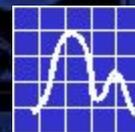


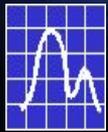
STATISTICA

Системный подход к анализу данных.

Владимир Боровиков



StatSoft® Russia



STATISTICA 2000

Windows 95, 98, NT, 2000 Ready

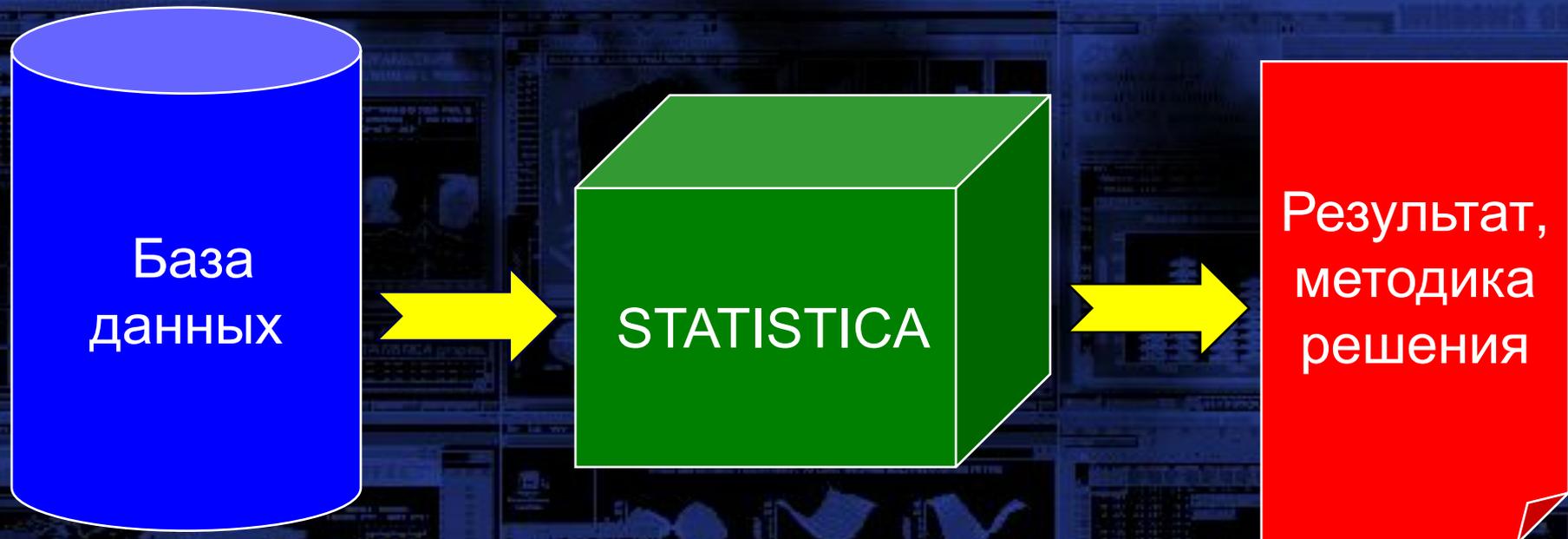
Полностью на русском языке

Около 30 Mb на винчестере

Существует Macintosh-версия

Клиент-серверный интерфейс

Законченные решения от StatSoft

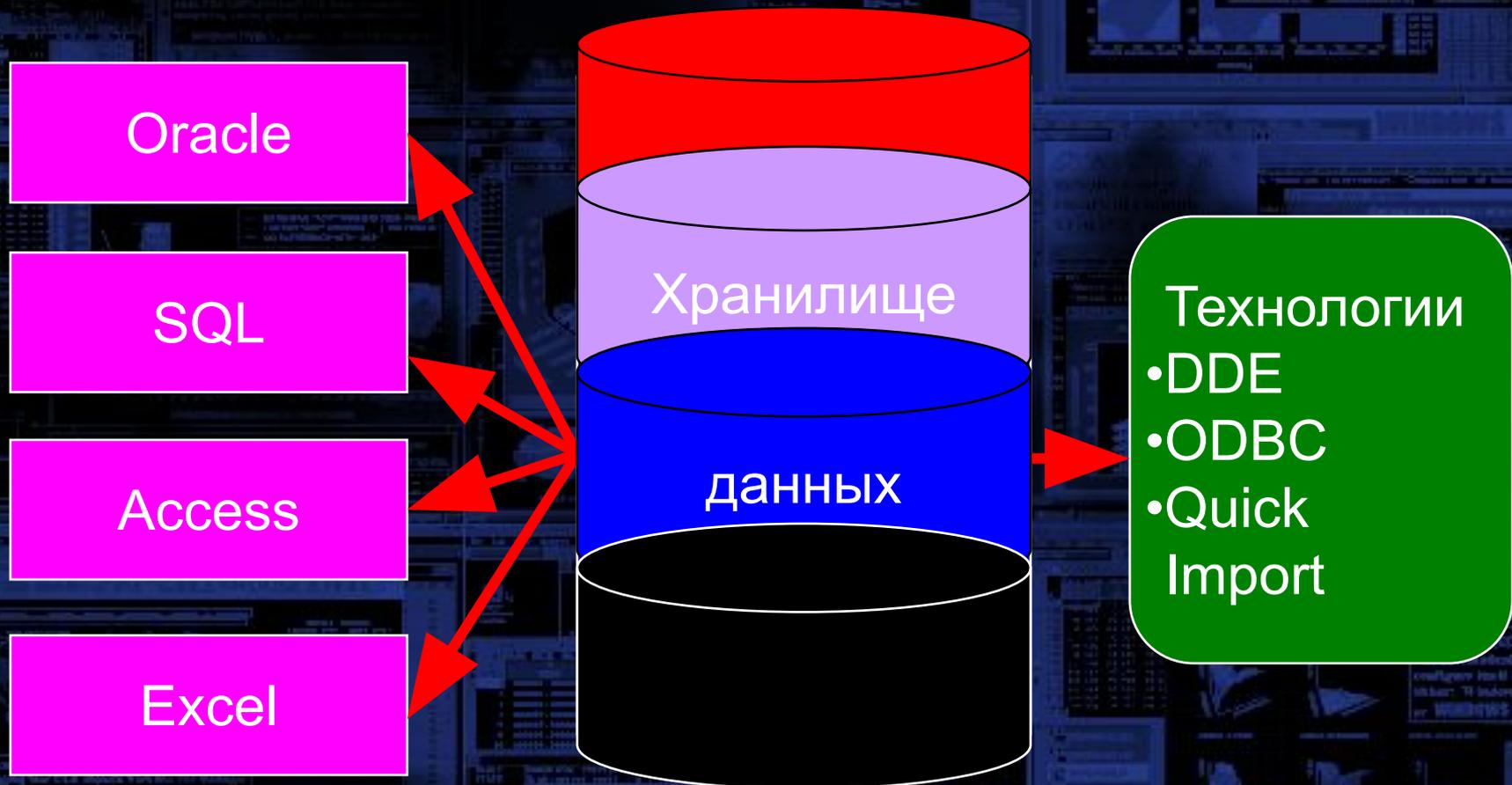




BusinessQuery
for
STATISTICAL



Законченные решения от StatSoft



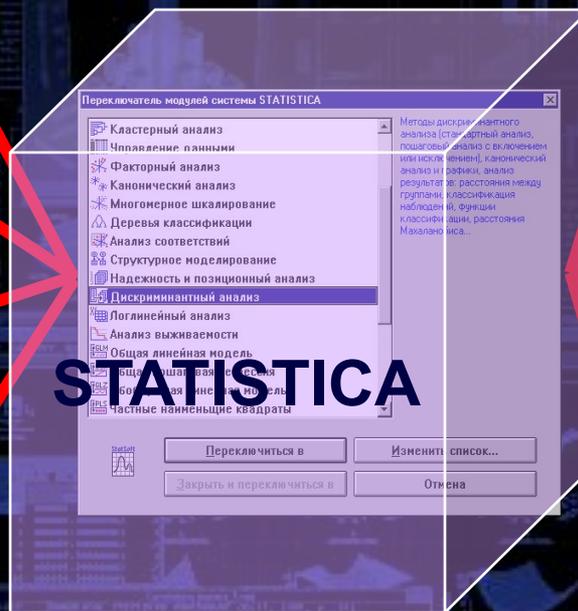
Законченные решения от StatSoft

Верификация
данных

Агрегировани
е
данных

Визуализация

Разведочный
анализ



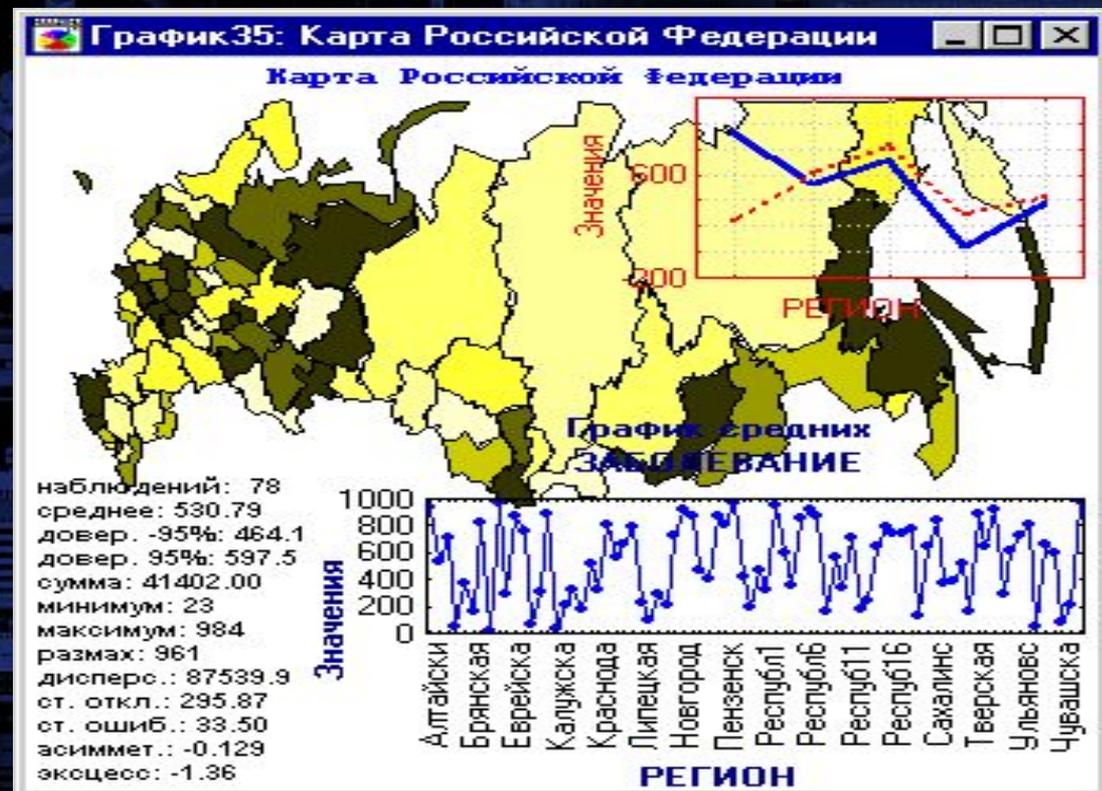
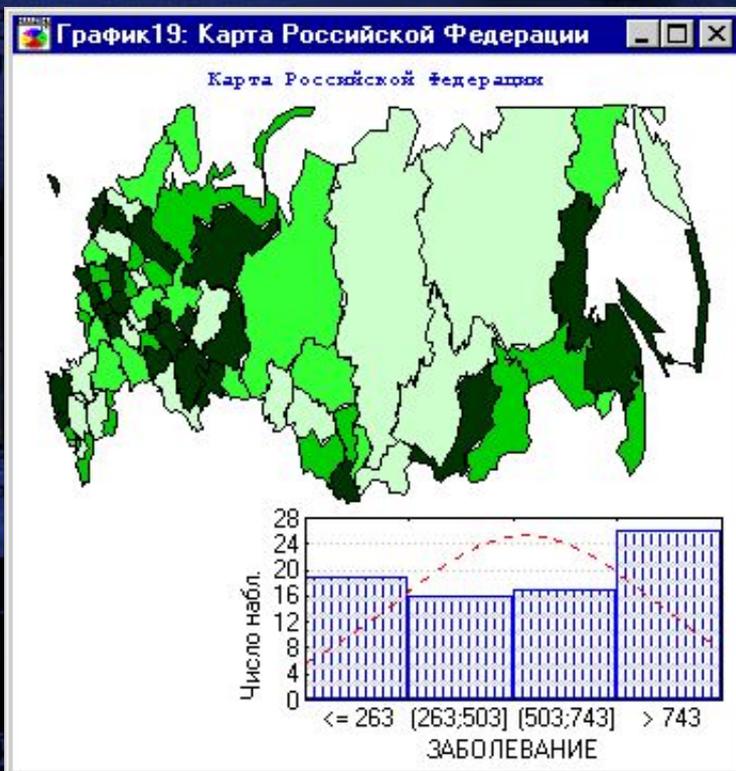
Оценка
репрезентативности
и
выборки

Применение
специальных
методов

Представление
результатов

Наложение результатов на географическую карту

Уникальный
проект
StatSoft



- **Департаменты ЦБ**
- **Страховая компания РОСНО**
- **АК АПРОСА (АЛМАЗЫ РОССИИ-САХА)**
- **Министерство Путей Сообщения**
- **Акционерное общество АВТОВАЗ**
- **Региональные подразделения Госкомстата**
- **Энергетическая компания Карелэнерго**
- **Авиатранспортное предприятие EASTLINE**
- **Телекоммуникационная компания**
“Северо-западное GSM”
- **Группа “Сибирский алюминий”**

Статистический Портал StatSoft

Учебно-образовательный и информационный Портал по анализу данных, визуализации, классификации, прогнозированию и разработке приложений в медицине, промышленности, экономике, финансах, образовании, интернете...

Портал StatSoft

Этот Портал создан под руководством известного в России и за рубежом специалиста в области анализа данных д-ра Владимира Боровикова, автора многочисленных книг по новейшим технологиям анализа данных, визуализации, классификации и прогнозированию. Полная, ясная и точная информация о современных методах анализа собрана на Портале и структурирована по темам, методам и областям применения, позволяет находить адекватные решения конкретных проблем. Профессиональность

$$y=161.05+180.311*\log_{10}(x)+\epsilon_{tpeps}$$

Визуальный анализ



Медицина

...одна из областей, активно применяются статистики и анализа

...емские исследования классическими исследованиями с использованием статистических методов. В этих исследованиях были собраны проявления проявлений такой болезни у 1929 2540 женщин...

