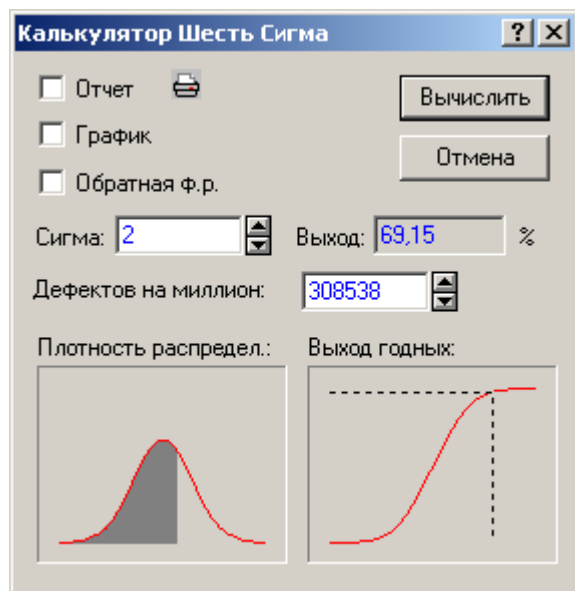
Односторонние и двусторонние пределы. На приведенных выше рисунках область выше верхней границы допуска (выше **ВГД**) определяет количество дефектов на один миллион возможностей произвести дефект. Во многих случаях все наблюдения, которые лежат ниже **НГД**, можно также считать дефектными. В этом случае рассматривается и левый хвост соответствующего (смещенного) нормального распределения.

Тем не менее, на практике обычно игнорируют левый хвост нормальной кривой. Это объясняется тем, что зачастую процесс по своей “природе” имеет односторонние границы допуска (например, очень маленькое время запаздывания не является дефектом в отличие от большого, малое количество жалоб пользователей не является проблемой).

Если вы хотите выполнить детальный анализ областей под кривой нормального распределения, используйте **Вероятностный калькулятор STATISTICA**, который

содержит различные распределения (включая нормальное) и опции для вычисления одно- и двусторонних р-значений.

Выход. В качестве выхода процесса можно рассматривать число “не дефектов”. *Калькулятор Шесть Сигма STATISTICA* вычисляет число дефектов на миллион возможностей и выход, выраженный как $S * 100\%$, где S - площадь области под нормальной кривой, которая находится ниже верхней границы допуска.



Как работает стратегия?

Эффективность концепции “Шесть Сигма” основана на её “эмпирическом” подходе (процедура использует количественные данные о том, как работает система). Процесс реализации так называемого “Проекта улучшения качества в рамках концепции Шесть Сигма” можно условно разделить на следующие стадии: **Определить, Измерить, Анализировать, Улучшить** и **Управлять**. Их называют стадиями “Шесть Сигма *DMAIC*”.

Фаза **Определить** направлена на определение целей проекта и выявление ключевых вопросов, на которые необходимо обратить внимание, чтобы достичь большего уровня Сигма.

Цель фазы **Измерить** – собрать информацию о текущем положении дел, получить ключевые данные о производительности процесса и оценить масштабы проблем.

Цель этапа **Анализировать** – выявить основные причины проблем с качеством и подтвердить результаты с использованием соответствующих средств анализа данных.

Цель фазы **Улучшить** – применение специальных средств устранения проблем (основных причин), выявленных на предыдущем этапе (этапе **Анализировать**).

Цель этапа **Управлять** – оценить результаты предыдущей фазы (фазы **Улучшить**) и наблюдать за текущим процессом.

Существует модификация последовательности стадий Шесть Сигма **DMAIC**, которая называется **DMADV** и применяется при проектировании новых процессов.

В последовательности стадий **DMADV** фаза **Определить** идентична одноименной фазе из **DMAIC**. Фаза **Измерить** направлена на оценивание потребностей клиента и/или рынка. Фаза **Анализировать** направлена на анализ свойств процесса. Фазы **Улучшить** и **Управлять** заменяются соответственно на фазы **Проектировать** (проектирование процесса в соответствии с потребностями заказчика/рынка) и **Проверить** (проверка качества проектирования и возможности удовлетворить критериям, сформулированным на стадии **Проектировать**).

Каждая из этих фаз подразумевает использование специальных аналитических (количественных) методов из большого числа процедур, рекомендуемых концепцией “Шесть Сигма” (в зависимости от природы процесса).

Для получения дополнительной информацией о концепции “Шесть Сигма” обратитесь к следующим авторитетным изданиям, включающим исчерпывающее обсуждение методологии “Шесть Сигма” и её приложений: “*Six Sigma: The Breakthrough management Strategy*” (2000) by M. J. Harry and P. Schroeder и “*The Six Sigma Handbook*” (2001) by T. Pyzdek.

“Шесть Сигма” и *STATISTICA*

Система *STATISTICA* специально спроектирована в соответствии со всеми стадиями Шесть Сигма **DMAIC** и удовлетворяет требованиям по сбору необходимых данных. Таким образом, она является фундаментом для внедрения программ “Шесть Сигма” любой сложности в организациях различного масштаба.

В системе *STATISTICA* доступен исчерпывающий набор средств “Шесть Сигма”, который включает такие инструменты как **Калькулятор Шесть Сигма**, Шесть Сигма - ориентированные отчеты с выводом составных графиков, **Диаграммы причин и следствий**.

С помощью меню **Анализ - Промышленная статистика и Шесть Сигма** системы *STATISTICA* вы можете получить доступ к набору типовых аналитических средств *STATISTICA*, которые доказали свою эффективность в промышленных и бизнес-приложениях. Данные инструменты организованы в группы методов в соответствии с этапами стратегии Шесть Сигма **DMAIC**. Имеется возможность открыть пользовательскую панель инструментов “Шесть Сигма” с пятью подменю, которые отражают пять этапов **DMAIC**.

