

Министерство образования Российской Федерации

Тамбовский государственный технический университет

ВВЕДЕНИЕ В СИСТЕМУ STATISTICA

Методические указания
к лабораторным занятиям для студентов 2 и 3 курсов
дневного отделения специальности 351400

Тамбов • Издательство ТГТУ • 2002

УДК 002.55:31
ББК Уя 73-5
Л17

Утверждено Редакционно-издательским советом университета

Рецензент
Кандидат технических наук, доцент
В. Г. Серегина

Л17 **Введение в систему STATISTICA:** Мет. указ. /
Сост.: Т. Я. Лазарева, Р. Н. Абалуев. Тамбов: Изд-во
Тамб. гос. техн. ун-та, 2002. 32 с.

Даны начальные сведения по работе в системе
STATISTICA, предназначенной для статистической
обработки данных в среде Windows.

Методические указания предназначены для
студентов 2 и 3 курсов дневного отделения
специальности 351400.

УДК 002.55:31
ББК Уя 73-5

© Тамбовский государственный
технический университет
(ТГТУ), 2002

ВВЕДЕНИЕ В СИСТЕМУ STATISTICA



• Издательство ТГТУ •

Учебное издание

ВВЕДЕНИЕ В СИСТЕМУ STATISTICA

Методические указания

С о с т а в и т е л и:

ЛАЗАРЕВА Татьяна Яковлевна,
АБАЛУЕВ Роман Николаевич

Редактор В. Н. Митрофанова

Подписано к печати 4.10.2002

Гарнитура Times New Roman. Формат 60 × 84/16. Бумага газетная.

Печать офсетная. Объем: 1,86 усл. печ. л.; 1,76 уч.-изд. л.

Тираж 100 экз. С. 668

Издательско-полиграфический центр ТГТУ
392000, Тамбов, Советская, 106, к. 14

ВВЕДЕНИЕ

Статистика – это наука, позволяющая увидеть закономерности в хаосе случайных данных, выделить устойчивые связи в них, определить наши действия с тем, чтобы увеличить долю правильно принятых решений среди всех принимаемых нами.

Статистика родилась из потребностей практики. Применяя статистические методы, можно найти закономерности в случайных данных, которые будут использоваться в дальнейшем при исследовании тех или иных процессов, явлений. Практическое применение статистических методов является чрезвычайно сложным делом, требует больших интеллектуальных усилий и временных затрат. С появлением мощных персональных ЭВМ бурно развивается рынок статистического программного обеспечения, открывающего путь к новым технологиям статистической обработки данных, максимально сокращающий рутинные процедуры.

Несомненным лидером на этом рынке является система STATISTICA, производимая фирмой StatSoft Inc. (США), представляющая собой интегрированную систему статистического анализа и обработки данных. Она состоит из следующих основных компонент: электронных таблиц для ввода и задания исходных данных, а также специальных таблиц для вывода численных результатов анализа; графической системы для визуализации данных и результатов статистического анализа; набора специализированных статистических модулей; специального инструментария для подготовки отчетов; встроенных языков программирования STATISTICA Command Language и STATISTICA BASIC, позволяющих расширить стандартные возможности системы.

Основными специализированными модулями системы являются: основные статистики и таблицы, непараметрическая статистика, дисперсионный анализ, множественная регрессия, нелинейное оценивание, кластерный анализ, факторный анализ, дискриминантный функциональный анализ, каноническая корреляция и др.

Широкое применение статистические методы нашли в экономике, в связи с этим студенты специальности 351400 "Прикладная информатика в экономике" должны знать и уметь пользоваться системой STATISTICA. Представляемые Вашему вниманию методические указания по использованию данной системы предназначены для студентов, а также могут быть использованы аспирантами, инженерами при решении задач статистического анализа. Знакомство с системой STATISTICA организовано на практических занятиях по курсам "Теория вероятностей и математическая статистика", "Эконометрика".

Тема 1 КРАТКОЕ ТЕХНИЧЕСКОЕ ОПИСАНИЕ СИСТЕМЫ STATISTICA

Цель занятия: Знакомство с основными возможностями системы STATISTICA.

Общие положения

Статистический анализ данных в системе STATISTICA разбивается на следующие основные этапы [1]:

- 1 Ввод данных в электронную таблицу исходных данных и их предварительное преобразование (построение необходимых выборок, ранжирование и т.д.).
- 2 Визуализация данных с помощью одного из типов графиков.
- 3 Применение конкретной процедуры статистической обработки.
- 4 Вывод результатов анализа в виде графиков и электронных таблиц с численной и текстовой информацией.
- 5 Подготовка и печать отчета.

STATISTICA работает с четырьмя различными типами документов; это

– электронная таблица *Spreadsheet*, предназначенная для ввода исходных данных и их преобразования;

– электронная таблица *Scrollsheet* для вывода численных и текстовых результатов анализа;

– график – документ в специальном графическом формате для визуализации и графического представления численной информации;

– отчет – документ в формате *RTF* (расширенный текстовый формат) для вывода текстовой и графической информации.

Одним из возможных и самых простых способов взаимодействия с системой является интерактивный режим работы, когда взаимодействие с системой осуществляется при помощи последовательного выбора различных команд из меню.

Работа с системой STATISTICA начинается с ее запуска.

ЗАПУСК STATISTICA

Запустить ОС Windows. После чего нажать кнопку "Пуск" и выбрать в меню "Программы" папку, которая содержит систему STATISTICA. В этой папке выбрать ярлык программы STATISTICA и щелкнуть по нем мышью (рис. 1).