[Непараметрические методы](#), [Планирование эксперимента](#), [Факторный анализ](#), [Статистический словарь](#), [Поиск в учебнике](#), [Таблицы распределений](#).

Углубленный учебник по статистике и анализу данных (учебник с формулами)

Позволяет углубленно и всесторонне изучить методы статистики и анализа данных, начиная с основных вероятностных моделей и формул. Этот учебник основан на знаменитом учебнике Иваница Г.И., Морозова Ю.И. "Математическая статистика", который

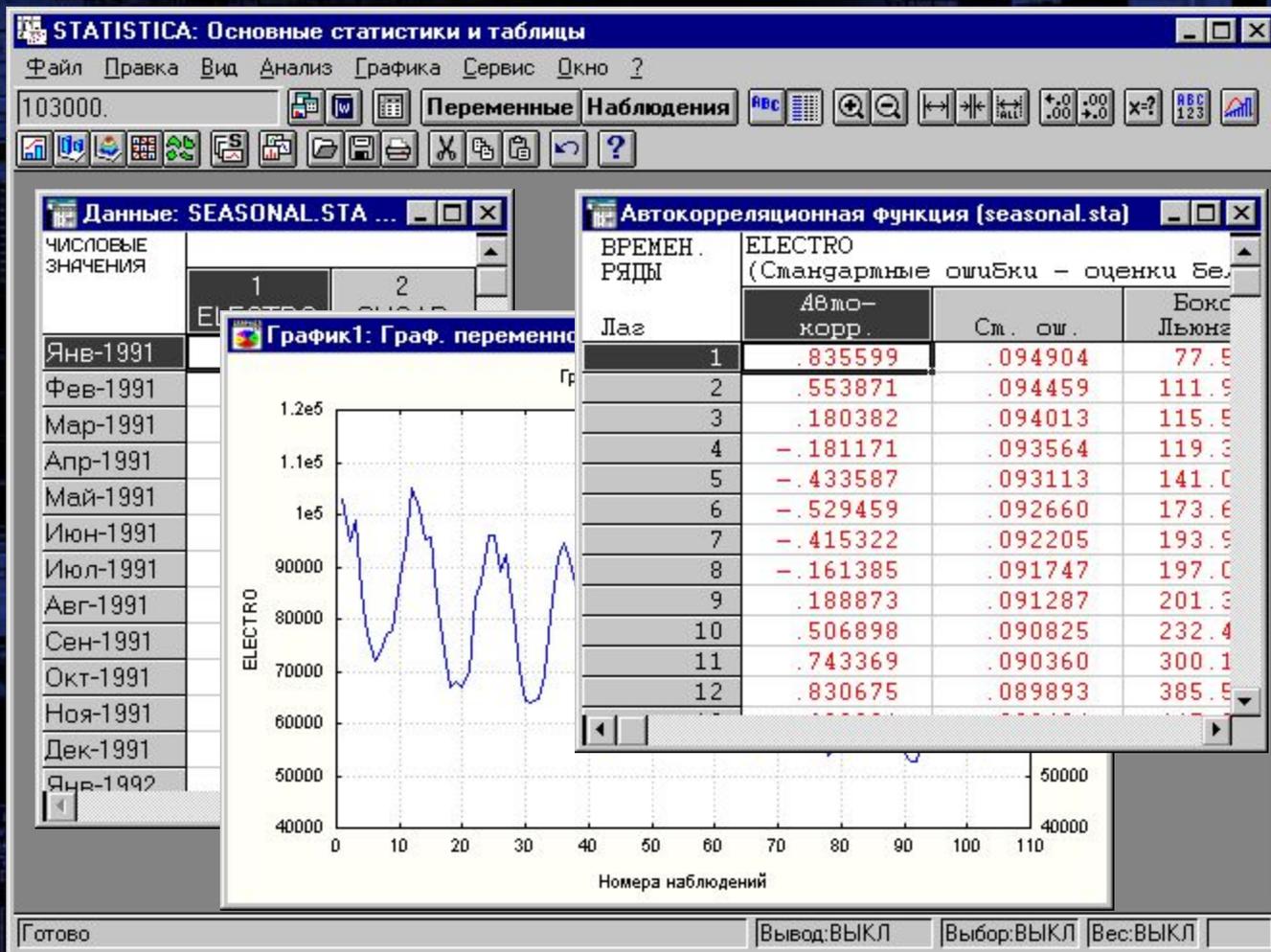
[Модели Кокса](#) используются в медицине для оценки условного риска заболевания при заданных значениях исходных признаков...

Статистика находит применение при исследовании действия различных препаратов в [фармакологии](#) и [токсикологии](#)...

В работе практического врача



Структура пакета

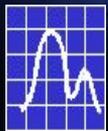


Основное
окно

Таблицы
данных

Графики

Таблицы
результатов



Импорт данных

Прямое преобразование файлов из большинства популярных форматов:

- + Excel
- + Lotus (1-2-3, Symphony)
- + Quattro Pro
- + xBase (dBase, FoxPro, Clipper)
- + Paradox
- + ASCII



Импорт данных

Быстро
и просто

ЧИСЛОВЫЕ ЗНАЧЕНИЯ	Данные: SEASONAL.STA ...	
	1 ELECTRO	2 SUGAR
Янв-1991	103000.0	224000.0
Фев-1991	95000.00	117000.0
Мар-1991	99000.00	172000.0
Апр-1991	85000.00	198000.0
Май-1991	77000.00	250000.0
Июн-1991	72000.00	262000.0
Июл-1991	74000.00	189000.0
Авг-1991	77000.00	110000.0
Сен-1991	78000.00	353000.0
Окт-1991	88000.00	682000.0
Ноя-1991	94000.00	424000.0
Дек-1991	105000.0	146000.0
Янв-1992	101200.0	64100.00
Фев-1992	95300.00	57600.00
Мар-1992	95500.00	143800.0
Апр-1992	84000.00	331900.0
Май-1992	74000.00	376100.0
Июн-1992	67000.00	289500.0
Июл-1992	68000.00	313900.0

Выбор импортируемого файла

Имя файла: DATA_TEST.XLS

Папки: d:\#users\cbr

DATA_TEST.XLS

Выбор листа для импорта

Список листов в файле Excel:

Быстрый импорт из Excel - Параметры

OK

Отмена

Network...

Сохранение импортированного файла

Имя файла: :#users\cbr\Data test.STA

Папки: d:\#users\cbr

BLACK.STA
BLACK_IS.STA
SEASONAL.STA

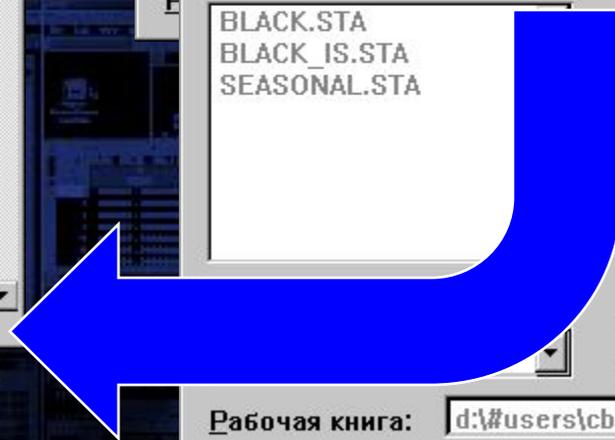
Диски: d:

Рабочая книга: d:\#users\cbr\Data test.xls

OK

Отмена

Network...





Импорт данных

Поддержка интерфейса открытых баз данных Microsoft ODBC:

- + MS Access
- + MS SQL Server
- + Oracle
- + Sybase
- + IBM DB2/2, DB2/6000
- + сотни других популярных СУБД



Импорт данных

Удобно

Схема данных

CSMINN
NODEID
MONID

CSMOPD
MONID
SEQNO

CSPINN
NODEID
PROFID

CSNODE_1
NODEID
NAME

CSODAT
CHRTID
OPTID
ANAL

Select Data Source

File Data Source Machine Data Source

Импорт через ODBC - Выбор полей и условий выбора строк

Таблицы:

- CSVALUE
- CSVDAT
- CSVINL
- CSVLRN
- Data
- Specs
- CSALDATV
- CSCHRTV
- CSCINNV

Поля:

- Product
- Center Line
- Upper Control Limit
- Lower Control Limit

Поля для импорта:

- Data.Product
- Specs.Center Line
- Data.Test1
- Specs.Product

Добавить >> Вырезать

Выбор строк

- Выбор всех
- КУДА... (использовать выражение для выбора строк, с использованием символа % как максиса SQL)

Data.Product=Specs

Данные: QCTEST.STA 3п * 3...

	1 CENTER L	2 TEST1	3 PRODUCT
1	100.000	98.927	BN111
2	100.000	102.039	BN111
3	100.000	95.203	BN111
4	100.000	102.932	BN111
5	100.000	100.092	BN111
6	100.000	98.798	BN111
7	100.000	106.468	BN111
8	100.000	107.068	BN111
9	100.000	112.649	BN111
10	100.000	103.227	BN111
11	50.000	46.884	BN222
12	50.000	50.548	BN222
13	50.000	46.711	BN222
14	50.000	49.084	BN222
15	50.000	50.622	BN222
16	75.000	78.339	BN333
17	75.000	75.760	BN333
18	75.000	79.222	BN333
19	75.000	74.288	BN333



Подготовка данных

Данные Стандартизация Подмножества Формат

Данные: ADSTUDY STA 25n * 50n

Адвертайзинг Эффективность Исследования

	1	2	3	4	5	6	7	8	9	10	11
ГЕНДЕР	ADVERT	MEAS	MEAS	MEAS	MEAS	MEAS	MEAS	MEAS	MEAS	MEAS	MEAS
J. Baker	MALE	PEPSI	9	1	6	8	1	2	1	3	8
A. Smith	MALE	COKE	6	7	1	8	0	0	6	8	9
M. Brown	FEMALE	COKE	9	8	2	9	8	8	0	3	2
C. Mayer	MALE	PEPSI	7	9	0	5	9	9	6	6	0
M. West	MALE	PEPSI	7	1	6	2	8	9	6	4	0
D. Young	FEMALE	COKE	6	0	0	8	3	1	0	7	8
S. Bird	FEMALE	COKE	7	4	3	2	5	7	1	6	9
D. Flynn	MALE	PEPSI	9	9	2	6	6	8	7	5	2
J. Owen	FEMALE	PEPSI	7	8	2	3	6	9	1	3	2
H. Morrow	MALE	PEPSI	6	6	2	8	3	6	4	2	3
F. East	FEMALE	PEPSI	4	6	6	5	6	8	7	2	5
C. Clint	MALE	COKE	7	3	3	7	0	6	5	0	4
I. Neil	MALE	PEPSI	6	2	3	1	8	1	4	2	6
G. Boss	MALE	COKE	7	2	4	8	1	2	6	8	9
K. Record	FEMALE	PEPSI	6	2	7	5	7	2	5	0	6
T. Bush	FEMALE	PEPSI	3	2	5	4	4	4	3	7	6
P. Squire	MALE	COKE	2	9	9	3	1	4	5	5	4
L. Mynard	FEMALE	PEPSI	1	0	7	5	2	4	2	6	9

Стандартизация значений

Переменные: нет

Наблюдения: ВСЕ

Вес: ВЫКЛ

Значения выбранных переменных будут стандартизованы (в пределах их диапазона).

OK Отмена

Создать подмножество

Переменные: нет

Наблюдения: ВСЕ

OK Отмена

Переменная 1

Имя: GENDER Код ПД: -9999

Формат отображения

Ширина столб.: 6 Дес. разр.: 0

Тип:	Представление:
Число	1,000; -1,000
Дата	1,000; -1,000
Время	1,000; (1,000)
Научный	1,000; (1,000)
Денежный	
Проценты	

Длинное имя (метка, связь или формула с Функциями):

Gender of the subjects (May 15, 1996).

Примеры: Метка: Валовой доход в 1991 Формулы: = v1 + v2;
Связь: @Excelc:\file.xls!2c2:14c4 = (v1>0)*AGE + v3

OK Отмена

Фильтрация

Условия выбора наблюдений

Изменить/добавить условия выбора:

Включ., если:

Искл., если:

Заголовок (для файла условий выбора):

Переменные

Состояние

ВКЛ Операторы: = <> < > <= >= NOT AND OR
Переменные: имена или v1, v2...

ВЫКЛ Номер наблюдения: v0
Примеры: v1 = 0 OR v2 >= 0 (v7<1 OR v9='YES') and v4<>0

Открыть Сохранить...

OK Отмена

Текстовые значения

Диспетчер текстовых значений - GENDER

Текст	Число	Метка значения
MALE	1	
FEMALE	2	

OK Отмена

Удалить строку

Очистить все

Сортировать

Текст - возр.

Текст - убыв.

Числа - возр.

Числа - убыв.

Перенумеровать

С поз. курсора

Все

Копировать из

Переменной...

Имен набл.

Автозаполнение

Автопроверка

Чтобы создать новое "текст-числовое" присвоение, введите текст и числ. значения в первые два поля строки, например MALE и 1 (метка значения необязательна).

Если последняя строка будет заполнена, новая строка добавится автоматически.