STATISTICA Data Miner

Что такое Data Mining? Термин Data Mining переводится как “извлечение информации” или “добыча данных”. Нередко рядом с Data Mining встречаются слова knowledge discovery – “обнаружение знаний” и Data Warehouse – “хранилище данных”. Возникновение указанных терминов, которые являются неотъемлемой частью Data Mining, связано с новым витком в развитии средств и методов обработки и хранения данных.

Цель Data mining состоит в выявлении скрытых правил и закономерностей в наборах данных, образно говоря, в “нахождении золотых песчинок в огромной куче руды”. Дело в том, что человеческий разум сам по себе не приспособлен для восприятия больших массивов разнородной информации. В среднем человек, за исключением некоторых индивидуумов, не способен улавливать более двух-трех взаимосвязей даже в небольших выборках.

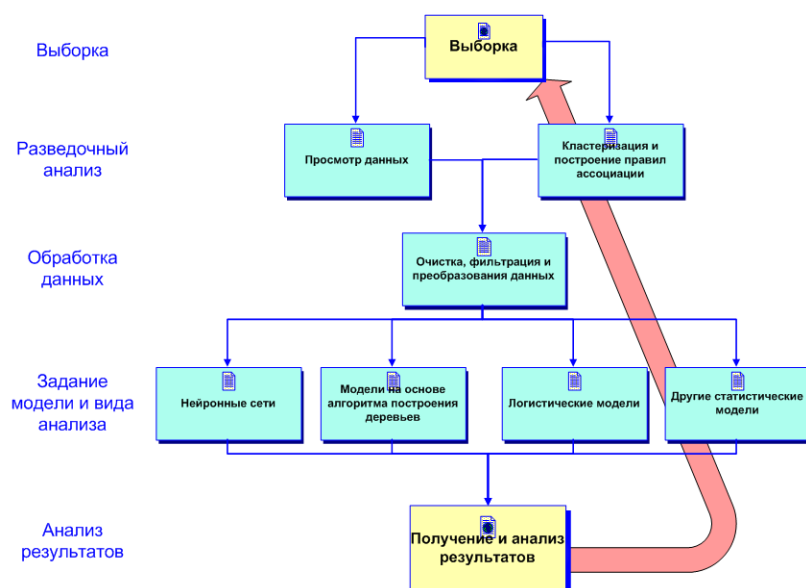
Но и традиционная статистика, долгое время претендовавшая на роль основного инструмента анализа данных, также нередко пасует при решении задач из реальной жизни. Она оперирует усредненными характеристиками выборки, которые часто являются фиктивными величинами (средней платежеспособностью клиента, когда в зависимости от функции риска или функции потерь вам необходимо уметь прогнозировать состоятельность и намерения клиента; средней интенсивностью сигнала, тогда как вам интересны характерные особенности и предпосылки пиков сигнала и т.д.).

Поэтому методы математической статистики оказываются полезными главным образом для проверки заранее сформулированных гипотез, тогда как определение гипотезы иногда бывает достаточно сложной и трудоемкой задачей. Современные технологии Data Mining обрабатывают информацию с целью автоматического поиска шаблонов (паттернов), характерных для каких-либо фрагментов неоднородных многомерных данных.

В отличие от оперативной аналитической обработки данных (OLAP) в Data Mining время формулировки гипотез и выявления необычных (unexpected) шаблонов переложено с человека на компьютер.

Data Mining – это совокупность большого числа различных методов обнаружения знаний. Выбор метода часто зависит от типа имеющихся данных и от того, какую информацию вы пытаетесь получить. Среди имеющихся методов можно выделить следующие: ассоциация (объединение), классификация, кластеризация, анализ временных рядов и прогнозирование, нейронные сети и т.д.

Парадигма Data Mining. Для более полного представления о Data Mining следует рассматривать это понятие как процесс анализа данных, в котором на разных этапах применяются разные методы.



Данный подход облегчает начальное знакомство с возможностями и методами Data Mining. Data Mining представляет собой процесс анализа данных, начиная с выбора, хранения данных и заканчивая анализом результатов, полученных в ходе выполнения данного процесса. Весь процесс можно разбить на 5 основных частей – **Выборка, Разведочный анализ, Обработка и трансформация данных, Задание и применение модели или вида анализа, Анализ результатов.**

Анализируя результаты, полученные после каждой стадии процесса анализа, вы можете перенастроить модель или структуру следующей стадии. В случае получения

неудовлетворительных результатов вы всегда можете вернуться назад и провести некоторые стадии анализа заново либо провести весь процесс с самого начала.

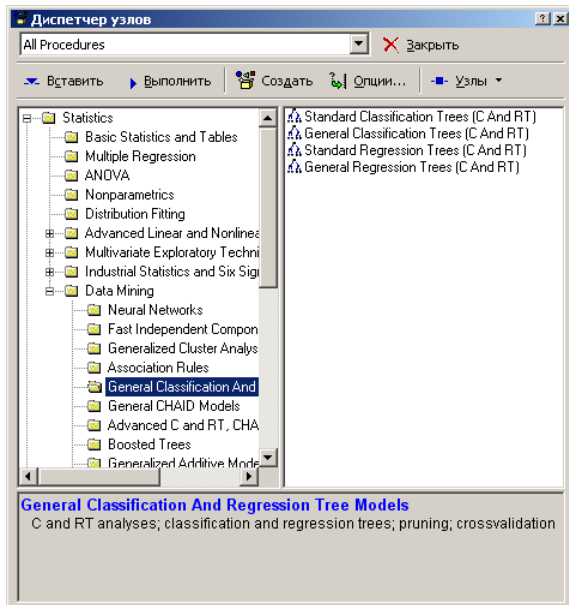
Возможности *STATISTICA Data Miner*. Данная система спроектирована и реализована как универсальное и всестороннее средство анализа данных – от взаимодействия с различными базами данных до создания готовых отчетов.