Рис. 1 Запуск системы STATISTICA

После инсталляции системы папка STATISTICA будет содержать следующие элементы:

- *Setup* – установка (с помощью вспомогательной программы Reinstall.exe можно установить дополнительные модули, удалить ненужные модули и настроить конфигурацию по конкретной рабочей станции);
- *Help* – Справка (Справочная система STATISTICA);
- *Readme* – Важные замечания (Информация об инсталляции и деинсталляции программы, а также о последних изменениях, внесенных в систему и не вошедших в печатную документацию);
- *ModuleSwitcher* (обозначен STATISTICA)– Переключатель модулей (вызов статистического модуля, необходимого для анализа данных);

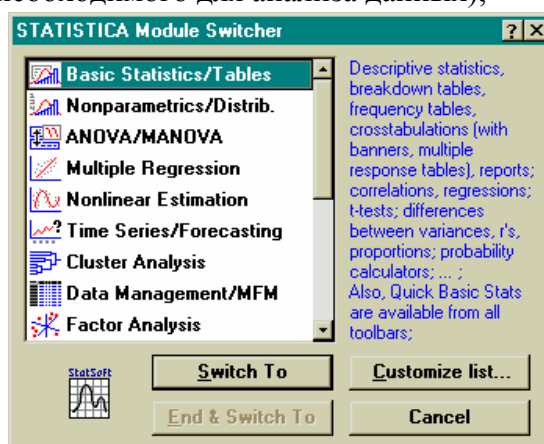


Рис. 2 Переключатель модулей

– *Basic Statistics and Tables* – Основные статистики и таблицы (один из статистических модулей системы, позволяющий вычислить основные статистические характеристики исходных данных).

После запуска STATISTICA на экране появляется "Переключатель модулей" (рис. 2).

В левой части диалогового окна расположен прокручиваемый список всех модулей, доступных в данной версии STATISTICA. В правой части дается краткое описание статистических процедур, содержащихся в выбранном модуле. Выбор конкретного модуля осуществляется либо двойным щелчком на названии модуля, либо одним щелчком и

нажатием одной из кнопок *Switch to – Переключиться на...*, *End&Switch to – Закончить и переключиться на...*.

Кнопка *Customize list... – Настройка списка...* позволяет расположить список используемых в работе модулей в верхней части списка. После нажатия на нее появляется следующее диалоговое окно (рис. 3).

В левой части данного окна располагается список модулей. В правой части кнопки:

- *Append* – *добавить* (выделенные модули перемещаются в конец списка);
- *Replace* – *заменить* (Список модулей заменяется на список выделенных модулей в окне *Customize List of Modules*);
- *Add / Remove* – *установка/удаление* (используется для изменения конфигурации системы).

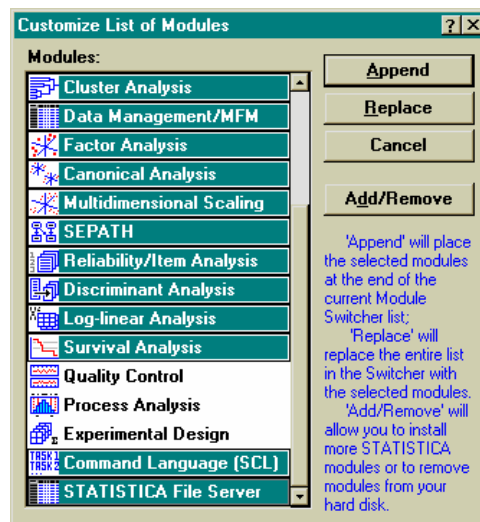


Рис. 3 Настройка списка модулей в "Переключателе модулей"

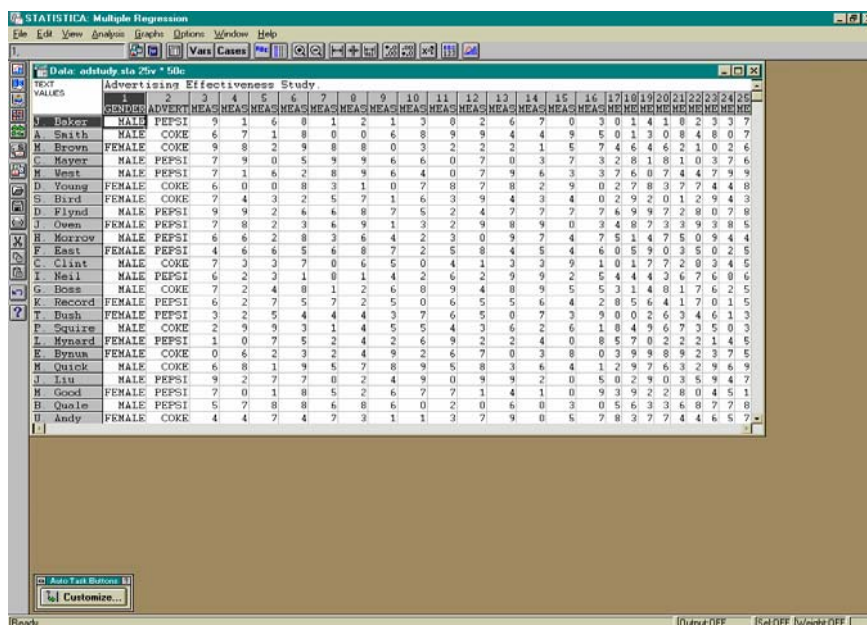


Рис. 4 Общий вид окна STATISTICA

Основные элементы окна STATISTICA

Общий вид окна STATISTICA представлен на рис. 4.

"Строка заголовка" содержит название активного статистического модуля, на рис. 4 – это STATISTICA *Multiple Regression*. В правой части этой строки расположены кнопки в направлении слева направо – минимизации размеров окна, восстановления окна и закрытия окна.