Вес

Задание веса

Вес из переменной:

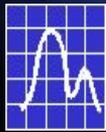
GENDER

Состояние

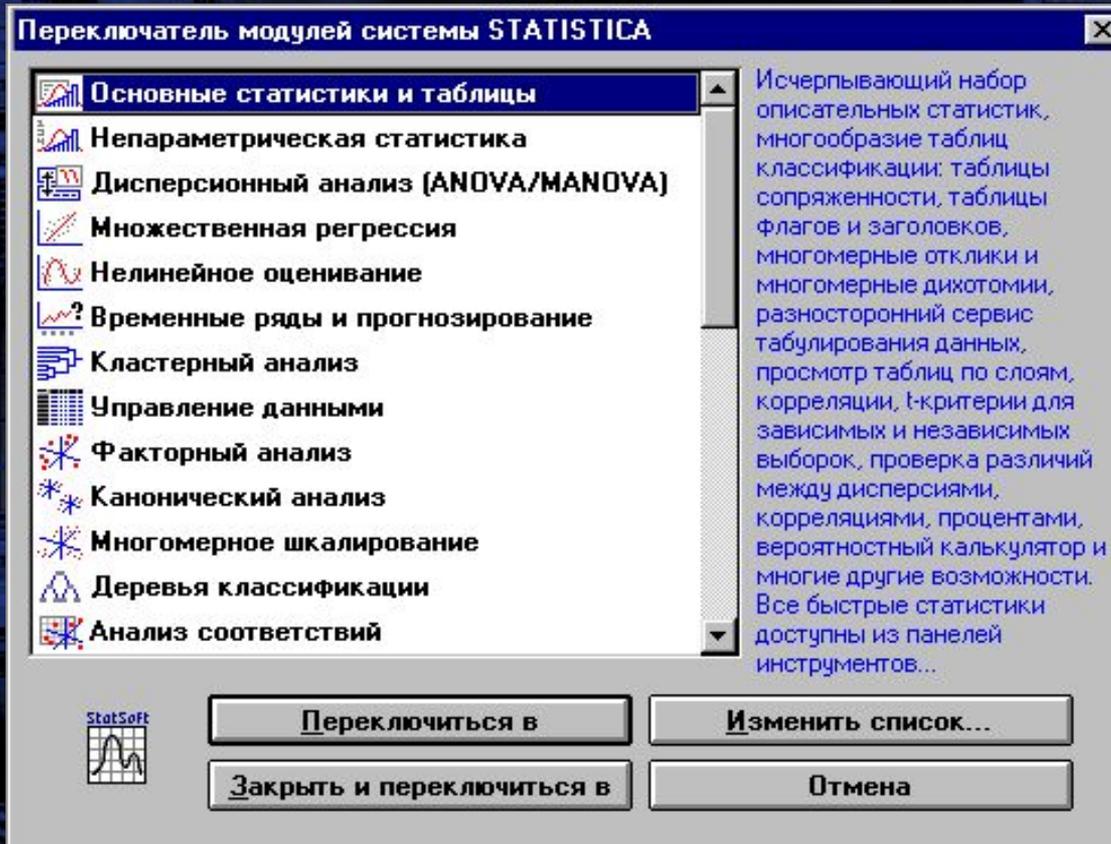
ВКЛ Дважды щелкните на поле ввода, чтобы выбрать переменную из списка.

ВЫКЛ

OK Отмена

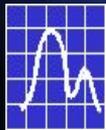


Анализ данных



STATISTICA
включает все
методы
статистического
анализа данных:
от классических
до самых
современных

Для удобства
анализа методы
разделены на
модули



Анализ данных

Переключатель модулей системы STATISTICA

- Дисперсионный анализ (ANOVA/MANOVA)
- Множественная регрессия
- Нелинейное оценивание
- Временные ряды и прогнозирование
- Кластерный анализ
- Управление данными
- Факторный анализ**
- Канонический анализ
- Многомерное шкалирование
- Деревья классификации
- Анализ соответствий
- Структурное моделирование
- Надежность и позиционный анализ
- Дискриминантный анализ
- Логлинейный анализ
- Анализ выживаемости

Всесторонний набор методов факторного анализа и анализа главных компонент; общности, задаваемые пользователем или определяемые регрессией, метод максимального правдоподобия, ортогональные и косоугольные факторы, иерархический анализ косоугольных факторов, вращения, критерий и график каменистой осыпи...
Расширенный факторный анализ реализован в модуле 'Структурное моделирование', см. также 'Наименьшие частные квадраты'.

StatSoft

Переключиться в Изменить список...

Закреть и переключиться в Отмена

Факторный
анализ



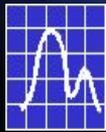
Анализ данных

Просто

Естественно

Эффективно

Полностью на русском языке!



Интерактивный анализ

Весь анализ в системе проводится с

ИСПОЛЬЗОВАНИЕМ НАГЛЯДНЫХ ДИАЛОГОВЫХ ОКОН,

ценариям обработки данных.

Переключатель модулей системы STATISTICA

- Основные статистики и таблицы
- Непараметрическая статистика
- Дисперсионный анализ (ANOVA/MANOVA)
- Множественная регрессия
- Нелинейное оценивание
- Временные ряды и прогнозирование
- Кластерный анализ
- Управление данными
- Факторный анализ
- Канонический анализ
- Многомерное шкалирование
- Деревья классификации
- Анализ соответствий
- Структурное моделирование
- Основные статистики и таблицы

Исчерпывающий набор описательных статистик, многообразие таблиц классификации, таблицы сопряженности, таблицы флагов и заголовков, многомерные отклики и многомерные дихотомии, разносторонний сервис табулирования данных, просмотр таблиц по слоям, корреляции, t-критерии для зависимых и независимых выборок, проверка различий между дисперсиями, корреляциями, процентами, вероятностный калькулятор и многие другие возможности. Все быстрые статистики доступны из панелей инструментов...

Внутригрупповые описательные статистики и корреляции - Результаты

ЗАВИСИМЫЕ: 1 перемен.: AGGRESSN

ГРУППИР.: 1-GENDER (2): BOYS GIRLS

Итоговая таблица средних

Статистики

- Число наблюдений
- Суммы
- Стандартные отклонения

OK Отмена

Внутригрупповые описательные статистики и корреляции

Анализ: **Подробный анализ выбранных таблиц**

Переменные

Группирующие: **GENDER**
Зависимые: **AGGRESSN**

Коды для группирующих переменных: **нет**

Построчное удаление ПД

Итоговая таблица средних (aggressn.sta)

Далее... N=24 (Нет пропусков в завис. перем.)

GENDER	AGGRESSN
BOYS	80.75000
GIRLS	26.66667
Всего	53.70833

Мартин Дисперсион Брауна-Форсайта (ОД)

Внутригрупповые корреляции Опции

Переупорядочить факторы в таблице

Графики взаимодействий

Категоризованные диаграммы рассеяния

Графики средних и станд. отклонений



Интерактивный анализ

Весь анализ в системе проводится с использованием наглядных диалоговых окон, следующих типовым сценариям обработки данных.

Множественная регрессия

Переменные:

Независимые: ВОЗРАСТ-МЕЧ_ОТР
Зависимые: БИОХИМИЯ

Файл ввода: Исходные данные

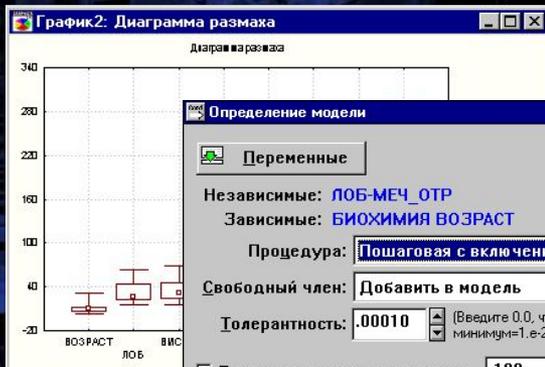
Удаление ПД: Построчное

Тип регрессии: Стандартная

Провести анализ по умолчанию (не пошаговый)
 Показывать описательные статистики, корр. матрицы
 Вычисления с повышенной точностью
 Пакетная обработка/печать
 Печатать результаты анализа остатков

Задать все переменные для анализа; дополнительные модели (зависимые/независимые переменные) могут быть заданы позже.

OK Отмена Данные



Определение модели

Переменные

Независимые: ЛОБ-МЕЧ_ОТР
Зависимые: БИОХИМИЯ ВОЗРАСТ

Процедура: Пошаговая с включением

Свободный член: Добавить в модель

Толерантность: .00010 (Введите 0.0, чтобы минимум=1.e-25)

Гребневая регрессия; лямбда: .100

Пошаговая множественная регрессия:

F-включить: 1.00 F-исключить: 0.00

Число шагов: 10

Отображение результатов: Только итоги

Пакетная обработка/печать
 Печатать результаты анализа остатков

Промотреть описательные статистики



Просмотр описательных статистик

Пропущ. данные построчно удалены

49 набл. обработано
28 набл. принято

Средние и std.откл. Диаграмма размаха Std.откл.=Сумма квадратов/М

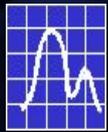
Корреляции Матричный график Сохранить корреляции Отмена

Ковариации

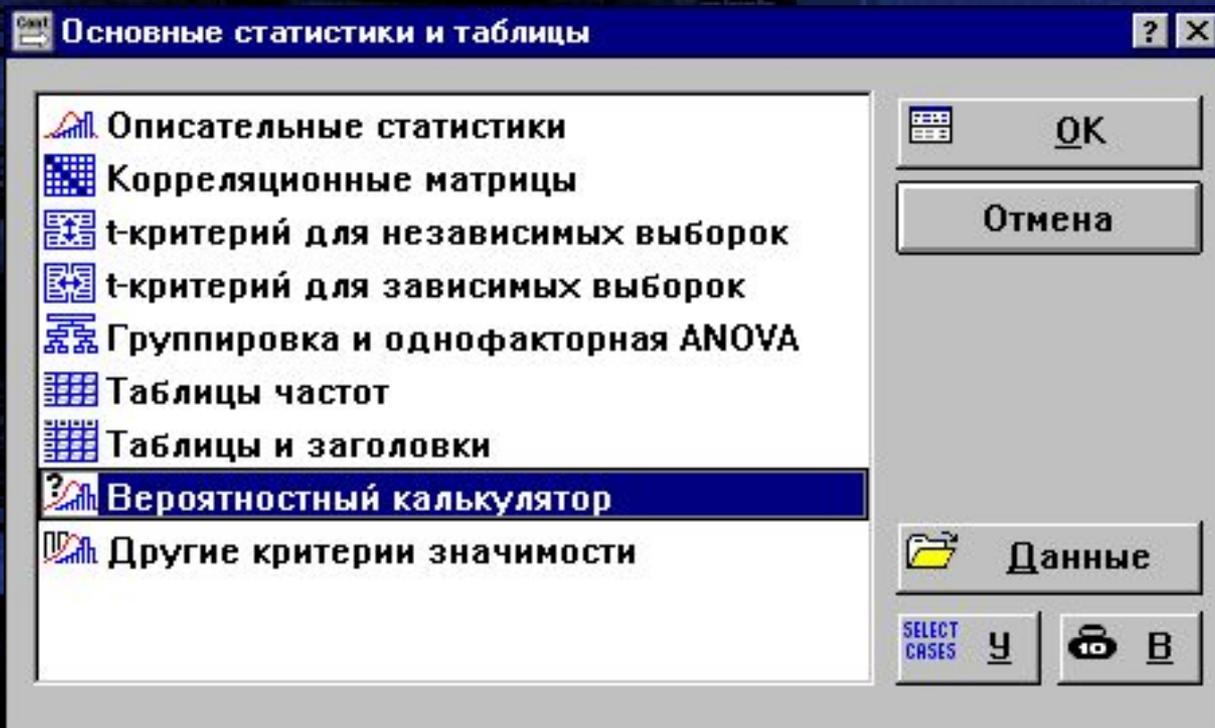
Итоги гребневой регрессии для зав. переменной: БИОХИМИЯ

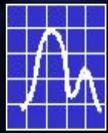
МНОЖЕСТВ РЕГРЕС. 1= 10000 R= .95076477 R2= .90395365 Скооррекц. R2= .88724993
F(4,23)=54.117 p<.000000 Станд. ошибка оценки : 30.007

N=28	БЕТА	Ст. ов. БЕТА	B	Ст. ов. B	t(23)	p-уров.
Св. член			-85.7534	16.52745	-5.18854	.000029
ЛАДОНЫ	.417725	.117398	3.5193	.98905	3.55820	.001673
ЛОБ	.198922	.148716	1.1467	.85726	1.33760	.194103
ПЛЕЧО	.176975	.131506	1.0459	.77717	1.34575	.191499
ВИСОК	.175132	.153821	.9809	.86152	1.13855	.266613



Основные статистики и таблицы





Вероятностный калькулятор

Калькулятор вероятностных распределений [?] [X]

Распределение

- Бета
- Коши
- Хи-квадрат
- Экспоненциальное
- Экстрем. значения
- F
- Гамма
- Лапласа
- Логнормальное
- Логистическое
- Парето
- Релея
- t (Стьюдента)**
- Вейбулла
- Z (нормальное)

Фиксир. масштаб

Обратная ф.р. Печать

Двусторонняя График

[1 - ф.р.]

Вычислить

Выход

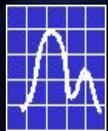
t:

Ст. св.:

p:

Функция плотности:

Функция распределения (ф.р.):



Вероятностный калькулятор

Калькулятор вероятностных распределений

Распределение

- Бета
- Коши
- Chi-квадрат**
- Экспоненциальное
- Экстрем. значения
- F
- Гамма
- Лапласа
- Логнормальное
- Логистическое
- Парето
- Релея
- t (Стьюдента)
- Вейбулла
- Z (нормальное)

Фиксир. масштаб

Обратная ф.р. Печать

Двусторонняя График

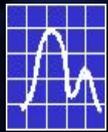
[1 - ф.р.]