Чтобы описать все возможности данного пакета, потребуется написать целую книгу. Здесь мы ограничимся только кратким описанием имеющихся в данном пакете основных средств Data Mining.

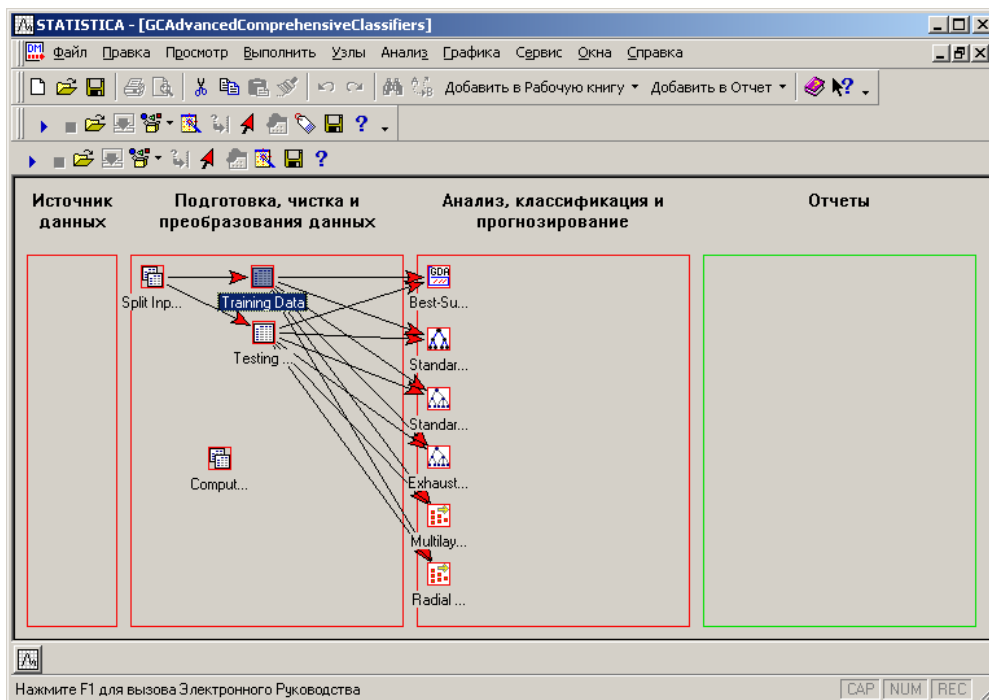
Отличительные черты *STATISTICA Data Miner*:

- Наиболее полный пакет методов Data Mining на рынке программного обеспечения.
- Большой набор готовых решений.
- Удобный пользовательский интерфейс, полностью интегрированный с MS Office.
- Мощные средства разведочного анализа.
- Полностью оптимизированный пакет для работы с огромным объемом информации.
- Гибкий механизм управления.
- Многозадачность системы.
- Чрезвычайно быстрое и эффективное развертывание.
- Открытая СОМ-архитектура, неограниченные возможности автоматизации и поддержки пользовательских приложений (использование промышленного стандарта VB(является встроенным языком), Java, C/C++).

Сердцем *STATISTICA Data Miner* является **Диспетчер узлов**, содержащий более 300 основных процедур, специально оптимизированных под задачи Data Mining, средства логической связи между ними и управления потоками данных. **Диспетчер узлов** позволяет вам конструировать собственные аналитические методы.








Рабочее пространство *STATISTICA* Data Miner состоит из четырех основных частей:



- **Источник данных.** В данной части пользователь идентифицирует источник данных для анализа, будь то файл данных или запрос из базы данных.
- **Подготовка, очистка и преобразования данных.** Здесь данные преобразуются, фильтруются, группируются и т.д.
- **Анализ, классификация и прогнозирование.** Здесь пользователь может при помощи браузера или готовых моделей задать необходимые виды анализа данных, таких как прогнозирование, классификация, моделирование и т.д.
- **Отчеты.** В данной части пользователь может просмотреть, задать вид и настроить результаты анализа (например, Рабочая книга, Отчет или электронная таблица).

Средства анализа STATISTICA Data Miner. В пакете предлагается исчерпывающий набор процедур и методов визуализации.

Средства анализа STATISTICA Data Miner можно классифицировать на пять основных классов:

-  **Бурение и расслоение данных.** Набор процедур, позволяющий разбивать, группировать переменные, вычислять описательные статистики, строить исследовательские графики и т.д.
-  **Углубленные методы классификации.** Полный пакет процедур классификации, включающий в себя обобщенные линейные модели, деревья классификации, регрессионные деревья, кластерный анализ и т.д.
-  **Разведчик общих многомерных моделей.** Содержит линейные, нелинейные, обобщенные регрессионные модели и элементы анализа деревьев классификации.
-  **Обобщенное прогнозирование.** Включает модели АРПСС, сезонные модели АРПСС, экспоненциальное сглаживание, спектральный анализ Фурье, сезонную декомпозицию, прогнозирование при помощи нейронных сетей и т.д.
-  **Нейросетевой разведчик.** Содержит наиболее полный пакет процедур нейросетевого анализа.