Вторую строчку окна занимает "строка меню", содержащая команды меню, упорядоченные по функциональному признаку. Выбор команды из этого, так называемого, выпадающего меню наиболее просто осуществлять при помощи мыши. Так как STATISTICA работает с четырьмя различными типами документов, которые соответствуют основным структурным компонентам системы, то каждый документ имеет свое меню и свою панель инструментов. Например, для электронной таблицы "SpreadSheet" ввода исходных данных меню содержит следующие команды:

File – файл, Edit – правка, View – вид, Analysis – Анализ, Graphs – графика, Options – опции, Window – окно, Help – справка.

"Панель инструментов" расположена в третьей строке и в левом первом столбце, она содержит кнопки для быстрого доступа к наиболее часто используемым командам меню.

Большую часть основного окна STATISTICA занимает рабочая область, на которую выводятся документы, с которыми работает пользователь. Обычно это следующие документы: электронные таблицы с исходными данными, электронные таблицы с результатами анализа, набор графиков, отчет, текстовый файл с программами на языках SCL и STATISTICA BASIC. Каждый документ выводится в своем собственном окне.

Окно документа состоит из строки заголовка, рабочей области для ввода и(или) отображения информации и инструментов по управлению окном.

Наряду с выпадающими меню STATISTICA имеет контекстные меню, вызываемые при помощи правой кнопки мыши и позволяющие получать быстрый доступ к наиболее часто используемым командам для работы с тем или иным объектом в активном окне. Так контекстное меню электронной таблицы с исходными данными обеспечивает доступ к следующим группам команд:

- построение различного типа графиков для выделенного блока значений;
- вычисление основных статистик;
- задание спецификаций переменных, операции с текстовыми значениями;
- преобразование структуры электронной таблицы (добавление, удаление строки и т.д.);
- операции с выделенным блоком значений;
- операции с "Буфером обмена Windows".

Порядок проведения занятия

- 1 Изучить основные возможности системы STATISTICA.
- 2 Войти в систему STATISTICA.
- 3 Изучить элементы окна STATISTICA: строка заголовка, меню, панель инструмента, рабочая область.

Контрольные вопросы

- 1 Какими возможностями располагает система STATISTICA?

- 2 Какие статистические модули имеются в системе?
- 3 Какие элементы входят в папку системы STATISTICA?
- 4 Что представляет собой строка заголовка?
- 5 Какие команды входят в строку меню?
- 6 Что такое контекстное меню?
- 7 Что представляет собой рабочая область STATISTICA?
- 8 Дать характеристику окна документа.

Литература: [1], [2], [3].

ТЕМА 2 ЭЛЕКТРОННЫЕ ТАБЛИЦЫ С ИСХОДНЫМИ ДАННЫМИ

Цель занятия: 1 Знакомство с электронными таблицами с исходными данными для проведения статистического анализа.

2 Освоить заполнение электронных таблиц с исходными данными.

Общие положения

Данные в системе STATISTICA организованы в виде электронной таблицы. Таблица с исходными данными является одним из типов документа. Каждый документ выводится в своем собственном окне в рабочей области системы. Как только это окно становится активным, изменяется панель инструментов и меню, появляются команды, доступные для этого типа документов.

Строки и столбцы электронной таблицы показываются в STATISTICA случаями и переменными соответственно. В столбцах находятся переменные, а в строках – значения, которые они принимают. Значения переменных могут быть числовые и текстовые.

Для работы с таблицами исходных данных существует большое количество инструментов, которые доступны при помощи выпадающих и контекстных меню и из панели инструментов, включая:

- операции, которые изменяют структуру электронной таблицы (добавление, удаление, копирование, перемещение переменных и случаев);
- операции по заданию спецификаций (имен, форматов и т.д.) для переменных и случаев;
- операции с выделенным блоком значений;
- операции, реализованные при помощи *Drag-and-Drop* – *Перетащить* и *Опустить*, включая операции по копированию, перемещению и автозаполнению блока и др.

Каждая переменная – *variables* имеет свое имя, формат и другие атрибуты, которые называются спецификацией и задаются пользователем. Переменная представляет собой наблюдаемую величину. Результаты наблюдений записываются в строках таблицы – случаях *Cases*. Ими могут быть даты наблюдений, либо другие имена естественным образом, возникающие в конкретной задаче.

Электронная таблица с исходными данными в STATISTICA называется *SpreadSheet*. В этой таблице можно задать спецификации переменных, которые включают в себя:

- формат отображения данных (количество десятичных разрядов в представлении чисел, формат представления дат, времени, соответствие между численными и текстовыми значениями и др.);
- код, приписываемый пропущенным данным;
- длинные имена переменных и комментарии к ним;
- метки для текстовых значений переменных, содержащих длинные текстовые значения – более восьми символов;
- формулы, используемые для определения, перекодировки или преобразования значений переменных;
- динамические связи между файлом данных STATISTICA и другим Windows – совместимым файлом с использованием механизма DDE.

Кроме атрибутов переменных можно задать также общую информацию обо всей таблице в целом, включая информационную часть и атрибуты внешнего оформления таблицы: информационную строку, отображаемую в поле заголовка таблицы; несколько строк комментариев или заметки о конкретном файле в виде текстового параграфа; список файлов, связанных с таблицей и называемых "Рабочей книгой" – "Workbooks", цвет, шрифт, размер, различных структурных элементов таблицы.