Chi-квад.: Ст. св.:

p:

Функция плотности:

Функция распределения (ф.р.):



Вероятностный калькулятор

Калькулятор вероятностных распределений

Распределение

- Бета
- Коши
- Хи-квадрат
- Экспоненциальное
- Экстрем. значения
- F
- Гамма**
- Лапласа
- Логнормальное
- Логистическое
- Парето
- Релея
- t (Стьюдента)
- Вейбулла
- Z (нормальное)

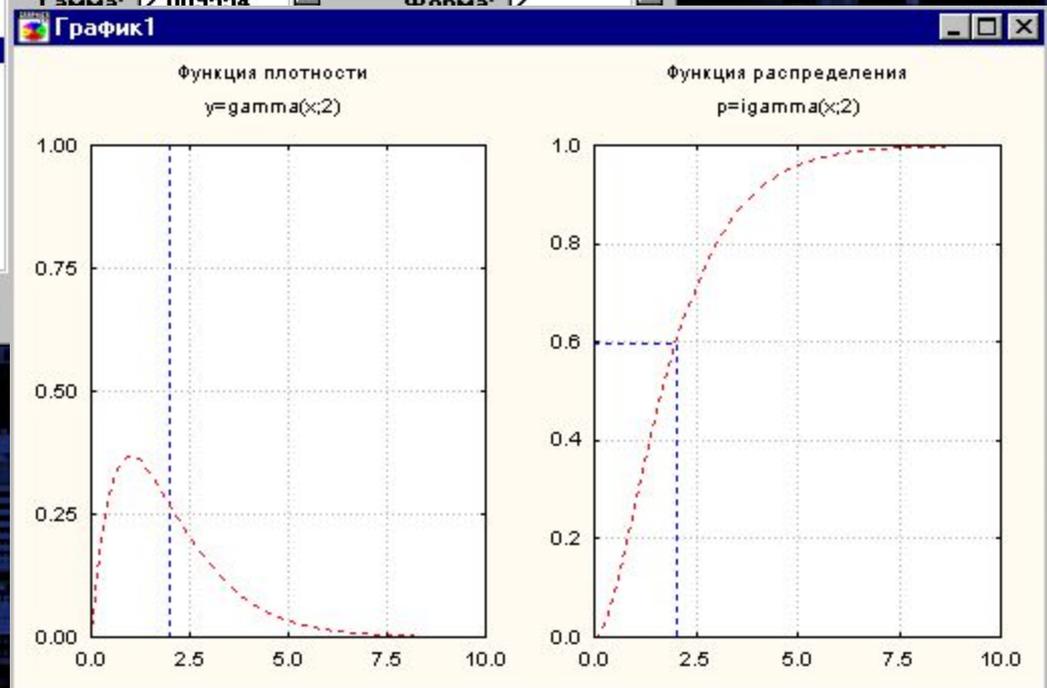
Фиксир. масштаб

Обратная ф.р. Печать График

Двусторонняя [1 - ф.р.]

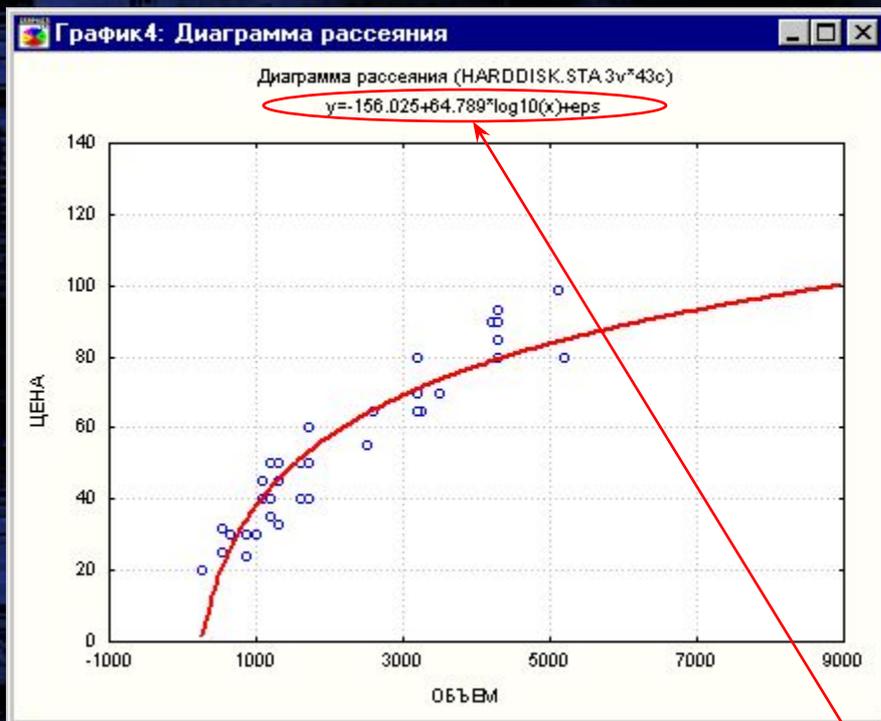
Гамма: 2.009554 Форма: 2

Вычислить **Выход**





Интерактивный анализ



Кисть

Результат изменился

Закрашивание

Обновить

Выход

Отменить все

Автообновление

Зависимость 1

Операция

- Маркировать
- Пометить
- Выключить
- Отменить выбор

Обратить

Кисть

- Точка
- Прямоугольник
- Лассо

Движение

Дополнительно...



Интерактивный анализ



Кисть

Точки автоматически помечены

Закрашивание

Обновить

Выход

Отменить все

Автообновление

Зависимость 1

Операция

- Маркировать
- Пометить
- Выключить
- Отменить выбор

Обратить

Кисть

- Точка
- Прямоугольник
- Лассо

Движение

Дополнительно...



Программы-мастера

Размещение графиков

Мастер автоматического р

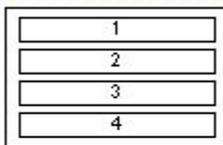
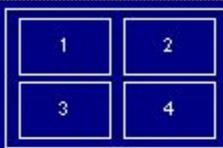
Графики (файлы или окна) для

- 1: График4: Гистограмма
- 2: График3: Полунорм. ве
- 3: График2: Автокорреля
- 4: График1: Граф. переме

Связанные

Мастер автоматического

Автоматическое разме

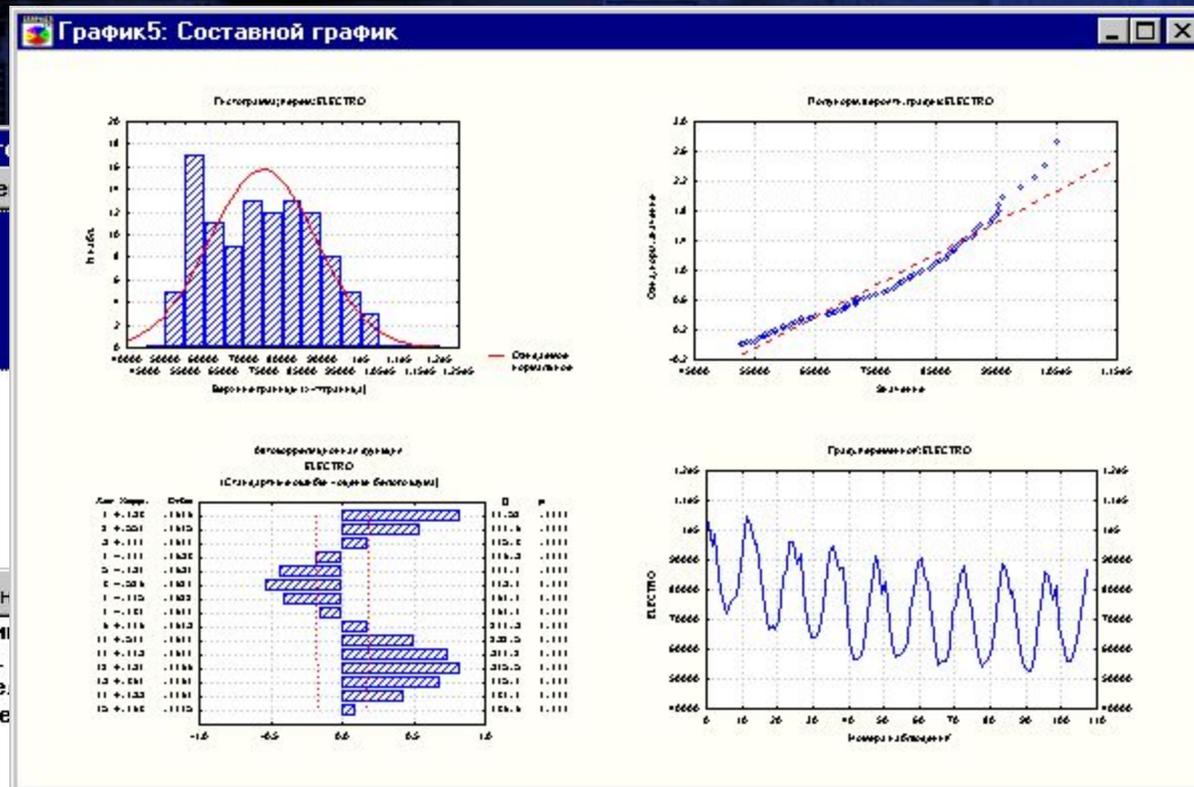


Графики: Чтобы измен

- 1: График4: Гистограм
- 2: График3: Полунорм.
- 3: График2: Автокорре.
- 4: График1: Граф. пере

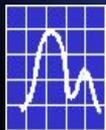
Связанные

Внедренные



Изотропное

Изменить список



Программы-мастера

SQL-запросы
к базам данных

Select Data Source [?] [X]

File Data Source Machine Data Source

Data Source Name	Type	D
MQIS	System	M
MS Access 97 Database	User	
MS Access Database	User	
SALES	System	
SENS	System	
SEWSS	System	

Visu Login

Data Source: SALES

Authorization

Login name:

Password:

Импорт через ODBC - Выбор полей и условий выбора строк [?] [X]

Таблицы:

- ORDERS**
- PARTS
- DISTINCT 1
- SHOW
- SHOW_READY

Поля:

- ID**
- ORDER_ID**
- PART_NUM**
- QUANTITY**
- DATETIME**

Поля для импорта:

- ORDERS.ORDER_ID
- ORDERS.PART_NUM
- ORDERS.QUANTITY
- ORDERS.DATETIME

OK

Отмена

Изменить запрос SQL

Выбор строк

Выбор всех строк

КУДА... (условия выб

Данные: NEW.STA 4п * 5н

	1	2	3	4
	ORDER_ID	PART_NUM	QUANTITY	DATETIME
1	1	1	2	2/18/00 12:00:00 AM
2	1	5	2	2/18/00 12:00:00 AM
3	1	8	4	2/18/00 12:00:00 AM
4	1	9	6	2/18/00 12:00:00 AM
5	1	5	3	2/18/00 12:00:00 AM

цы, тени

бы для списками



Программы-мастера

Программирование на STATISTICA BASIC

Язык STATISTICA BASIC: Маст

STATISTICA: Электронное руководство

File Edit Bookmark Help

Contents Search Back History Печать Словарь По

Графическая библиотека языка STATISTICA
GraphCopyToClipboard (Graph)

Параметр:

Graph дескриптор графика

Функция *GraphCopyToClipboard* копирует указанный график в

Пример:

```
RandomAccess;  
graphname$ := 'c:\working.stg';  
a_graph := OpenGraph(graphname$);  
GraphCopyToClipboard(a_graph);  
DeleteGraph(a_graph);
```

Эта программа открывает предварительно сохраненный график, копирует его в буфер обмена и затем закрывает.

См. также разделы: [PrintGraph](#), [GetGraph](#), [SaveGraph](#), [OpenGraph](#), [DeleteGraph](#) и [NewGraph](#).

Полный список графических функций см. в разделе [Графическая библиотека языка STATISTICA BASIC](#).

```
Язык STATISTICA BASIC: D:\#USERS\cbr\NEW.STB  
0 1 2 3 4  
RandomAccess;  
  
graphname$ := 'c:\working.stg';  
a_graph := OpenGraph(graphname$);  
GraphCopyToClipboard(a_graph);  
  
DeleteGraph(a_graph);
```

Новая строка



(GetGraph (0));

Чтобы начать новую строку, нажмите Ctrl+Enter.



Автоматизация

Данные: autofill.STA 10п * 10н

	1 VAR1	2 VAR2	3 VAR3	4 VAR4	5 VAR5	6 VAR6	7 VAR7	8 VAR8	9 VAR9	10 VAR10
1	1	1.50	Yes	6-Сен-02	1.15					
2	2	1.55	Yes	10-Сен-02	2.10					
3	3	1.60	Yes	14-Сен-02	.14					
4	4	1.65	Yes	18-Сен-02	3.00					
5	5	1.70	Yes	22-Сен-02	3.70					
6										
7										
8										
9										
10										

Данные: autofill.STA 10п * 10н

	1 VAR1	2 VAR2	3 VAR3	4 VAR4	5 VAR5	6 VAR6	7 VAR7	8 VAR8	9 VAR9	10 VAR10
1	1	1.50	Yes	6-Сен-02	1.15					
2	2	1.55	Yes	10-Сен-02	2.10					
3	3	1.60	Yes	14-Сен-02	.14					
4	4	1.65	Yes	18-Сен-02	3.00					
5	5	1.70	Yes	22-Сен-02	3.70					
6	6	1.75	Yes	26-Сен-02	3.82					
7	7	1.80	Yes	30-Сен-02	4.42					
8	8	1.85	Yes	4-ОКТ-02	5.02					
9	9	1.90	Yes	8-ОКТ-02	5.62					
10	10	1.95	Yes	12-ОКТ-02	6.22					

Статистики

- Медиана и квантили
- Доверит. границы средних

Интервал: %

Другие статистики

0.77



Автоматизация

Поиск лучшей модели

Автоматический поиск лучших параметров [? X]

Максимальное число итераций: [OK]

Условия сходимости: [Отмена]

Безусловное оценивание параметров

Индикаторы согласия

Среднеквадратическая ошибка
 Средняя абсолютная ошибка
 Средняя абсол. относ. ошибка

Программа минимизирует выбранный индикатор согласия (т.е., найдет соответствующие лучшие параметры), начав с начальных значений параметров.