Приведенные выше средства анализа являются комбинацией модулей других продуктов StatSoft. Кроме них STATISTICA Data Miner содержит набор специализированных процедур, которые дополняют линейку инструментов Data Mining:

- **Обобщенные методы кластерного анализа.** Данный модуль предназначен для обработки больших наборов данных. Он позволяет кластеризовывать как непрерывные, так и категориальные переменные. Кроме того, он обеспечивает все необходимые функциональные возможности для распознавания образов.

- **Правила связи.** Модуль является реализацией так называемого априорного алгоритма обнаружения правил ассоциации. Например, результат работы этого алгоритма мог бы быть следующим: клиент после покупки продукта А в 95 случаях из 100 в течении следующих двух недель после этого заказывает продукт В или С.
- **Общие деревья классификации и регрессии.** Модуль является полной реализацией методов, разработанных Breiman, Friedman, Olshen и Stone (методов классификации и регрессионных деревьев). Кроме этого, модуль содержит разного рода доработки и дополнения, например, оптимизации алгоритмов для больших объемов данных и т.д.
- **Общие CHAID модели.** Данный модуль является оптимизацией математической модели CHAID для больших объемов данных.
- **Интерактивные деревья (C&RT, CHAID).** Помимо модулей автоматического построения разного рода деревьев STATISTICA Data Miner включает средства для формирования таких деревьев в интерактивном режиме.
- **Растущие деревья классификации и регрессии.** Последние исследования аналитических алгоритмов показывают, что для некоторых задач построения “сложных” оценок, прогнозов и классификаций использование последовательно растущих простых деревьев дает более точные результаты, чем нейронные сети или сложные цельные деревья. Данный модуль реализует алгоритм построения простых растущих деревьев.
- **Обобщенные аддитивные модели.** Набор методов, разработанных и популяризованных Hastie и Tibshirani. Детальное рассмотрение этих методов вы также может найти в работах Schimek (2000).
- **Многомерные адаптивные сплайны (MAR-сплайны).** Данный модуль основан на реализации методики, предложенной Friedman (1991; Multivariate Adaptive Regression Splines, Annals of Statistics, 19, 1-141). В STATISTICA Data Miner расширены опции MARSPLINES для того, чтобы приспособить задачи регрессии и классификации к непрерывным и категориальным предикторам.
- **Быстрое построение прогноза.** Модуль позволяет строить за короткое время классификационные и прогнозирующие модели для большого объема данных. Полученные результаты могут быть непосредственно сохранены во внешней базе данных.
- **Качество подгонки, классификации, прогноза.** Данный модуль производит вычисления различных статистических критериев согласия, как для непрерывных переменных, так и для категориальных.

- **Отсевание признаков.** Данный модуль автоматически выбирает подмножества переменных из заданного файла данных для последующего анализа. Например, модуль может обработать около миллиона входных переменных с целью определения предикторов для регрессии или классификации.

Data Mining включает огромный набор различных аналитических процедур, что затрудняет работу с ним пользователей, которые недостаточно глубоко разбираются в методах анализа данных.

Компания StatSoft нашла выход из этой ситуации. Кроме общих методов анализа в *STATISTICA Data Miner* были встроены готовые модули анализа данных, предназначенные для решения наиболее важных и популярных задач – задач прогнозирования, классификации, создания правил ассоциации и т.д. Поэтому данный пакет могут использовать как профессионалы, так и обычные пользователи, обладающие небольшими опытом и знаниями в анализе данных и математической статистике.