Для удобства работы с текстовыми переменными в системе STATISTICA реализован механизм "двойной записи", согласно которому, каждому текстовому значению переменной ставится в соответствие некоторое число, т.е. устанавливается соответствие вида "число = текстовое значение" либо автоматически, либо пользователем. При работе с данными всегда можно переключиться с текстовой на числовую форму и наоборот, что позволяет удобно вводить текстовые значения, выполнять необходимые преобразования, выполнять любой статистический анализ над текстовыми переменными. Например, переменные "Пол" и "Образование" содержат текстовые значения. STATISTICA автоматически ставит текстовому значению некоторое число. Для переменной "Пол": 1 = муж, 2 = жен; для переменной "Образование": 1 = среднее, 2 = высшее, 3 = нет. Переключение между формами представления данных осуществляется с помощью кнопки *Text Values – Текстовые значения* на панели инструментов, при этом в информационном поле таблицы, находящемся в левом верхнем углу, появляется строка *Numeric Values – Численные значения* или *Text Values – Текстовые значения*.

Следует отметить, что максимальная длина текстового значения не должна превышать восьми символов, в противном случае необходимо использовать

При пропуске значений переменной для внутреннего их хранения STATISTICA пустым ячейкам приписывает некоторый специальный код – *Missing Data Code – Код пропущенных данных*, значение которого по умолчанию равно 99999.

Пример создания файла с исходными данными

Создание файла с исходными данными удобно рассмотреть на следующем примере.

В результате социологического исследования был произведен опрос группы населения, в результате которого фиксировались следующие параметры: "Пол" (*Муж* – мужской, *Жен* – женский), "Возраст", "Образование" (*Нет*, *Среднее*, *Высшее*), "Средний доход". Результаты опроса занесены в табл. № 1.

1 Исходные данные социологического опроса

Имена опрошенных (названия случаев)	Пол	Возраст	Образование	Доход
Иванов И.	Муж	30	Среднее	535
Радченко И.	Жен	41	Высшее	340
Береговой П.	Муж	52	Нет	250
Часов Н.	Муж	25	Высшее	400
Петрова Н.	Жен	32	Среднее	200
Гвоздева А.	Жен	20	Среднее	220

Требуется создать электронную таблицу из четырех переменных (столбцов) и шести случаев (строк), ввести в нее текстовые и числовые значения.

Наиболее простым способом создания нового файла с исходными данными является способ с использованием команды *New Data – Создать данные* из меню *File – Файл*, которая доступна также по комбинации клавиш CTRL + N. Для создания файла с данными необходимо выполнить следующие шаги:

1 Имя файла.

После выбора команды *New Data* из меню *File* появляется диалоговое окно, в котором необходимо набрать имя файла. В рассматриваемом примере имя файла пусть будет *socio.sta*, причем расширение *sta – STATISTICA* добавит автоматически. Нажимается кнопка *OK*. Перед пользователем открывается электронная таблица *SpreadSheet* размером 10 на 10, т.е. состоящая из 10 переменных, которые по умолчанию имеют имена VAR1, VAR2, ..., VAR10 и 10 пронумерованных случаев, которые не имеют имен. В заголовке окна с электронной таблицей автоматически отображается имя файла и его размер (*socio.sta 10v*10c*).

2 Изменение структуры таблицы.

В электронной таблице необходимо удалить ненужные переменные VAR5–VAR10 и случаи с 7 по 10.

Для удаления переменных необходимо выполнить ряд операций в следующей последовательности:

– переменные VAR5–VAR10 выделяются при помощи мыши;

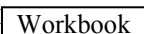
– нажимается кнопка – *vari* – *переменные*;

– выбирается команда *Delete – удалить*, появляется диалоговое окно *Delete Variables – Удаление переменных*;

– указывается в диалоговом окне диапазон удаляемых переменных с (*From variables*) VAR5 по (*To Variables*) VAR10.

Для удаления случаев используется – *Cases – Случаи* и команда *Delete – Удалить*. В появившемся диалоговом окне *Delete Cases – Удалить случаи* указывается диапазон удаляемых случаев с (*From Cases*) 7 по (*To Cases*) 10.

3 Заголовок таблицы. Комментарии к данным.


Заголовок таблицы "Социологический опрос" состоит из одной строки и выводится в первой строке таблицы. Дополнительная информация (комментарии) в таблице не выводится, но ее всегда можно просмотреть в специальном окне. Для задания этой информации надо дважды щелкнуть по заголовку таблицы или на кнопку  – *Рабочая книга*, после этого открывается диалоговое окно, в текстовые поля которого вводится необходимая информация.

В результате выполненных шагов таблица готова к вводу данных.

4 Имена переменных и случаев.

Существует несколько способов задания атрибутов переменной и атрибутов случаев.


Вызов диалогового окна задания атрибутов переменной можно осуществлять:

- дважды щелкнуть по имени переменной в электронной таблице, например, VAR1;
- установить указатель мыши на любую ячейку в этой переменной, нажать кнопку – "Переменные"  на панели инструментов, выбрать команду *Current Spec – Текущие спецификации*;
- в меню *Edit – Правка* выбрать команду *Variables, Current Spec*.

В появившемся диалоговом окне "Спецификации переменной" задаются следующие значения: *Name – Имя переменной* (Пол), длина имени не должна превышать восьми символов, поэтому имя переменной VAR3 – Образование необходимо сократить – Образ; *Display Format – Формат отображения* – число разрядов после десятичной точки задается параметром *Decimals*, а значащие параметром *Column Width*, по умолчанию принят формат – 8 значащих цифр, 3 после десятичной точки.

Для быстрого перехода от одной переменной к другой переменной следует использовать кнопки быстрого перехода: к следующей переменной {>>}, к предыдущей переменной {<<}.

Вызов окна задания имен случаев можно осуществить:

- дважды щелкнуть на номере или имени случая в электронной таблице;
- щелкнуть на кнопку – *Cases*  – *Случаи* на панели инструментов и выбрать команду *Names – Имена случаев*;
- в меню *Edit – Правка* выбрать команду *Cases – Случаи, Names – Имена*.