Начальные значения параметров

Альфа: Дельта:
Гамма: Фи:

Определение модели [?] [X]

Переменные [OK]

Независимые: **БИО_ПРЯМ-ВОЗРАСТ**
Зависимые: **БИОХИМИЯ**

Процедура:

Свободный член:

Толерантность: (Введите 0.0, чтобы задать минимум=1.e-25)

Гребневая регрессия; лямбда:

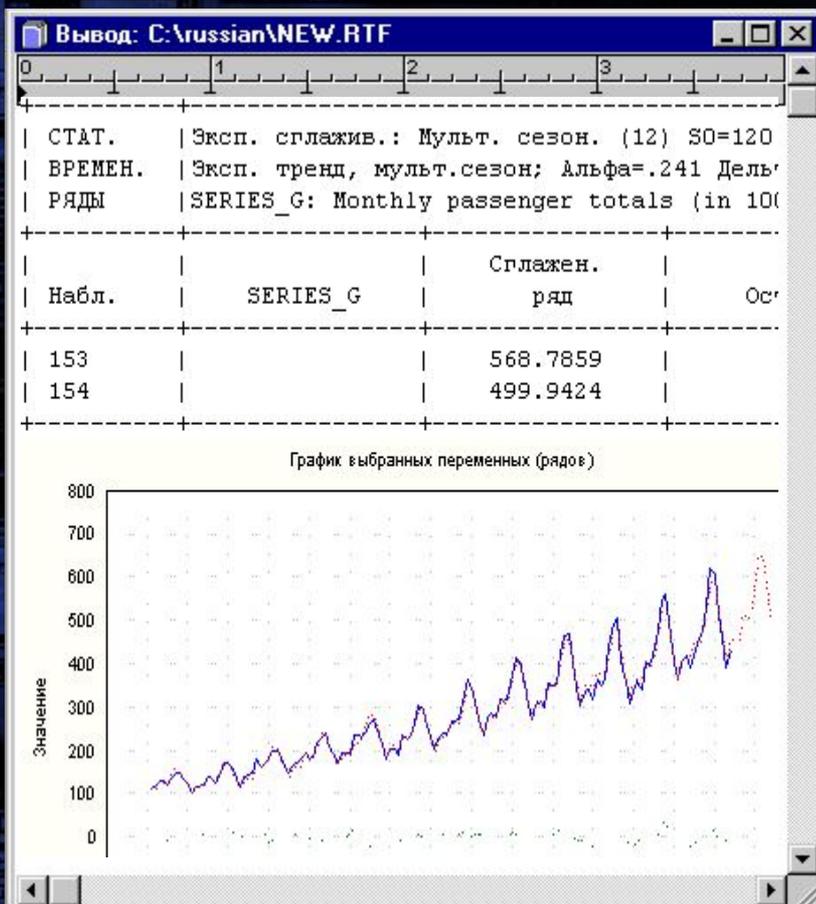
Пошаговая множественная регрессия:

F-включить: F-исключить:
Число шагов:
Отображение результатов:

Пакетная обработка/печать
 Печатать результаты анализа остатков



Автоотчет



Параметры страницы/вывода

Текст/таблицы результатов и данных Графики

Вывод **Тип отчета** **Страница**

Окно **Линии**

Заголовок вывода

Дата и время Разместить заголовки по центру

Автоотчет

Авт. дополнение содержания окна текста/вывода

Авт. выход из таблиц результатов и графиков

Авт. печать всех таблиц результатов (автоотчет)

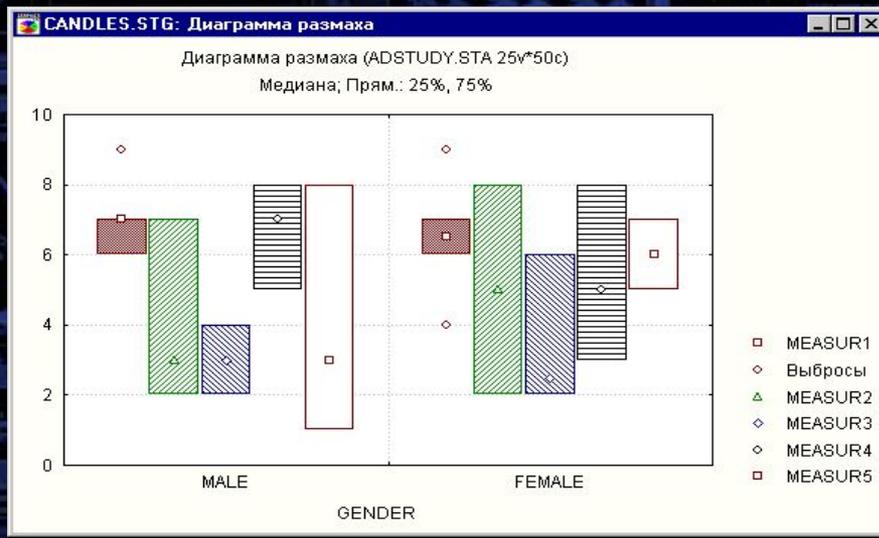
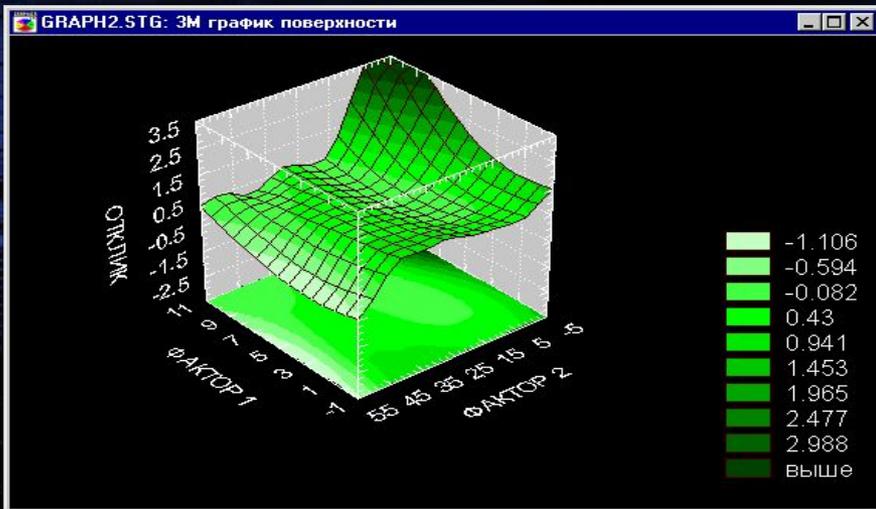
Авт. печать/выдача страниц при каждом выводе

Авт. выдача каждой заполненной страницы

Печать таблицы в окне текста/вывода

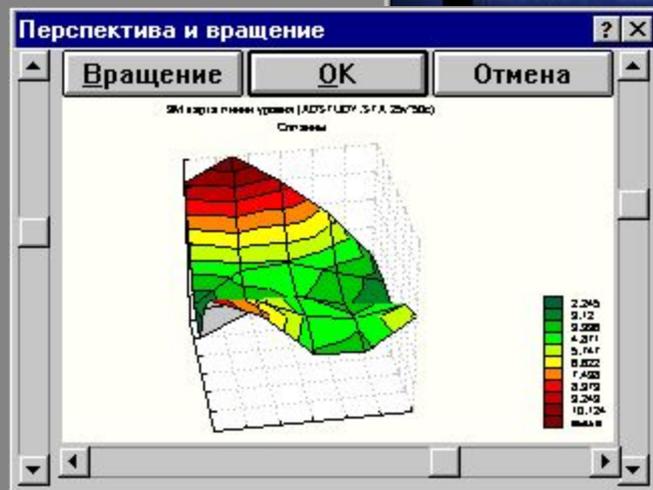
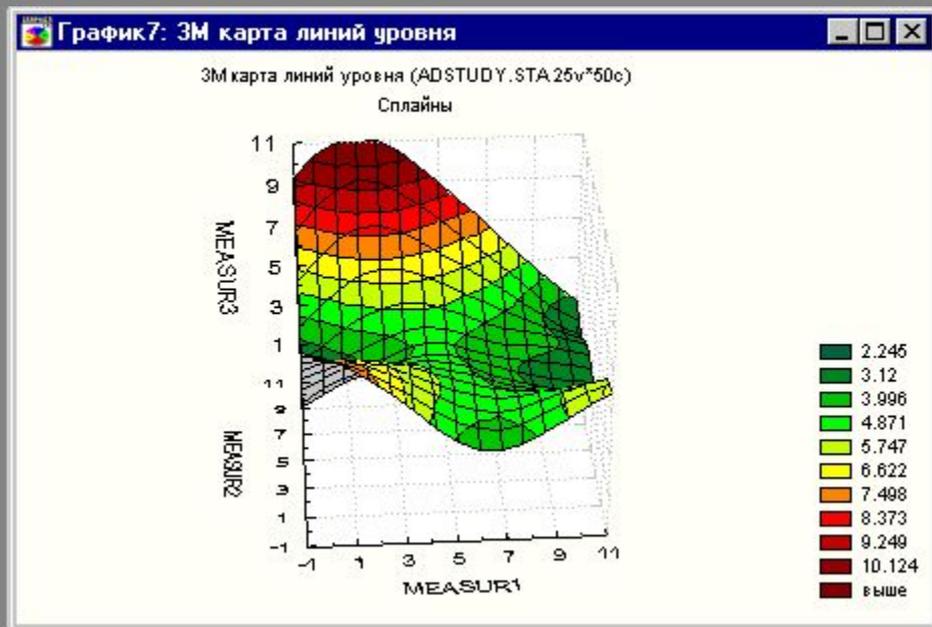
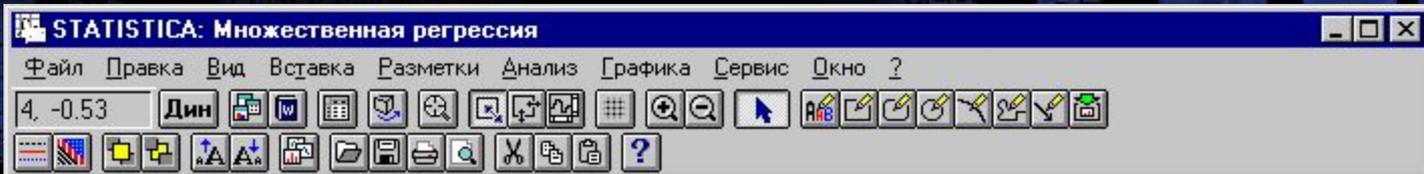


Графики





Работа с графикой



Готово

Вывод: ВЫКЛ

Выбор: ВЫКЛ

Вес: ВЫКЛ



Работа с графикой

STATISTICA: Множественная регрессия

Файл Правка Вид Вставка Разметки Анализ Графика Сервис Окно ?

4, -0.53

Общая разметка: 3М диаграммы рассеяния и поверхности

График7: 3М карта линий у...

3М карта линий ур...

Тип графика:
- Диаграмма отклонений
- Трассировочный график
- **График поверхности**

Категоризация

ЗАГОЛОВКИ

Заголовок 1: 3М карта линий у...
3М карта линий ур...

ОСИ

X: Макс.: 11 Разметка оси: Авто
Y: Шаг: 2 Тип: Линейная
Z: Мин.: 1

Объемный текст

ПОДГОНКА (ПОВЕРХНОСТИ И КОНТУРЫ)

Сглаживание сплайнами

3М графики: дополнительные свойства

УГОЛ ЗРЕНИЯ
X: 1.036745
Y: -15.58605
Z: 10

ПРОПОРЦИИ ОСЕЙ
X: 1
Y: 1
Z: 1

СТИЛЬ ПОВЕРХНОСТИ
Верхняя: Затененная Сплошная
Сверху: Снизу:

УРОВНИ ПОВЕРХНОСТИ/КОНТУРА
 Авто Другие
Число уровней: 10
Уровень отклонения: 0
Число спектральных плоскостей: 5

ЛИНИИ ПОВЕРХНОСТИ
 Показать скрытое
X-линия Y-линия

КОНТУР ПОВЕРХНОСТИ
 Отобразить
Расстояние от мин. Z: 20 % от Z

ПРАВКА ЗАГЛОВОКОВ

Заголовок 1: 3М карта линий ур...

Заголовок 2: @F{Слайны@}

Заголовок 3:

ЛИНИИ СЕТКИ

XY XZ YZ

Нет Нет Нет

УРОВНИ ПОВЕРХНОСТИ И КОНТУРА

Стиль линий уровня:
 Линии Зоны

Уровень	Контур. линия	Зонная линия	Зона
1: 2.44723			
2: 3.12025			
3: 3.99578			
4: 4.87130			
5: 5.74683			
6: 6.62236			
7: 7.49789			
8: 8.37341			
9: 9.24894			
10: 10.1244			

Зона над верхним уровнем:

Нарисовать линии

ШРИФТ

Шрифт: Arial

Начертание: Обычный

Зачеркивание
 Подчеркивание

Цвет: Black

Применять в качестве общего шрифта Использовать общий шрифт

До настоящего момента использовался общий шрифт

Вывод: ВЫКЛ



Гибкий интерфейс

Описательные статистики (adstudy)

ОСНОВНЫЕ СТАТИСТ.	N набл.	Среднее	доверит. -95.000%	доверит. +95.000%	Медиана	Сумма	Минимум
MEASUR1	50	5.900000	5.227345	6.572655	6.000000	295.0000	0.00
MEASUR2	50	4.540000	3.719507	5.360493	5.000000	227.0000	0.00
MEASUR3	50	4.140000	3.365389	4.914611	3.500000	207.0000	0.00
MEASUR4	50	4.840000	4.764281	6.275719	6.000000	276.0000	0.00
MEASUR5	50	4.764281	3.211469	4.708531	4.500000	198.0000	0.00
MEASUR6	50	4.764281	3.981898	5.698102	4.500000	242.0000	0.00
MEASUR7	50	4.660000	3.950704	5.369296	5.000000	233.0000	0.00
MEASUR8	50	3.720000	2.922263	4.517737	3.000000	186.0000	0.00
MEASUR9	50	4.160000	3.294248	5.025752	4.000000	208.0000	0.00
MEASUR10	50	3.940000	3.072252	4.807748	4.000000	197.0000	0.00
MEASUR11	50	5.840000	4.310000	5.860000	6.000000	292.0000	0.00
MEASUR12	50	4.840000	3.760000	5.920000	5.000000	242.0000	0.00
MEASUR13	50	4.840000	3.760000	5.920000	5.000000	242.0000	0.00

Кнопки автозадач

- Сезонная Корректировка
- Долгосрочный Прогноз
- Среднесрочный Прогноз
- Краткосрочный Прогноз

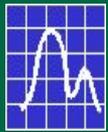
Кнопки автозадач

- РЕГИОНАЛЬНАЯ БЕЗРАБОТИЦА
- БЕЗРАБОТИЦА
- Региональная безработица
- АНАЛИЗ БЕЗРАБОТИЦЫ
- РАСЧЕТ БЮДЖЕТА
- БЮДЖЕТ1
- БЮДЖЕТ2
- РОЗНИЧНЫЙ ТОВАРОБОРОТ
- СЕЗОННОСТЬ РТО
- БЮДЖЕТ1997
- БЮДЖЕТ1996
- БЮДЖЕТ1995
- ЖИЛИЩНО-КОММУНАЛЬНОЕ ХОЗ-ВО
- ПРАВООХРАНИТЕЛЬНЫЕ ОРГАНЫ
- СЕЛЬСКОЕ ХОЗЯЙСТВО
- ОБРАЗОВАНИЕ
- ЗДРАВООХРАНЕНИЕ
- КУЛЬТУРА
- ДРУГИЕ ОТРАСЛИ
- Настройка