E

ПРИЛОЖЕНИЕ

СЕМЕЙСТВО ПРОДУКТОВ
STATISTICA



Продукты STATSOFT

По применению

Добыча данных

STATISTICA Добытчик данных

STATISTICA Текстовый добытчик

STATISTICA Добытчик качества

STATISTICA Нейронные сети

STATISTICA OLAP

Хранилища данных

STATISTICA Data Warehouse

Базовые модули STATISTICA

STATISTICA Многомерный разведочный анализ

STATISTICA Углубленные методы анализа

Анализ данных

STATISTICA Нейронные сети

STATISTICA Анализ мощности

SEDAS

STATISTICA OLAP

Управление документами

STATISTICA Data Warehouse

STATISTICA Document Management System

STATISTICA OLAP

Контроль качества

STATISTICA Карты контроля качества

STATISTICA Анализ процессов

STATISTICA Планирование экспериментов

STATISTICA Добытчик качества

SEWSS

Продукты STATSOFT

По типу

Однопользовательские

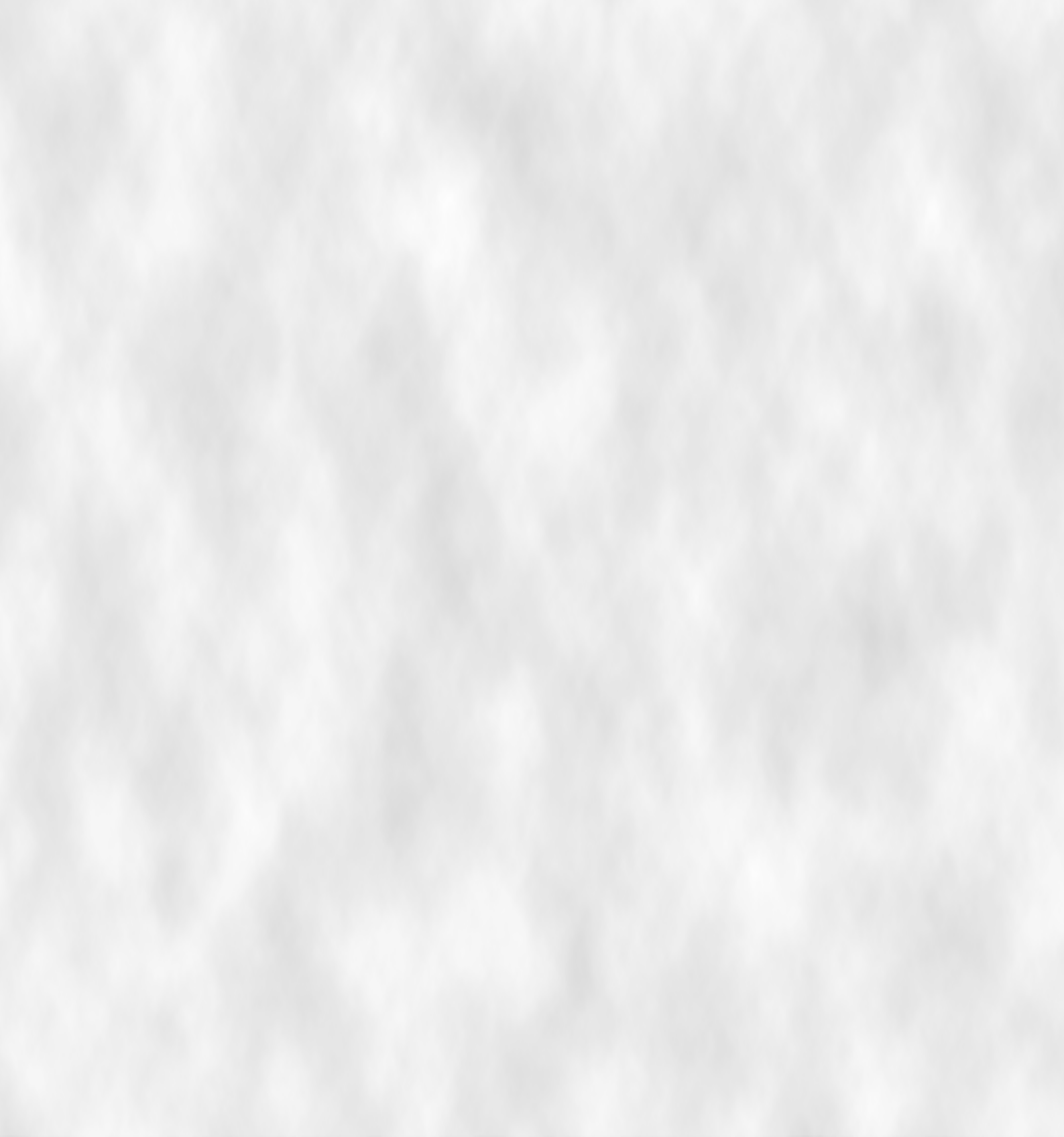
- Базовые модули STATISTICA
- STATISTICA Многомерный разведочный анализ
- STATISTICA Углубленные методы анализа
- STATISTICA Нейронные сети
- STATISTICA Анализ мощности
- STATISTICA Анализ процессов
- STATISTICA Планирование экспериментов
- STATISTICA Карты контроля качества
- STATISTICA Добытчик данных
- STATISTICA Текстовый добытчик
- STATISTICA Добытчик качества
- STATISTICA Document Management System
- STATISTICA OLAP

Корпоративные

- STATISTICA Data Warehouse
- STATISTICA Document Management System
- STATISTICA OLAP
- STATISTICA Добытчик данных
- STATISTICA Текстовый добытчик
- STATISTICA Добытчик качества
- STATISTICA Enterprise-wide Data Analysis System (SEDAS)
- STATISTICA Enterprise-wide SPC System (SEWSS)

Интернет-технологии







СЕМЕЙСТВО ПРОДУКТОВ *STATISTICA*

Основные системные характеристики. Все продукты *STATISTICA* имеют полностью настраиваемый пользовательский интерфейс, поддерживают гибкое управление выводом, высококачественные Отчеты, технологии OLE, ActiveX и Web.

Все продукты *STATISTICA* включают средства оптимизации доступа к большим хранилищам данных, интерактивные средства запросов к базам данных и широкий выбор методов импорта и экспорта.

Все программы поддерживают многозадачность, позволяют использовать несколько исходных файлов данных одновременно, а также получать доступ к наборам данных практически неограниченного объема.

В каждый продукт *STATISTICA* встроены средства интерактивной визуализации, графические и чертежные инструменты высокого качества, а также полный набор средств автоматизации и профессиональная среда разработчика Visual Basic.

Базовые модули *STATISTICA* (отдельный продукт). Предоставляет широкий набор основных статистик вместе с удобным пользовательским интерфейсом. Сочетает в себе простоту и мощность технологии *STATISTICA* (см. выше).

- Все графические средства *STATISTICA*
- Основные статистики и таблицы
- Подгонка распределений
- Множественная регрессия
- Дисперсионный анализ
- Непараметрическая статистика и др.

STATISTICA Углубленные методы анализа (дополнительный продукт).

Предоставляет широкий выбор передовых средств моделирования и прогнозирования, включая возможности автоматического выбора модели и инструменты интерактивной визуализации наивысшего качества.

- Общие линейные модели
- Обобщенные линейные и нелинейные модели
- Общие регрессионные модели
- Общие модели частных наименьших квадратов
- Компоненты дисперсии
- Анализ выживаемости
- Нелинейное оценивание
- Множественная нелинейная регрессия
- Логлинейный анализ таблиц частот
- Временные ряды и прогнозирование
- Моделирование структурными уравнениями и др.

STATISTICA Многомерный разведочный анализ (дополнительный продукт).

Предоставляет широкий выбор методов исследования различных типов данных со всевозможными интерактивными графическими средствами.