В открывшемся диалоговом окне *Case Name Manager – Менеджер имен случаев* задаются имена случаев.

5 Ввод данных.

Установить указатель мыши на ячейку, в которую необходимо ввести данные. Щелкнуть мышью, после этого ячейка становится активной (в диалоговом окне она заключается в рамку с более яркой границей по сравнению с другими ячейками). Вводится необходимое значение с клавиатуры, после чего осуществляется переход к другой ячейке либо с помощью клавиши ENTER (курсор перемещается вправо или вверх в зависимости от установок, определенных в постоянных параметрах системы).

6 Ввод текстовых значений.

Текстовые значения вводятся двумя способами:

- текст вводится непосредственно в таблицу;
- вводятся числовые значения, например, 1 для Муж, 2 для Жен; после этого в любую ячейку со значением 1 вводится текстовое значение Муж, после чего все значения 1 заменяются на текстовое значение для этой переменной.

7 Исправление данных.

Для исправления данных существует два способа:

– установить указатель мыши в ячейку, в которой необходимо исправить значение, щелкнуть мышью, ввести новое значение. Прежнее содержимое ячейки будет удалено и заменено на вновь введенное;

– установить указатель мыши в ячейку, в которой необходимо исправить значение, дважды щелкнуть мышью, исправить значение без его удаления. Альтернативный способ доступа к этой возможности – команда – *Cell – Ячейка* из меню *Edit – Правка*.

Файл исходных данных в системе STATISTICA создан.

8 Построение графиков.

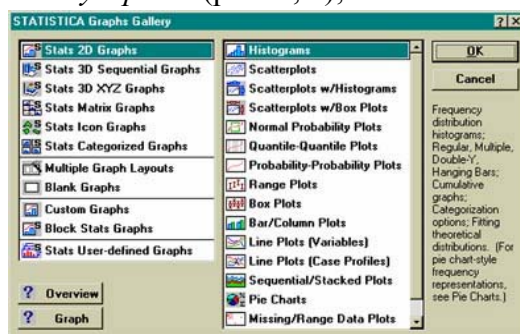
Важной характеристикой системы комплексной статистической обработки данных является наличие средств всесторонней графической поддержки процесса анализа данных. STATISTICA включает в себя большое количество разнообразных категорий и типов графиков. Это всевозможные графики на плоскости и в пространстве: гистограммы, матричные, категоризованные графики, диаграммы рассеяния, пиктографики и др.

Графические средства системы используются для визуализации численных и текстовых значений электронной таблицы с исходными данными. Для этого используются *Статистические графики – Stats Graphs* и *Пользовательские графики – CustomGraph*. Построение графиков удобно рассмотреть на примере.

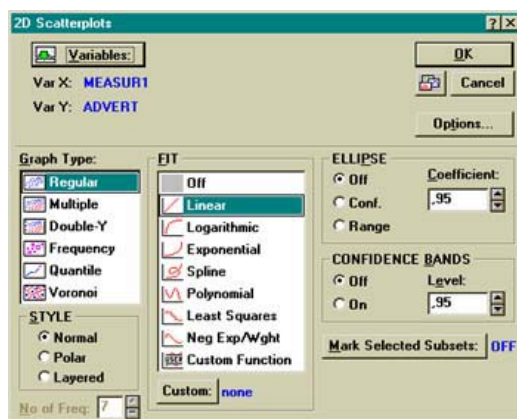
Пусть имеются наблюдения об уровне затрат на рекламу двух типов в квартал и объеме продаж некоторой продукции в соответствующий период. Эти данные занесены в электронную таблицу, состоящую из четырех переменных и 40 случаев. Требуется построить различные типы графиков.

Построение статистического графика состоит из следующих шагов:

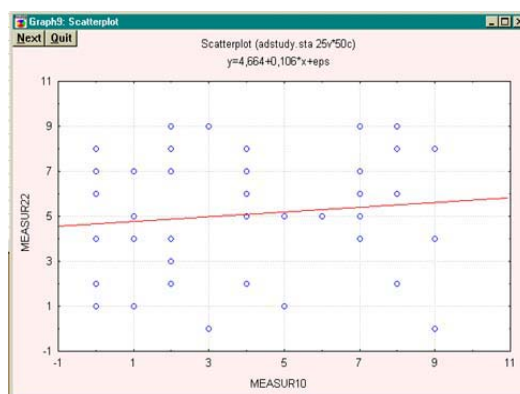
– выбор типа графика. Для быстрого доступа ко всем типам графиков из группы *Статистические графики* удобно использовать панель инструментов. Кнопка *STATISTICA Graphs Gallery – Галерея графиков STATISTICA* на панели инструментов электронной таблицы открывает диалоговое окно, в котором необходимо выбрать нужную категорию графиков (см. рис. 5, а), например, *Stats 2D Graphs – Статистические двумерные графики*, в этой группе выбрать необходимый тип графика – *Scatterplots Диаграмма рассеяния* и вид точечного графика *Regular – Регулярный* (рис. 5, б);



а)



б)



в)

Рис. 5 Диалоговые окна определения графиков

– выбор переменных. После задания вида графика появляется диалоговое окно, в котором необходимо выбрать переменные, зависимость между которыми необходимо изобразить графически. Для этого с помощью кнопки выбираются переменные Реклама 1 и Объем, откладываемые соответственно в осей OX и OY . Нажмите кнопку

– построение графика. На экране в отдельном окне появится график (рис. 5, в); настройка элементов графика. В STATISTICA имеется возможность настройки всех элементов графика. Для этого необходимо установить указатель мыши на необходимый элемент и либо дважды щелкнуть на нем, либо один раз нажать на правую кнопку мыши и выбрать команду из контекстного меню:

а) изменение заголовка графика. Изменить заголовок графика со стандартного на "Зависимость объема продаж от рекламы". Двойной щелчок на заголовке вызывает диалоговое окно *Edit Titles – Правка заголовков*. Внесите в него требуемый текст и выберите необходимый шрифт. Нажмите. Введенный заголовок сразу же отобразится на графике;

б) выбор вида точек. Измените вид точек, при помощи которых от аются данные. Дважды щелкните на какой-либо точке на графике. Появится диалоговое окно *Point pattern – Вид точек*. В прокручиваемом списке выбирается размер точки, например 7 и ее вид в списке *Pattern* – азцы. Нажмите .