Кнопки автозадач

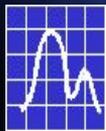
- Контрольная группа
- Ввод параметров
- Анализ выживаемости
- Лаборатория
- Отчет за месяц
- Графики
- Факторы риска

Basic Statistics



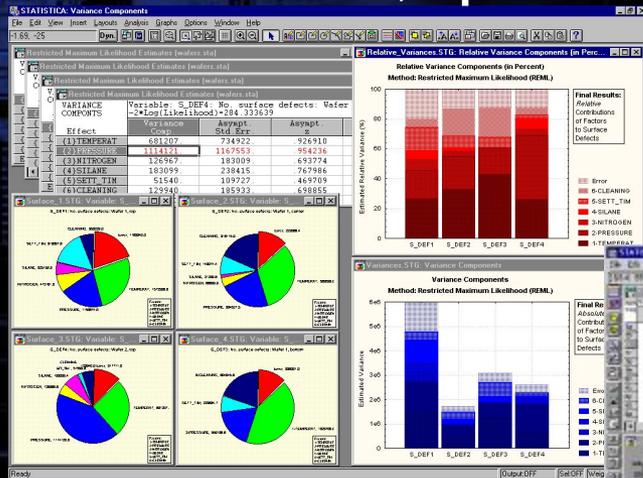
7 ДОСТОИНСТВ СИСТЕМЫ

- Знакомый Windows-интерфейс
- Русский язык
- Полный набор статистических методов
- Сотни типов графиков
- Облегчающие работу программы-мастера
- Объекты и соответствующие им операции
- Настройка аналитической среды в соответствии с пожеланиями пользователей

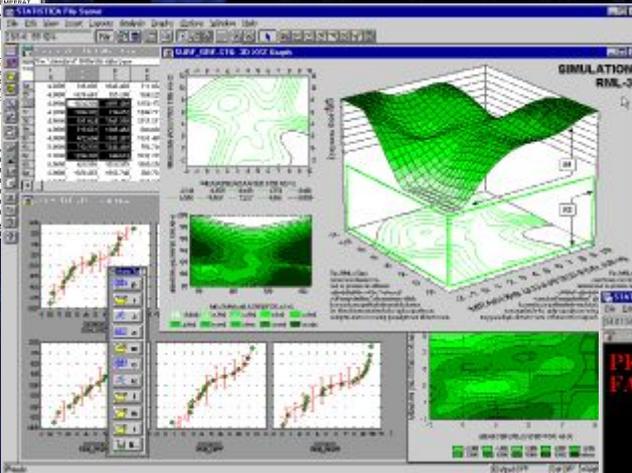


Области применения

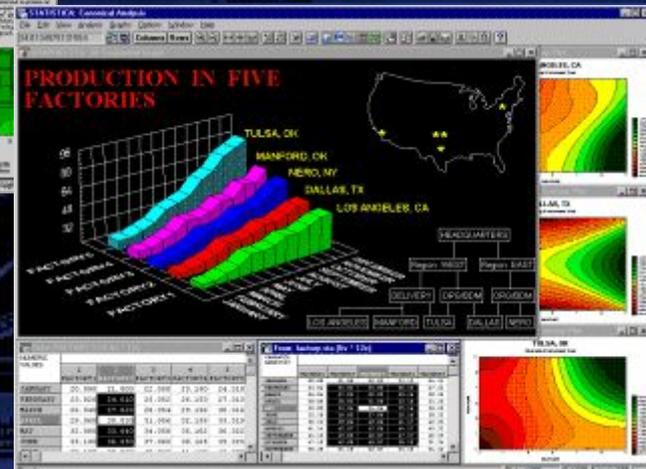
Экономика, финансы



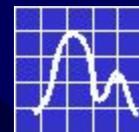
Научные исследования



Производство



**Пример:
Оценка эффективности
модернизации жилых зданий**



StatSoft® Russia

Описание объекта и цель исследования

Анализируемый объект представляет собой жилое здание, подвергнутое модернизации с целью экономии ресурсов.

Описание объекта и цель исследования

Целью исследования является оценка экономии ресурсов, полученная за счет модернизации.

Основные параметры объекта

- Q_SUM – суммарное потребление тепла
- QEL – потребление электроэнергии
- T_AP – температура в квартирах
- T_ENV – температура окружающей среды
- TSW – температура сетевой воды

Этапы исследования

Первый этап

- импорт данных из Excel в STATISTICA
- верификация данных
- удаление выбросов (чистка данных)
- агрегирование данных

Этапы исследования

Второй этап

- **Разведочный анализ данных**
- **Построение зависимостей**
- **Разработка методики расчета экономии ресурсов**

Этапы исследования

Третий этап

- **Оценка точности вычисления средней температуры дома по выбранным квартирам (оценка репрезентативности выборки)**

Использование модулей системы STATISTICA

Основные статистики и таблицы
(дескриптивный анализ, группировка)

- *Нелинейное оценивание* (нахождение точек изменения зависимостей)
- *Множественная регрессия* (нахождение зависимостей)

- *Анализ временных рядов* (нахождения сезонных составляющих)
- *Кластерный анализ* (нахождения групп квартир, близких по температуре)
- *Планирование Эксперимента*
(оптимальное размещение датчиков в доме, позволяющее максимально точно оценить среднюю температуру)

Предварительный этап анализа

Excel

Буфер обмена

STATISTICA

Данные: НС_41_F_STA 42п * 14148п

ТЕКСТОВЫЕ ЗНАЧЕНИЯ	1 ДАТА_ВР	2 ДАТА	3 ДЕНЬ	4 МЕСЯЦ	5 ГОД	6 ВРЕМЯ	7 QEL
1	98/8/14 10:00:00	14Авг-98	14	Aug	1998	10.00	9.140
2	98/8/14 11:00:00	14Авг-98	14	Aug	1998	11.00	8.170
3	98/8/14 12:00:00	14Авг-98	14	Aug	1998	12.00	8.730
4	98/8/14 13:00:00	14Авг-98	14	Aug	1998	13.00	7.590
5	98/8/14 14:00:00	14Авг-98	14	Aug	1998	14.00	6.836
6	98/8/14 15:00:00	14Авг-98	14	Aug	1998	15.00	7.890
7	98/8/14 16:00:00	14Авг-98	14	Aug	1998	16.00	7.990
8	98/8/14 17:00:00	14Авг-98	14	Aug	1998	17.00	7.570
9	98/8/14 18:00:00	14Авг-98	14	Aug	1998	18.00	10.560
10	98/8/14 19:00:00	14Авг-98	14	Aug	1998	19.00	9.880
11	98/8/14 20:00:00	14Авг-98	14	Aug	1998	20.00	13.180
12	98/8/14 21:00:00	14Авг-98	14	Aug	1998	21.00	15.550
13	98/8/14 22:00:00	14Авг-98	14	Aug	1998	22.00	15.000

Организация данных

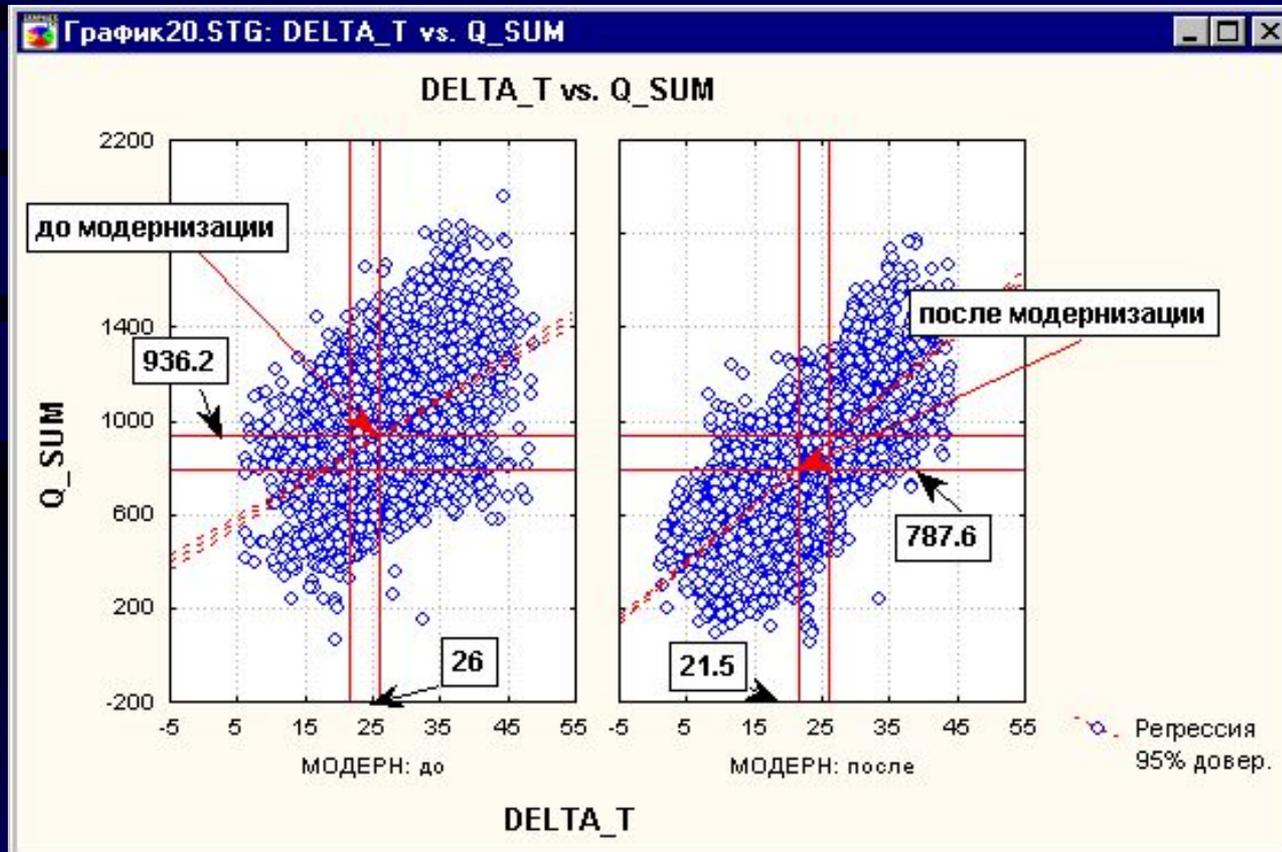
Верификация данных

Агрегирование
данных

- **На предварительном этапе найдены:**
 - *ложные* наблюдения
 - *неточность* временной шкалы (в данных присутствовали значения времени 0, 1, 2 ... 24 часов для некоторых суток)
 - *несоответствие* точек отсчета (до 22.04.1999 17:00 отсчет велся по Гринвичу, после этой даты - по местному времени)

Результаты анализа

Оценка значимости эффекта модернизации по исходным данным



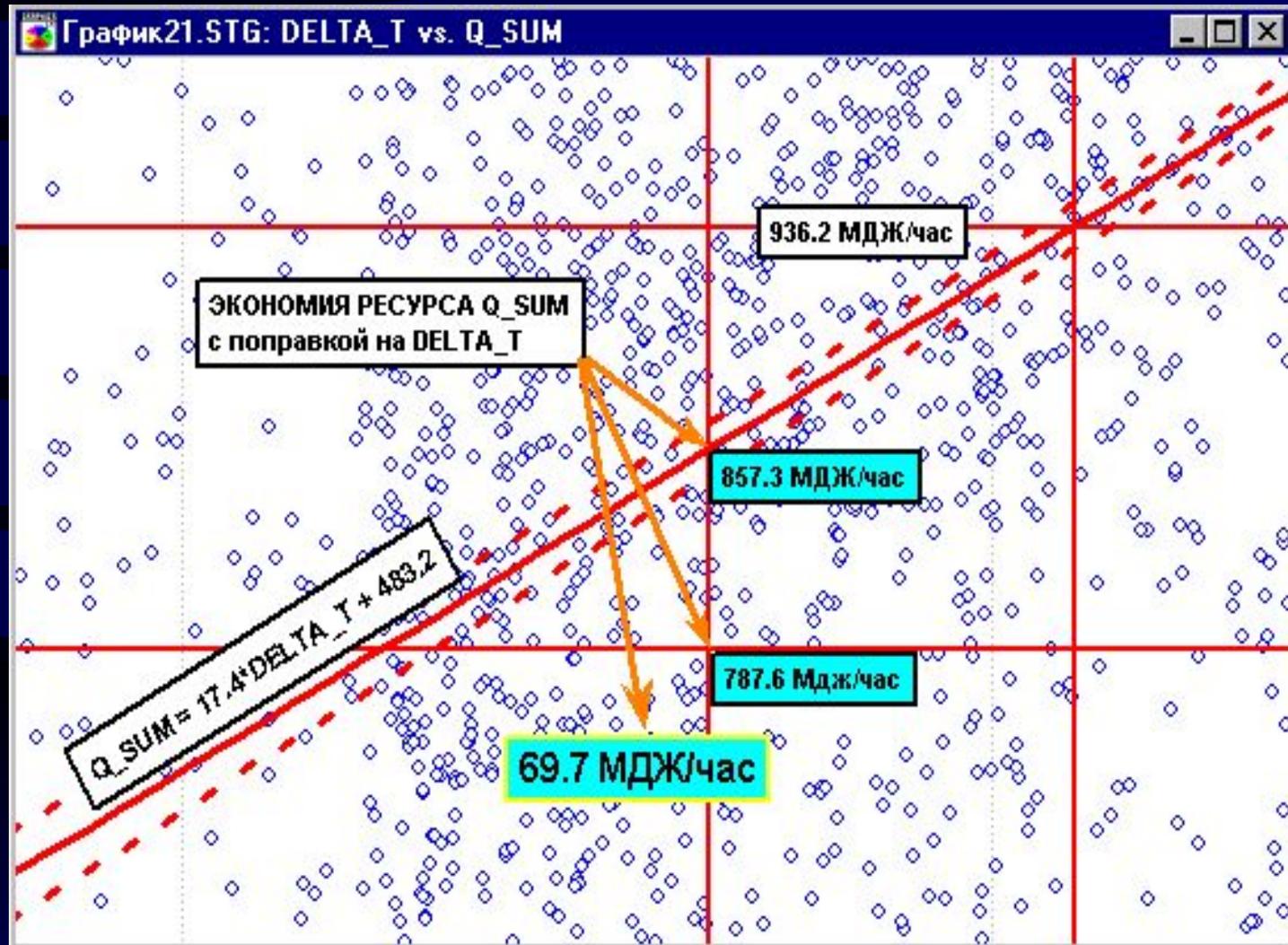
- Q_SUM – суммарное потребление тепла
- $DELTA_T$ разность: T_AP (температура квартирах) - T_ENV (температура окружающей среды)

Сложность сравнения состояла в том, что отопительный сезон после модернизации был теплее, чем до модернизации.