- Кластерный анализ
- Факторный анализ
- Анализ главных компонент и классификация
- Канонический анализ
- Надежность и позиционный анализ
- Деревья классификации
- Анализ соответствий
- Многомерное шкалирование
- Дискриминантный анализ
- Общие модели дискриминантного анализа и др.

STATISTICA Нейронные сети (отдельный или дополнительный продукт).

Предоставляет исчерпывающий набор методов построения нейронных сетей с помощью мастера решений и других автоматических инструментов; дополнительно предлагается генератор Си-кода.

- Мастер решений
- Многослойные персептроны
- Радиальные базисные и многие другие функции
- Самоорганизующаяся карта Кохонена
- Алгоритм обратного распространения
- Метод сопряженных градиентов
- Аналитические графики
- Повторные выборки (Кросс-проверка, Бутстреп)
- Анализ чувствительности, кривые операционных характеристик
- Ансамбли
- Автоматический поиск наилучшей сети
- Интерфейс API и др.

STATISTICA Анализ мощности (дополнительный продукт). Предельно точный и удобный для пользователя специализированный инструмент для всестороннего анализа статистической мощности и вычисления объема выборки.

- Оценка объема выборки
- Оценка доверительного интервала
- Калькулятор вероятностных распределений и др.

Промышленные решения, Шесть Сигма**STATISTICA Карты контроля качества (отдельный или дополнительный продукт).**

Предоставляет полностью настраиваемые (в частности, доступные из других приложений), простые в использовании графики контроля качества, а также набор средств автоматизации и горячих клавиш для упрощения обычной работы (инструмент для реализации методов Шесть Сигма).

- X- bar и/или R карты; S^2 , Np, P, U, C карты
- Карты Парето
- Индексы пригодности процесса и индексы производительности
- Карты скользящего среднего и скользящих размахов, EWMA карты
- Номинальные карты
- Целевые карты
- Краткие карты
- CuSum (Накопленные суммы)
- Критерии серий
- Многопоточные карты и др.

STATISTICA Анализ процессов (дополнительный продукт). Удобный пакет для анализа пригодности процессов, повторяемости и воспроизводимости измерений, а также других приложений контроля/повышения качества (инструмент для реализации методов Шесть Сигма).

- Анализ Вейбулла
- Анализ повторяемости и воспроизводимости измерений
- Компоненты дисперсии для случайных эффектов
- Планы выборочного контроля и др.

STATISTICA Планирование экспериментов (дополнительный продукт). Огромный выбор методов планирования экспериментов и связанных с ними графических средств, включая интерактивные профили желательности (инструмент для реализации методов Шесть Сигма).

- Дробные факторные планы
- Планы для смесей
- Латинские квадраты
- Поиск оптимальных 2^{k-p} планов
- Анализ остатков и преобразования
- Оптимизация переменных простого/многомерного отклика

- Центральные композиционные планы
- Планы Тагучи
- Минимальная аберрация и максимальная несмешанность
- Дробные 2^{k-p} факторные планы с блоками
- Поверхности с ограничениями
- D- и A-оптимальные планы
- Профили желательности и др.

Корпоративные системы *STATISTICA*

Корпоративные системы *STATISTICA* предлагают общие функциональные возможности, перечисленные выше. В дополнение к ним они предоставляют широкий выбор инструментов для коллективной работы, Web-ориентированный пользовательский интерфейс (с использованием поставляемого по отдельному заказу *STATISTICA Web Server*), специализированные базы данных и высоко оптимизированные средства доступа к корпоративным хранилищам. Последние включают опции для быстрой обработки больших объемов данных с удаленных серверов “на месте” (без создания локальной копии).

STATISTICA Enterprise-Wide Data Mining System (Data Miner). Наиболее полный пакет методов добычи данных на рынке программного обеспечения с исключительно простым и удобным пользовательским интерфейсом.

Включает интегрированные системы решений для широкого круга задач, как полностью готовые к использованию, так и легко настраиваемые. В системе предусмотрены дополнительные надстройки и возможность обучения.

Средства анализа *STATISTICA Data Miner* можно классифицировать на пять основных классов:

- Разметка/Разбиение и OLAP
- Классификация
- Обобщенные линейные, нелинейные и регрессионные модели
- Прогнозирование
- Нейросетевой анализ

STATISTICA Enterprise-Wide Data Analysis System (SEDAS). Интегрированная многопользовательская система для решения общих задач анализа данных и для бизнес-приложений в области маркетинга, финансов и т.д.

SEDAS может дополнительно предоставлять все функциональные возможности, доступные в каком-либо из перечисленных выше продуктов *STATISTICA* или во всех этих продуктах. Кроме того, предусмотрены:

- Интеграция с хранилищами данных
- Интуитивные средства запросов и фильтрации
- Простые в использовании функции администрирования
- Автоматическое распространение Отчетов
- Извещение о тревоге и др.

STATISTICA Enterprise-Wide SPC System (SEWSS). Предназначена для локальных и глобальных корпоративных приложений по контролю и улучшению качества, включая методику Шесть Сигма.

Предоставляет возможности сбора данных в реальном времени, оповещения о тревоге в случае наступления заданного события. Включает исчерпывающий набор аналитических средств для инженеров, сложные средства построения Отчетов для менеджеров, опции для построения Отчетов Шесть Сигма.

- Web-ориентированный пользовательский интерфейс и средства для создания Отчетов; интерактивные инструменты построения запросов
- Специальные пользовательские интерфейсы для операторов, инженеров, менеджеров, аналитиков и т.д.
- Групповое программное обеспечение для совместного использования запросов, специальных приложений и т.д.
- Система оповещения о нарушениях, включающая запросы причин/действий
- Масштабируемость, настраиваемость, совместимость с существующими СУБД и ERP системами и др.