Вид точек на графике изменится;

в) подгонка гладкой кривой. Пусть необходимо аппроксимировать наблюдаемые значения некоторым полиномом в смысле наименьших отклонений квадратов разности исходных данных и рассчитанных по полиному. Для этого указатель мыши ставится на

какую-либо точку построенного графика, нажатием на правую кнопку мыши вызывается контекстное меню, состоящее из двух пунктов. Выбирается пункт *Изменить расположение текущего индивидуального графика – Change series/plot layout*, далее в рамке *Fit – Подгонка* – способ приближения – при помощи полинома. В правом нижнем углу расположена рамка *Confidence Bands – Доверительный интервал*, в который переключатель отображения доверительного интервала устанавливается в положение *On – Включить*. Настройка графика закончена (рис. 6).

Построение Пользовательского графика производится в следующем порядке:

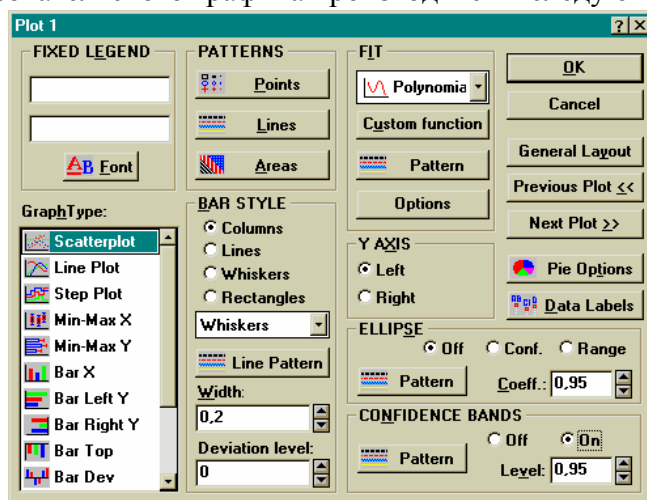


Рис. 6 Диалоговое окно, которое позволяет настроить все элементы графика, в том числе и подогнать гладкую кривую

– выделение блока значений. Выделяется какой-либо блок значений в исходной таблице с данными, например, две переменные и ряд случаев. Указатель мыши помещается на выделенный блок;

– построение графика. После нажатия на правую кнопку мыши открывается контекстное меню, из которого выбирается пункт *Custom Graphs – Пользовательские графики* и далее та категория графиков, которая естественным образом подходит для визуализации выделенных значений, например "3D Sequential Graphs – Трехмерные последовательные графики (диаграммы)". В диалоговом окне выбирается необходимый тип графика "Block Plot – Объемная диаграмма". После нажатия кнопки на экране появляется необходимый график;

– добавление текста в график. Необходимые пояснения и комментарии к графику можно сделать, находясь в окне с графиком. Для этого используется кнопка *Graphic Text Editor – Редактор текста в графике*.

В появившемся окне можно задать необходимый стиль выравнивания, изменить ориентацию текста, задать вид рамки и др., помимо непосредственного текста, который появится в левом верхнем углу графика после нажатия на кнопку . Этот текст можно перетащить в любое место графического окна, если подвести к нему указатель мыши и щелкнуть на левую кнопку;

– добавление стрелки в график. Добавляемая стрелка указывает на подмножество отображаемых случаев. Для этого используется кнопка *Arrow – Стрелка* на панели инструментов графика. В появившемся окне выбирается вид стрелки (форма, размер, цвет).

Порядок проведения занятия

- 1 Изучить порядок заполнения таблицы с исходными данными.
- 2 В соответствии с вариантом задания создать файл с исходными данными:
 - а) сделать заголовок таблицы, комментарии;
 - б) задать атрибуты переменных и случаев;
- 3 Построить статистические графики исходных данных.
- 4 Построить пользовательский график.

Контрольные вопросы

- 1 Что представляет собой электронная таблица с исходными данными?
- 2 Что такое атрибуты переменных и случаев?
- 3 Какие значения может принимать переменная?
- 4 Что значит изменить структуру таблицы?
- 5 Как исправить данные в электронной таблице?
- 6 Какие типы графиков имеются в системе STATISTICA?
- 7 Что представляет собой статистический график?
- 8 Назовите основные этапы построения статистического графика.
- 9 С какой целью строится пользовательский график?
- 10 Какие категории пользовательских графиков имеются в системе STATISTICA?

Литература: [1], [2], [3].

Тема 3 ВЫВОД РЕЗУЛЬТАТОВ СТАТИСТИЧЕСКОГО АНАЛИЗА

Цель занятия: 1 Знакомство с электронными таблицами вывода результатов статистического анализа.

- 2 Освоить вывод результатов статистического анализа в виде таблиц и графиков.

Общие положения

В результате обработки исходных данных при помощи статистических процедур, имеющих в системе STATISTICA, получают совокупность численной, текстовой и графической информации.