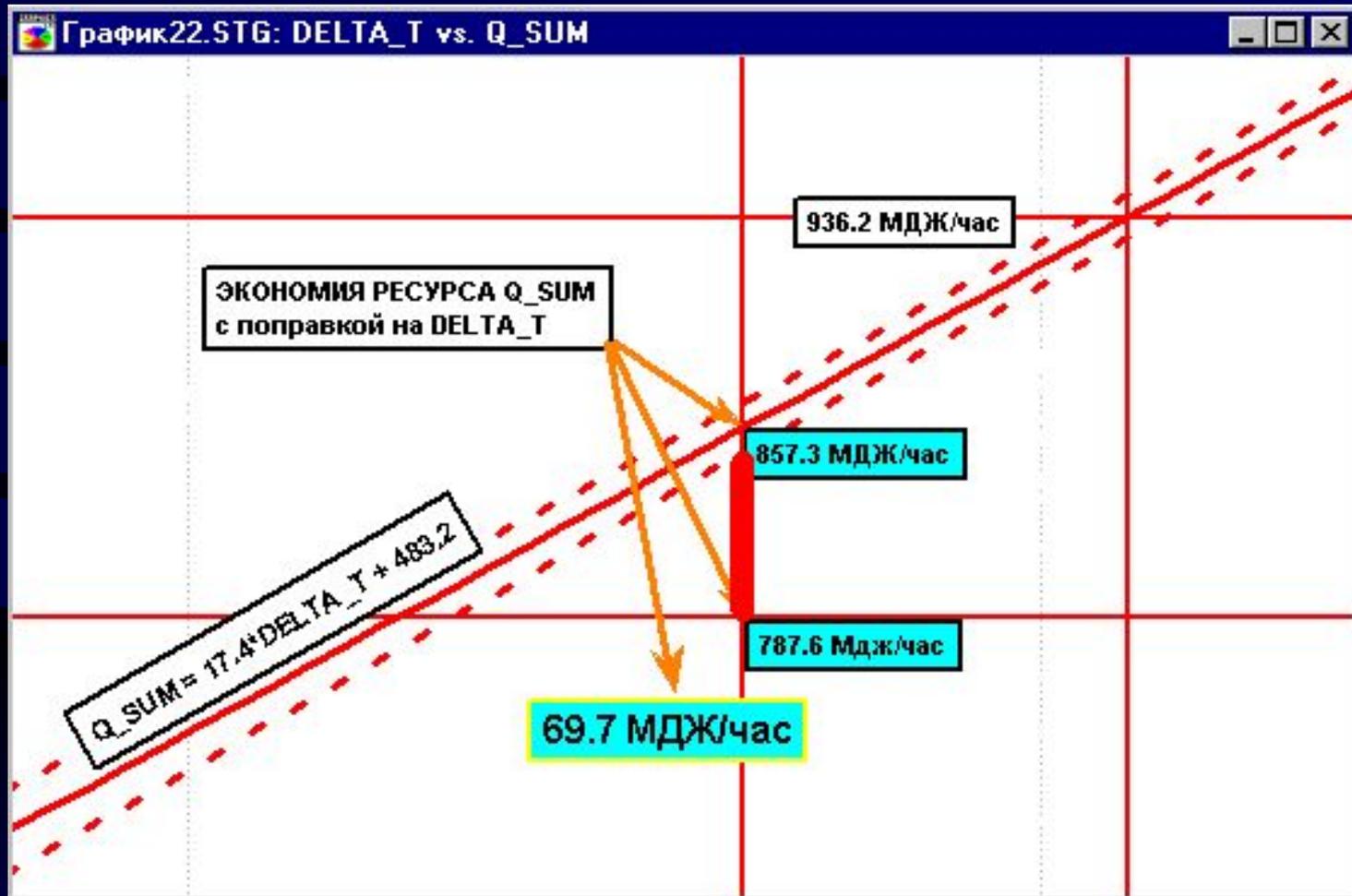
- Среднее значение DELTA_T на отопительном сезоне до модернизации равно 25.98.
95% доверительный интервал равен (25.7, 26.2), стандартная ошибка 0.1.
- Среднее значение DELTA_T на отопительном сезоне после модернизации равно 21.5.
95% доверительный интервал равен (21.3, 21.7), стандартная ошибка 0.9.
- Сдвиг средних значений составляет 4.4.
- Нужно провести коррекцию на сдвиг.

- **На следующих графиках показано, как провести коррекцию**

Результаты анализа



Результаты анализа



**Среднечасовая экономия
ресурса Q_SUM составляет
примерно
7%**

Дальнейшие результаты

Найдены оценки температур перехода из одного состояния комфортности в другое («недотоп-норма-перетоп») по характеру изменения потребления электроэнергии и расхода тепла на отопление

Системы	Температура перехода, °С	
	Недотоп – норма	Норма – перетоп
Теплоснабжение	18,8	23,3
Электроснабжение	19,8	23,3

Дальнейшие результаты

Определены зависимости суммарного потребления тепла Q при трех состояниях комфортности от разности температур ΔT в квартирах и на улице

Недотоп: $Q = 516 + 17 * \Delta T$

Норма: $Q = 401 + 20 * \Delta T$

Перетоп: $Q = 553 * \exp(0,02 * \Delta T)$

Дальнейшие результаты

Определены зависимости потребления электроэнергии E от разности температур ΔT в квартирах и на улице при разных состояниях комфортности

Недотоп: $E = 18,0 - 0,08 * \Delta T$

Норма и перетоп: $E = 11,3 + 0,09 * \Delta T$

Дальнейшие результаты

Исследованы зависимости расхода горячей и холодной воды от температуры теплоносителя в сети T_{hs}

Холодное водоснабжение: $V_{cw} = 0,016 * T_{hs}$

Горячее водоснабжение: $V_{dhw} = 0,015 * T_{hs}$

Дальнейшие результаты

Определены зависимости потребления тепла, и расхода холодной и горячей воды в зависимости от температуры теплоносителя в сети, а также зависимость потребления электроэнергии от температуры окружающей среды

Потребление тепла: $Q = -1145 + 19 * T_{hs}$

Расход холодной воды: $V_{cw} = -4,5 + 0,086 * T_{hs}$

Расход горячей воды: $V_{dhw} = -4,3 + 0,076 * T_{hs}$

Потребление
электроэнергии: $E = 12,9 - 0,19 * T_{os}$

**Оценка точности вычисления
средней температуры здания
(оценка репрезентативности выборки)**

**Рассматривается *пятиэтажное* жилое
здание с 6 подъездами.**

**Количество квартир: 79 (в крайних
подъездах по две квартиры на этаже, в
средних подъездах - по три квартиры на
этаже).**

**Датчики устанавливались в 16
квартирах..**

Измерения проводились:

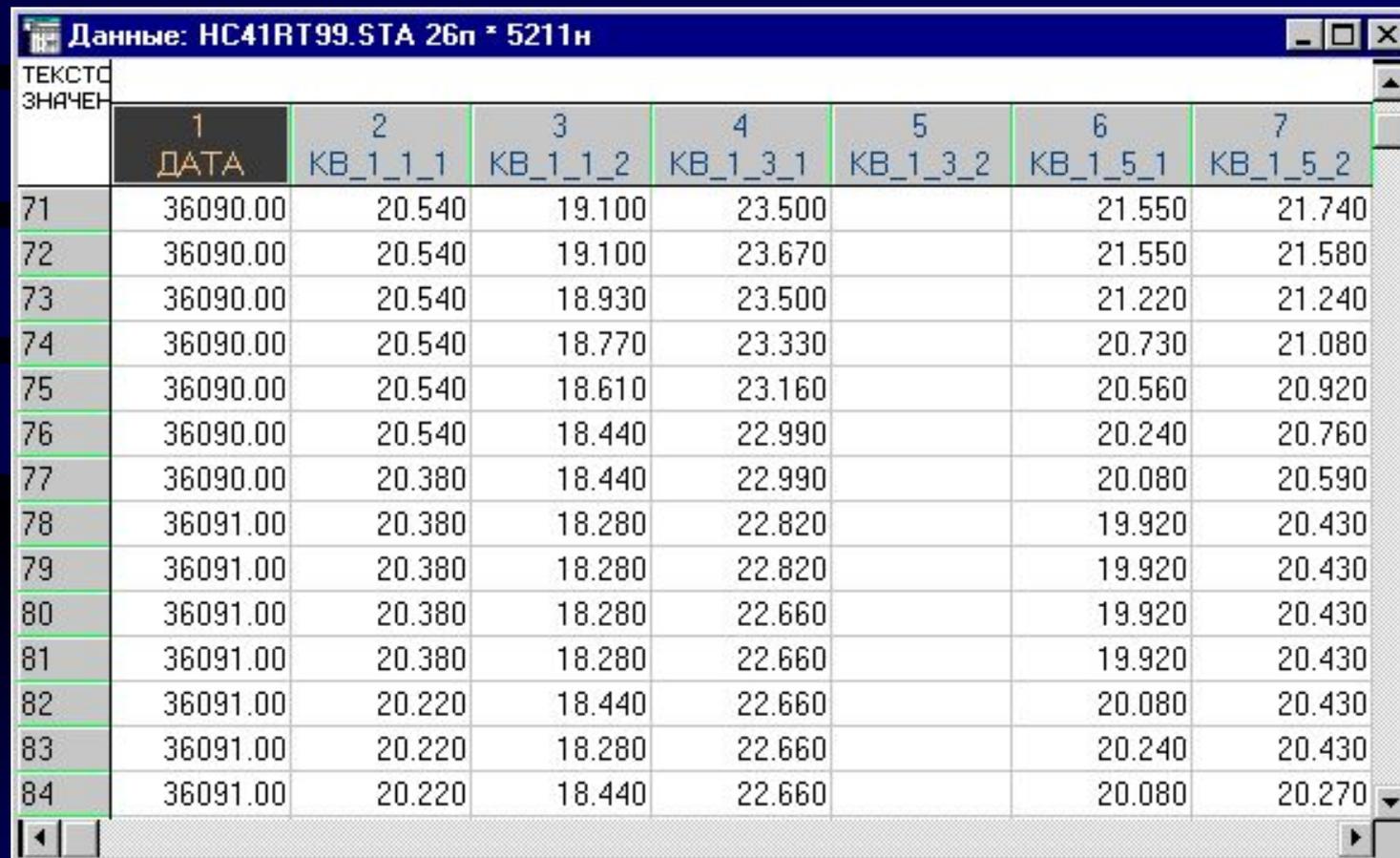
на 1, 3, 5 этажах

в 1-м, 3-м и 6-м подъездах

**Датчики устанавливались в 2-х квартирах
на каждом этаже.**

**На третьем этаже первого подъезда и на
пятом этаже шестого подъезда наблюдалось
только по одной квартире.**

Данные находятся в файле hc41rt99.sta:



Скриншот окна с данными из файла hc41rt99.sta. Таблица содержит 7 столбцов и 14 строк данных. Столбцы имеют следующие заголовки: 1 ДАТА, 2 КВ_1_1_1, 3 КВ_1_1_2, 4 КВ_1_3_1, 5 КВ_1_3_2, 6 КВ_1_5_1, 7 КВ_1_5_2. Строки пронумерованы от 71 до 84.

	1	2	3	4	5	6	7
ТЕКСТ	ДАТА	КВ_1_1_1	КВ_1_1_2	КВ_1_3_1	КВ_1_3_2	КВ_1_5_1	КВ_1_5_2
ЗНАЧЕН							
71	36090.00	20.540	19.100	23.500		21.550	21.740
72	36090.00	20.540	19.100	23.670		21.550	21.580
73	36090.00	20.540	18.930	23.500		21.220	21.240
74	36090.00	20.540	18.770	23.330		20.730	21.080
75	36090.00	20.540	18.610	23.160		20.560	20.920
76	36090.00	20.540	18.440	22.990		20.240	20.760
77	36090.00	20.380	18.440	22.990		20.080	20.590
78	36091.00	20.380	18.280	22.820		19.920	20.430
79	36091.00	20.380	18.280	22.820		19.920	20.430
80	36091.00	20.380	18.280	22.660		19.920	20.430
81	36091.00	20.380	18.280	22.660		19.920	20.430
82	36091.00	20.220	18.440	22.660		20.080	20.430
83	36091.00	20.220	18.280	22.660		20.240	20.430
84	36091.00	20.220	18.440	22.660		20.080	20.270

Описательный анализ

Далее...	N набл.	Среднее	Минимум	Максимум	Стд.откл.
KB_1_1_1	5211	19.57330	12.77000	28.44000	4.219331
KB_1_1_2	5211	20.21223	9.66000	28.48000	3.851137
KB_1_3_1	5173	19.46170	6.53000	29.88000	4.346816
KB_1_5_1	5211	20.52730	14.33000	27.39000	2.426025
KB_1_5_2	5211	20.76913	14.52000	27.42000	2.553897
KB_3_1_1	5211	20.41742	11.49000	25.75000	2.523740
KB_3_1_2	5211	20.68598	6.29000	29.53000	3.286092
KB_3_3_1	5211	23.34781	13.48000	25.92000	1.413080
KB_3_3_2	5211	23.65043	17.18000	28.66000	1.885842
KB_3_5_1	5211	24.70308	16.76000	27.81000	1.780534
KB_3_5_2	5211	23.74516	18.77000	32.63000	2.456055
KB_6_1_1	5057	21.39474	13.82000	27.39000	2.539762
KB_6_1_2	5057	20.01250	9.06000	28.62000	3.296055
KB_6_3_1	5057	23.21807	15.89000	29.43000	2.634394
KB_6_3_2	5057	22.56275	12.11000	30.48000	2.876231
KB_6_5_1	5211	17.15411	7.42000	25.21000	4.341535

Минимальная температура наблюдается в “крайних” квартирах: на первых этажах первого подъезда и на пятом этаже шестого подъезда.

Температура максимальна в третьем подъезде.

Средняя температура в квартирах колеблется от 17.1 до 24.7 градусов.