Основная литература

1. Боровиков В. П. Программа STATISTICA для студентов и инженеров. М., 2001.
2. Боровиков В. STATISTICA. Искусство анализа данных на компьютере: Для профессионалов. 2-е изд. (+CD). СПб., 2003.
3. Боровиков В. П., Ивченко Г. И. Прогнозирование в системе *STATISTICA*[®] в среде Windows. М., 2000.
4. Кендалл М., Стьюарт А. Теория распределений. М., 1966.
5. Кендалл М., Стьюарт А. Статистические выводы и связи. М., 1973.
6. Кендалл М., Стьюарт А. Многомерный статистический анализ и временные ряды. М., 1976.
7. Розанов Ю. А. Теория вероятностей, случайные процессы, математическая статистика. М., 1985.
8. Ширяев А. Н. Вероятность. М., 1980.

Дополнительная литература

1. Андерсон Т. Введение в многомерный статистический анализ. М., 1963.
2. Андерсон Т. Статистический анализ временных рядов. М., 1976.
3. Афифи А., Эйзен С. Статистический анализ. Подход с использованием ЭВМ. М., 1982.
4. Беляев Ю. К. Вероятностные методы выборочного контроля. М., 1975.
5. Беляев Ю. К., Чепурин Е. В. Основы математической статистики. М., 1983.
6. Бикел П., Доксам К. Математическая статистика. М., 1983. Вып. 1, 2.
7. Большев Л. Н., Смирнов Н. В. Таблицы математической статистики. М., 1983.
8. Боровков А. А. Теория вероятностей. М., 1976.
9. Боровков А. А. Математическая статистика. Оценка параметров, проверка гипотез. М., 1984.
10. Боровков А. А. Математическая статистика. Дополнительные главы. М., 1984.

11. Вальд А. Последовательный анализ. М., 1960.
12. Гаек Я., Шидак З. Теория ранговых критериев. М., 1974.
13. Гихман И. И., Скороход А. В. Введение в теорию случайных процессов. М., 1977.
14. Гнеденко Б. В. Курс теории вероятностей. М., 1974.
15. Де Гроот М. Оптимальные статистические решения. М., 1974.
16. Драйпер Н., Смит Г. Прикладной регрессионный анализ: В 2-х кн. М., 1987.
17. Дэйвид Г. Порядковые статистики. М., 1979.
18. Закс Ш. Теория статистических выводов. М., 1975.
19. Ивченко Г. И., Глибоченко А. Ф., Иванов В. А., Медведев Ю. И. Статистический анализ дискретных случайных последовательностей. М., 1984.
20. Ивченко Г. И., Медведев Ю. И. Математическая статистика: Учеб. пособие для вузов. М., 1992.
21. Климов Г. П. Теория вероятностей и математическая статистика. М., 1983.
22. Климов Г. П., Кузьмин А. Д. Вероятность, процессы, статистика. Задачи с решениями. М., 1985.
23. Кокс Д., Снелл Э. Прикладная статистика. Принципы и примеры. М., 1984.
24. Колмогоров А. Н. Основные понятия теории вероятностей. М., 1974.
25. Крамер Г. Математические методы статистики. М., 1975.
26. Леман Э. Проверка статистических гипотез. М., 1964.
27. Липцер Р. Ш., Ширяев А. Н. Статистика случайных процессов. М., 1974.
28. Лозэв М. Теория вероятностей. М., 1962.
29. Мостеллер Ф., Тьюки Дж. Анализ данных и регрессия. М., 1982.
30. Прикладная статистика: Классификация и снижение размерности: Справ. изд. / С. А. Айвазян, В. М. Бухштабер, И. С. Енюков, Л. Д. Мешелкин. М., 1989.
31. Рао С. Р. Линейные статистические методы и их применения. М., 1968.
32. Розанов Ю. А. Случайные процессы. М., 1979.
33. Розанов Ю. А. Введение в теорию случайных процессов. М., 1982.
34. Себер Дж. Линейный регрессионный анализ. М., 1980.

35. Севастьянов Б. А. Курс теории вероятностей и математической статистики. М., 1982.
36. Уилкс С. Математическая статистика. М., 1967.
37. Уиттекер Э. Т. и Ватсон Дж. Н. Курс современного анализа. Т. I, II. М., 1963.
38. Феллер В. Введение в теорию вероятностей и ее приложения. Т. I, II. М., 1984.
39. Хемминг Р. Численные методы для научных работников и инженеров. М., 1972.
40. Хеннан Э. Анализ временных рядов. М., 1964.
41. Хеннекен П. А., Тортра А. Теория вероятностей и некоторые ее приложения. М., 1974.
42. Чистяков В. П. Курс теории вероятностей. М., 1987.
43. Шеффе Г. Дисперсионный анализ. М., 1963.
44. Greene W. H. Econometric analysis. Third edition. Prentice-Hall International, 1997.