Численные и текстовые результаты статистического анализа выводятся в виде таблиц *scrollsheet*, являющихся удобным средством для вывода, анализа и хранения численных результатов исследования.

Таблица *scrollsheet* является одним из типов документов системы STATISTICA. Панель инструментов и основные приемы работы с таблицей *scrollsheet* напоминают основные приемы работы с таблицей исходных данных, но сами таблицы принципиально отличаются от последних. Основное различие заключается в назначении этих таблиц, в связи с чем

имеются различия и в инструментах для работы с этими таблицами, а также некоторые технические различия, например, одновременно может быть открыто несколько таблиц *scrollsheet*.

Большинство средств для работы с таблицей *scrollsheet* доступно при помощи кнопок на панели инструментов и контекстного меню. Они включают в себя:

- стандартные операции по редактированию таблицы;
- операции удобного просмотра;
- доступ к основным статикам и графическим возможностям системы.

При выводе целого ряда результатов необходимые значения можно выделить при помощи цвета, сама таблица может быть распечатана несколькими способами, кроме того в ней можно изменить все аспекты формата, отредактировать результат, добавить пустые строки и столбцы и .т.п.

Таблицы *scrollsheet* могут быть сохранены как:

- документ *scrollsheet*, который может быть открыт позже и использован в процессе анализа;
- стандартный файл данных STATISTICA, который может быть использован в качестве исходных данных;
- текстовый файл.

Информация выводится в виде последовательности, состоящей из набора таблиц *scrollsheet* и графиков и имеющей некоторую фиксированную длину, при превышении которой закрывается самый старый график или таблица. Естественно имеется возможность заблокировать удаление необходимых для дальнейшего анализа результатов. Кроме того, любая таблица *scrollsheet* или график могут быть автоматически направлены в отчет.

Пример вывода результатов

Вывод результатов статистического анализа удобно рассмотреть на примере расчета некоторых статистических характеристик в модуле *Basic statistic & tables – Основные статистики и таблицы*.

Пусть создан файл с исходными данными, содержащий объем, затраты и выручку от продажи продукции. Требуется рассчитать основные описательные статистики этих данных – среднее, максимум, минимум и др.

Для вывода результатов необходимо выполнить следующие шаги:

- 1 Запуск модуля и открытие файла с данными.

Запуск модуля *Basic statistic & tables – Основные статистики и таблицы* производится *Переключателем модулей*, который вызывается после запуска STATISTICA и дает прокручиваемый список модулей. После запуска модуля на рабочем пространстве STATISTICA открывается стартовая панель модуля (рис. 7). После нажатия на кнопку *Open Data – Открыть данные* и выбора файла с данными product.sta в диалоговом окне выбора файла на рабочем пространстве STATISTICA будет открыта электронная таблица с исходными данными.

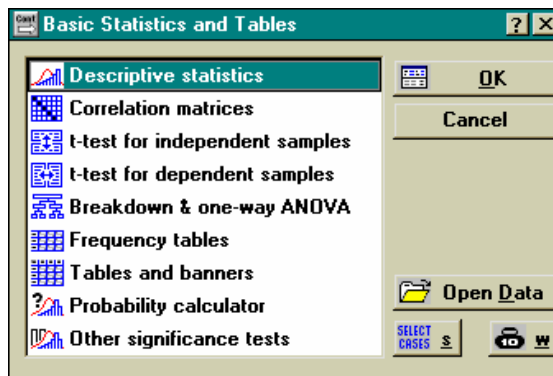


Рис. 7 Стартовая панель модуля "Basic statistic & tables"

2 Выбор процедуры статистического анализа.

Из стартовой панели модуля выбирается процедура *Descriptive statistics* – *Описательные статистики*. В появившемся диалоговом окне задаются имена переменных, для которых необходимо вычислить описательные статистики. Для этого пользуются кнопкой *Variables – Переменные*. После выбора переменных нажимается ОК.

3 Вывод результатов в виде таблиц *scrollsheet*.

После выполнения пунктов 1 и 2 автоматически создается таблица *scrollsheet*, в которую будут выведены результаты анализа. Эта таблица выводится в новом окне на рабочем пространстве STATISTICA (рис. 8).

Valid N	Mean	Minimum	Maximum	Std. Dev.
EXPERIENCE 25	16.88000	4.000000	32.00000	9.075241
SUCCESS 25	.44000	0.000000	1.00000	.506623

Рис. 8 Результаты расчета в электронной таблице *scrollsheet*

В столбцах этой таблицы расположены вычисленные статистики для соответствующих переменных. Имена строк и столбцов. Форматы вывода результатов. Ширина столбцов автоматически задаются системой. Внешне эта таблица напоминает таблицу с данными, но между тем имеет и некоторые различия:

- в левом верхнем углу расположена кнопка *Continue – Продолжить*;
- таблица имеет стандартное название и не имеет информационного заголовка;
- при активизации окна (щелкнуть мышью в любом месте этого окна) меняется содержание пунктов меню и панель инструментов.

4 Продолжение анализа. Вычисление корреляционной матрицы. Сохранение таблицы *scrollsheet*.

Для продолжения анализа необходимо нажать кнопку *Continue – Продолжить* и в появившемся диалоговом окне выбрать опцию *Correlation matrix – Корреляционная матрица*. После чего выбираются переменные, для которых требуется вычислить корреляционную матрицу, и нажимается кнопка ОК. В результате этих операций создается новая таблица *scrollsheet*, в которую выводятся корреляционная матрица, а значимые коэффициенты помечаются красным цветом.

5 Доступ к графическим возможностям.

В процессе анализа часто требуется представить результаты в графическом виде. Например отобразить корреляцию между переменными "Выручка" и "Затраты". С этой целью указатель мыши перемещается на соответствующий коэффициент корреляции и нажатием на правую кнопку вызывается контекстное меню, которое обеспечивает быстрый доступ к основным операциям с ячейками таблицы *scrollsheet*.