Для того чтобы найти группы квартир с близкой средней температурой, воспользуемся кластерным анализом, предварительно посуточно усреднив данные.

Кластерный анализ: кластеризация методом К средних

Переменные: 2-4 6-18

Кластер: Переменные (столбцы)

Число кластеров: 2

Число итераций: 10

Пропущенные данные: Построчное удаление

Начальные центры кластеров

- Выбрать набл., максимиз. начальные расстояния между кластерами
- Сортировать расстояния и выбрать набл. на постоянных интервалах
- Выбрать первые N (число кластеров) наблюдений

Пакетная обработка и печать

SELECT CASES Y 10 B

OK Отмена

Все квартиры разбились на 2 ярко
выраженных кластера:
теплые и холодные квартиры.

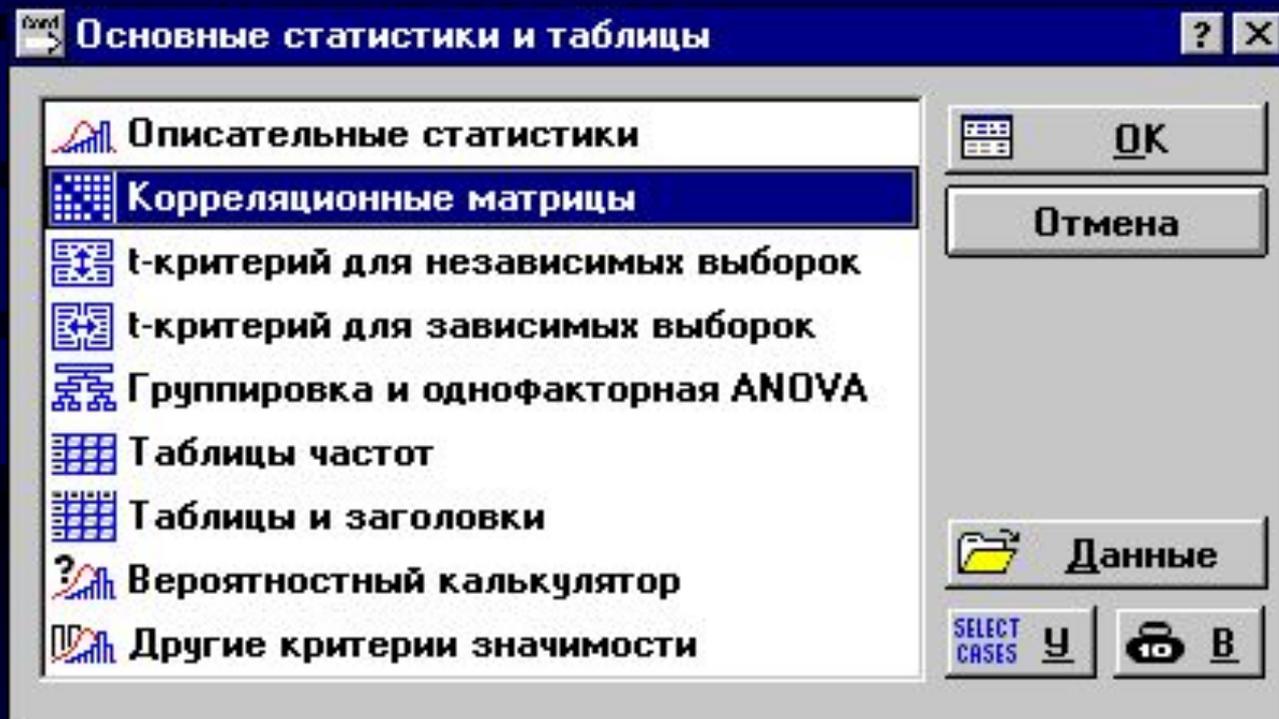


В среднем кластеры различаются на 4.6 градуса.

Среднее по теплым квартирам равно 23.5 градуса.

Среднее по холодным квартирам равно 19.9 градуса.

Вычислим корреляции между температурами в квартирах.



Из корреляционной матрицы видно, что между всеми переменными имеется сильная связь. Особенно сильная связь между переменными **кв_1_1_1** и **кв_1_1_2**.

Корреляции (hc41rt99.sta)							
Далее...	Отмеченные корреляции значимы на уровне $p < .05000$ N=5019 (Построчное удаление ПД)						
Перемен.	КВ_1_1_1	КВ_1_1_2	КВ_1_3_1	КВ_1_5_1	КВ_1_5_2	КВ_3_1_1	КВ_3_1_2
КВ_1_1_1	1.00	.94	.70	.87	.86	.67	.65
КВ_1_1_2	.94	1.00	.62	.84	.85	.69	.63
КВ_1_3_1	.70	.62	1.00	.63	.56	.71	.66
КВ_1_5_1	.87	.84	.63	1.00	.87	.59	.62
КВ_1_5_2	.86	.85	.56	.87	1.00	.61	.63
КВ_3_1_1	.67	.69	.71	.59	.61	1.00	.60
КВ_3_1_2	.65	.63	.66	.62	.63	.60	1.00
КВ_3_3_1	.60	.61	.57	.58	.59	.77	.54
КВ_3_3_2	.67	.65	.52	.63	.72	.66	.67
КВ_3_5_1	.65	.63	.68	.64	.68	.80	.64
КВ_3_5_2	.73	.72	.52	.69	.78	.62	.68
КВ_6_1_1	.85	.83	.77	.81	.82	.81	.69
КВ_6_1_2	.90	.87	.77	.85	.88	.76	.75
КВ_6_3_1	.83	.82	.74	.81	.85	.79	.72
КВ_6_3_2	.85	.86	.66	.82	.87	.72	.69

Для того, чтобы понять какие квартиры наиболее сильно *связаны* между собой воспользуемся кластерным анализом.

Кластерный анализ: объединение (древовидная кластеризация) ? X

Переменные: 2-4 6-18

Исходные данные: Исходные данные

Кластер: Переменные (столбцы)

Правило объединения (связи): Метод полной связи

Мера расстояния: 1-г Пирсона

д: 2 г: 2

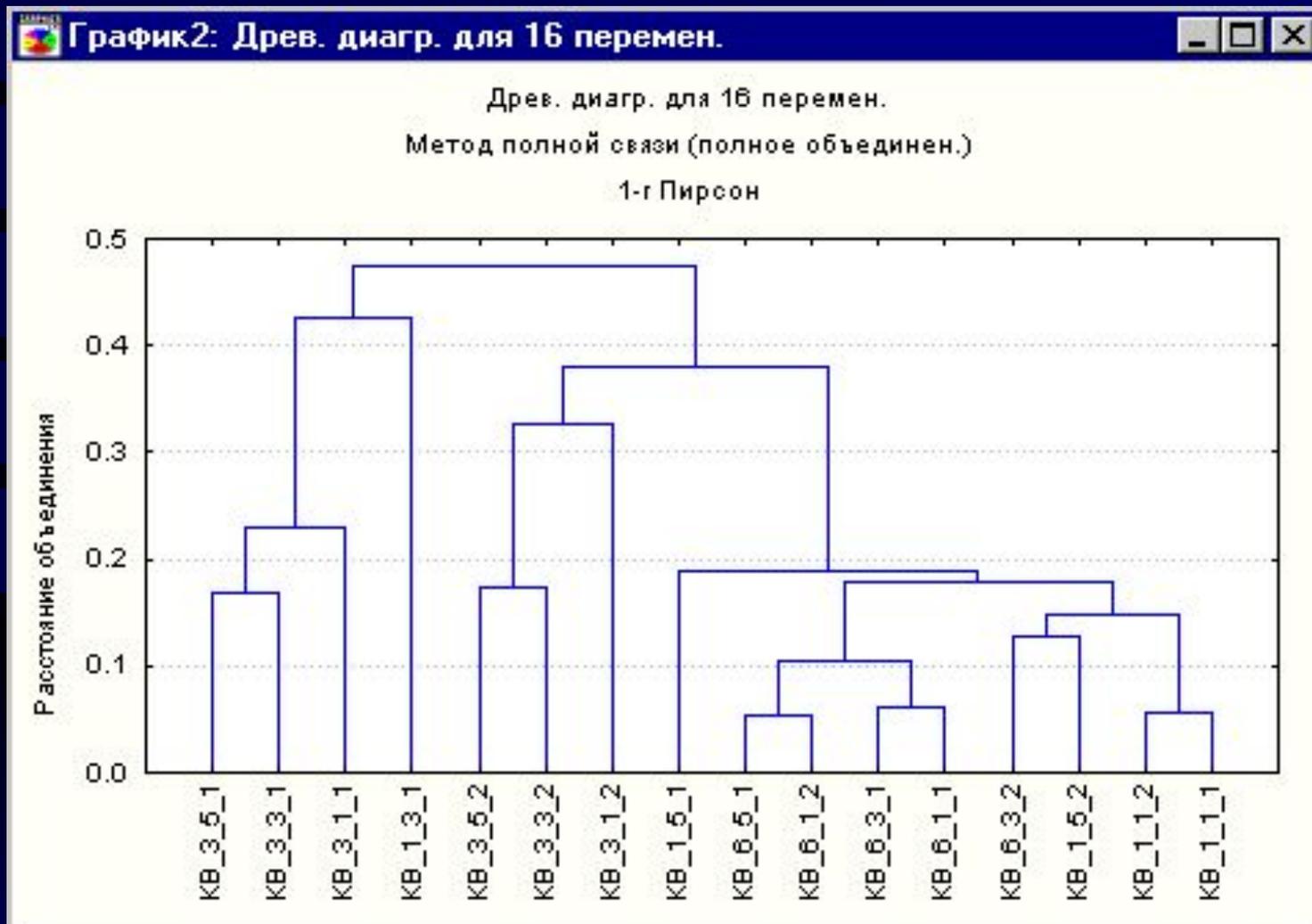
Пропущенные данные: Построчное удаление

Пакетная обработка и печать

SELECT CASES у 10 В

OK Отмена

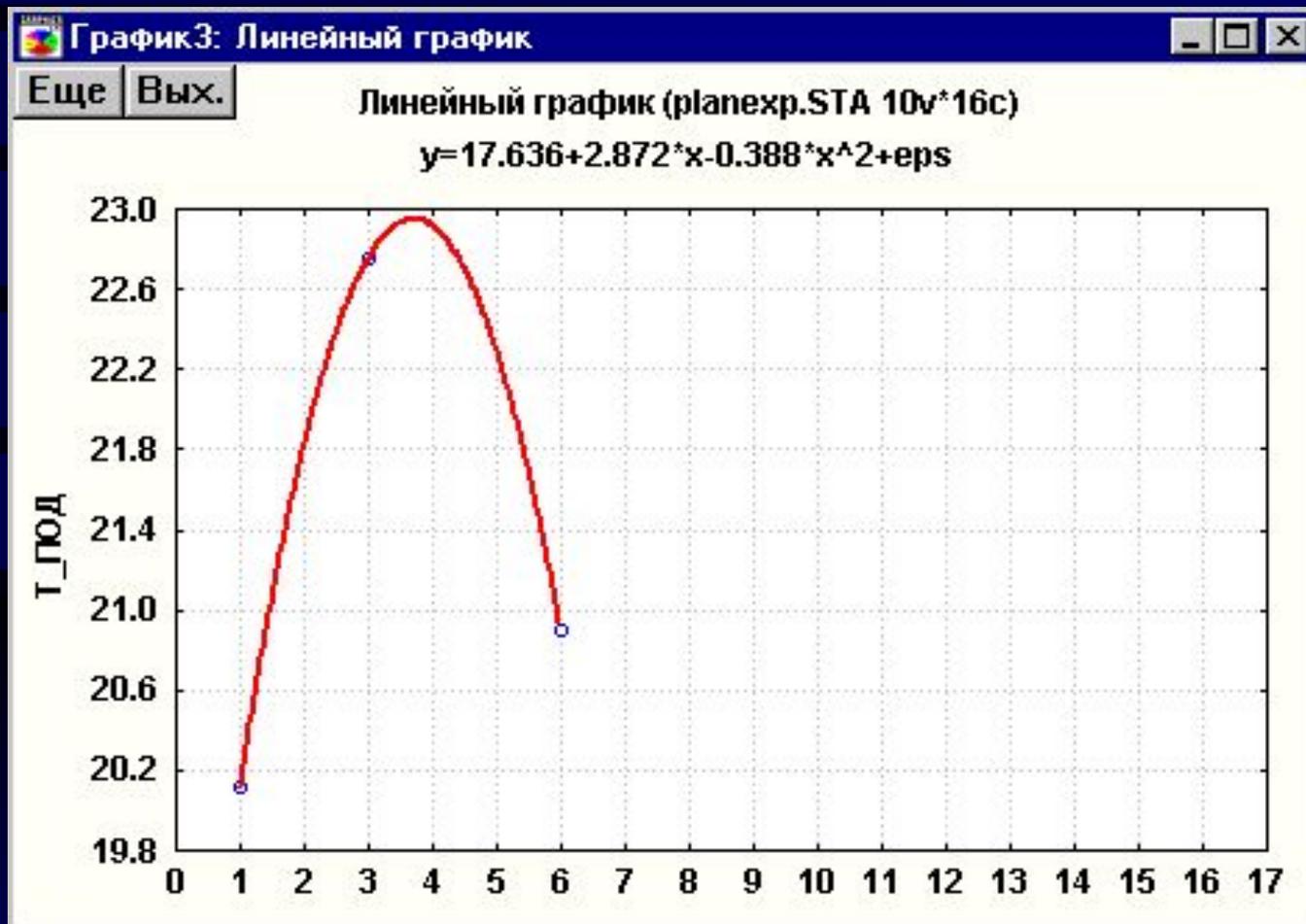
Квартиры образуют 4 кластера



**Зависимость температуры от номера
подъезда имеет вид:**

$$Y = 17.64 + 2.87 * X - 0.39 * X * X (*)$$

На графике зависимость имеет вид:



Используя формулу (*), вычислим среднюю температуру в зависимости от номера подъезда:

4 T_ПОД	5 T1_ПОД
20.120	20.120
	21.828
22.760	22.760
	22.916
	22.296
20.900	20.900

**Тогда средняя температура здания равна:
21.8**

**Средняя температура здания, вычисленная
по имеющимся данным, равна: 21.26**

Отсюда погрешность не превышает в среднем 0.54 градуса.

Итак, средняя температура здания оценивается с точностью половина градуса.