**ПРЕДМЕТНЫЙ
УКАЗАТЕЛЬ**

A

ActiveX, 109, 119
документы, 163
объекты, 163

C

C/C++, 162, 175

E

Excel, 163

J

Java, 162, 175

O

OLE объекты, 121

S

SEDAS, 7, 214
SEWSS, 7, 214
STATISTICA Advanced
Linear/Nonlinear Models,
210
STATISTICA Base, 5, 209
STATISTICA Data Miner, 197
STATISTICA Design of
Experiments, 6, 212
STATISTICA Enterprise
systems, 213
STATISTICA Enterprise-wide
Data Analysis System, 7,
214
STATISTICA Enterprise-wide
Data Mining System, 213
STATISTICA Enterprise-wide
SPC System, 7, 214

STATISTICA Multivariate
Exploratory Techniques,
210
STATISTICA Neural Networks,
5, 211
STATISTICA Power Analysis,
6, 211
STATISTICA Process Analysis,
6, 212
STATISTICA Quality Control
Charts, 6, 211
STATISTICA Query, 167
STATISTICA Visual Basic, 94,
123, 157
выполнение программ,
162
методы, 94
свойства, 94
создание графиков, 133
создание программ, 158
STATISTICA Web-сервер, 97,
177

V

Visual Basic
методы, 94
свойства, 94

W

Web, 96, 176
Web-сервер (*STATISTICA*),
97, 177
Word (Microsoft), 163

X

XML, 177

A

автономные окна, 106
аналитическое средства, 173

Б

браузер (Internet), 96, 176
быстрые меню, 48
быстрый анализ, 51
быстрый и углубленный
анализ, 51

B

вводный пример, 45
выбор файла данных, 46
вывод результатов, 49

Г

графики, 127
STATISTICA Visual Basic,
133
блоковые данные, 136,
140
исходные данные, 136,
137
меню, 142
настройка параметров,
128, 130, 152
пользовательские, 141
специализированные, 144
ссылки, 135
типы, 136
графики блоковых данных,
140
графики исходных данных,
137
графическое окно
настройка, 63
построение графиков для
электронных таблиц,
61

Д

дерево Отчета, 121
дисперсионный анализ -
пример, 69
график средних, 74

- задание кодов для
 межгрупповых
 факторов, 71
задание плана, 69
задание повторных
 измерений, 70
запуск модуля
 Дисперсионный
 анализ, 69
окно результатов, 72
переменные, 70
таблица всех эффектов,
 73
- Ж**
- журнал анализа, 158
- З**
- заголовок окна, 115
запись журнала анализа, 158
запись макросов, 158
запрос, 167
запуск *STATISTICA*, 45
- И**
- имена наблюдений, 116
имена переменных, 116
интерфейс пользователя, 79
 интерактивный, 83
 настройка, 79
 основные
 характеристики, 79
 сеанс анализа, 84
исходные таблицы данных,
 117
- К**
- комментарии в рабочих
 книгах, 103
контекстная подсказка, 53
контекстные меню, 48
копирование, 57
- копирование с заголовками,
 57
корреляции, пример, 45
 выбор переменных, 53
 выбор файла данных, 46
 диалоговое окно выбора
 переменной, 53
 настройка графика, 63
 открытие модуля
 Основных статистики
 и таблицы, 50
отчет по результатам, 57
парные и частные
 корреляции, 51
переменные, 53
построение графиков для
 электронных таблиц,
 61
рабочая книга
 результатов, 55
таблица результатов, 55
- М**
- макрос
 анализа, 158
 клавиатурный, 159
 Мастер-макрос, 158
Мастер-макрос, 158
матрица корреляций, 50
методология Шесть Сигма и
 STATISTICA, 191
многозадачный режим, 80
модуль Множественная
 регрессия, 17
модуль Непараметрическая
 статистика, 31
модуль Общий
 Дисперсионный анализ,
 23
модуль Основные статистики
 и таблицы, 8
модуль Подгонка
 распределений, 37
мультимедийные таблицы, 46
- Н**
- наблюдения, кнопка панели
 инструментов, 46
настройка, 79
 SVB, 94, 123, 157
 внешний вид
 STATISTICA, 147
временная и постоянная,
 150
графики, 63, 128, 152
другие приложения, 94
интерфейс пользователя,
 148
конфигурации для
 нескольких
 пользователей, 153
общие параметры
 системы, 151
панели инструментов, 92
различные
 конфигурации, 153
различные способы
 доступа к одним и
 тем же
 возможностям, 80
функционалирование
 STATISTICA, 147
наиболее новые методы анализа
 данных, 191
новые возможности, 187
- О**
- обзор функциональных
 возможностей, 173
аналитические средства,
 173
используемые
 технологии
 программного
 обеспечения, 175
общая философия работы
 в *STATISTICA*, 174
уникальные
 возможности, 173

общие стандартные
настройки, 151

Отзывы и награды, 177

Отчет, 57, 104, 119
дерево, 121

П

панель инструментов, 84, 91
настраиваемая, 92
пользовательская, 92

параметры

авто скрыть, 88
выводить при выборе, 89
скрывать при выборе, 89
скрыть поле Отчета, 90

переменная

выбор, 53
спецификации, 47

переменные, кнопка панели
инструментов, 46

печать таблиц данных, 57

получение подробной
справки, 181

пользовательские графики,
141

пример

1 - Корреляции, 45
2 - Дисперсионный
анализ, 69

Р

работа с пакетом *STATISTICA*
из других приложений,
94

Рабочая книга, 102

дерево, 111
исходные таблицы
данных, 118
обзор, 109

Рабочие книги анализа, 55

С

семейство продуктов
STATISTICA, 209

особенности, 209

семейство продуктов,
STATISTICA, 5, 209

создание графиков с
помощью *STATISTICA*
Visual Basic, 133

стартовая кнопка

STATISTICA, 90

стартовая панель, 84

статистический советник, 67

Т

таблица результатов, 55

техническая поддержка, 181

технология программного
обеспечения, 175

типы документов, 90

панели инструментов, 91

тонкий клиент, 177

У

углубленный анализ, 51

удаленные базы данных, 169

удаленные серверы, 169

уникальные возможности,
173

управление выводом, 85

управление выводом
результатов, 49

Ф

философия работы в
STATISTICA, 174

формулы (в электронных
таблицах), 47

Э

Электронное руководство, 66,
181

электронные таблицы, 46

метод перетащить и
отпустить, 65

разделение прокрутки, 64
формулы, 47