В контекстном меню выбирается опция *Quick STATS Graphs – Быстрые статистические графики*, а затем один из типов графика. Необходимый график появится на экране в отдельном графическом окне.

6 Ввод результатов анализа.

После вывода статистического графика на рабочем пространстве STATISTICA находятся четыре открытых документа:

- файл с исходными данными;
- две таблицы с результатами анализа (Descriptive statistics, Correlations);
- график.

Подобная ситуация типична для большинства видов статистического анализа. Если же расчеты содержат много промежуточных результатов, которые генерируют набор таблиц и графиков, то это приводит к тому, что все рабочее пространство будет заполнено, что крайне неудобно. В связи с этим в STATISTICA реализована возможность автоматического закрытия самых старых таблиц и графиков, т.е. тех которые были выведены первыми.

7 Создание отчета.

Часто бывает необходимо создавать файл регистрации всех результатов анализа, т.е. таблиц и графиков, появившихся на экране. Такой файл называется *отчетом*. При автоматическом выводе всех результатов анализа в отчет, последний получил название автоотчет.

Для создания автоотчета необходимо:

1 Выбрать опцию *Automatically Print All Scrollsheet – Автоматически печатать все*, в диалоговом окне *Page/Output Setup – Параметры страницы/вывода* задать:

- канал вывода (принтер, текстовый файл, окно с отчетом);
- заголовок отчета;
- опции автоматического создания отчета.

2 Все таблицы, появляющиеся на экране автоматически выводятся в файл с отчетом, который открыт в отдельном окне на рабочем пространстве системы STATISTICA.

Порядок проведения занятия

- 1 Изучить порядок вывода результатов в системе STATISTICA.

- 2 В соответствии с вариантом задания провести расчет некоторых статистических характеристик в модуле *Основные статистики и таблицы*.
- 3 Вывести результаты расчета в виде таблиц *scrollsheet*.
- 4 Вычислить корреляционную матрицу.
- 5 Построить статистические графики.
- 6 Создать автоотчет.

Контрольные вопросы

- 1 Что представляет собой таблица *scrollsheet*?
- 2 Как сохранить таблицу *scrollsheet*?
- 3 Каким образом осуществляется быстрый доступ к графическим возможностям из таблицы *scrollsheet*?
- 4 Что такое отчет?
- 5 Как создать автоотчет?

Литература: [1].

Тема 4 МНОЖЕСТВЕННАЯ РЕГРЕССИЯ

Цель занятия: 1 Знакомство с модулем системы STATISTICA *Multiple Regression – Множественная регрессия*.

- 2 Освоить процедуру построения линейной регрессии.
- 3 Освоить процедуру построения множественной регрессии.

Общие положения

Статистические процедуры системы сгруппированы в нескольких модулях, одним из которых является модуль *Multiple Regression – Множественная регрессия*, включающий в себя набор средств множественной линейной и фиксированной нелинейной регрессии. Работу этого модуля удобно рассмотреть на примере построения простой линейной регрессии, а затем на примере линейной множественной регрессии.

Простая линейная регрессия

Уравнение $y = \beta_1 x + \beta_0 + \varepsilon$ называется уравнением простой линейной регрессии, здесь обозначено x – независимая переменная, y – зависимая переменная, β_1 , β_0 – неизвестные параметры, описывающие прямую, которые требуется оценить по результатам наблюдений (x_i, y_i) , $i = \overline{1, n}$; ε – независимая случайная ошибка с нулевым средним, интерпретируемая как

ошибка наблюдений. Требуется построить уравнение простой регрессии по результатам наблюдения (x_i, y_i) . В этом случае получить точные значения параметров β_1, β_0 невозможно, так как требуется бесконечная выборка, поэтому получают их оценки b_1, b_0 , и уравнение регрессии записывается в виде $y = b_1x + b_0$ построение этого уравнения с помощью системы STATISTICA рассмотрим на примере данных [2] представляющих собой оптовые цены за одну бутылку на марочные вина (y) в зависимости от года закладки вина (x) (табл. 2).

После запуска системы необходимо выбрать модуль *Multiple Regression* – *Множественная регрессия*, после запуска которого появляется "Стартовая панель" модуля, содержащая основные операции, доступные в запущенном модуле, и различные параметры анализа. Подлежащие определению.

2 Исходные данные по закладке вина

Год (x)	Цена (y)
1890	50,00
1900	35,00
1920	25,00
1931	11,98
1934	15,00
1935	13,00
1940	6,98
1941	10,00
1944	5,99
1948	8,98
1950	6,98
1952	4,99
1955	5,98
1960	4,98

На первом шаге необходимо создать электронную таблицу с исходными данными, в качестве переменных в рассматриваемом примере выступают "Год" закладки вина и его "Цена" на аукционе. В рассматриваемом примере удобнее работать с новыми переменными: возраст и логарифм цены, которые связаны с исходными переменными соотношениями: возраст = 1972 – год, Цена_Лог = $\ln(\text{Цена})$. Формулы преобразования переменных задаются в диалоговом окне спецификаций переменных. Эти переменные принимают численные значения, их имеет смысл отобразить на графике. Для вызова графических возможностей системы следует воспользоваться меню *Graphics* – *Графика*. Построим двумерные диаграммы рассеяния. В диалоговом окне при помощи кнопки *Variables* – *Переменные* выбираются необходимые переменные (log, Цена), которые будут отражаться графически, а также выбирается тип графика. После нажатия на кнопку *OK* график выводится в отдельном окне на рабочем пространстве системы.

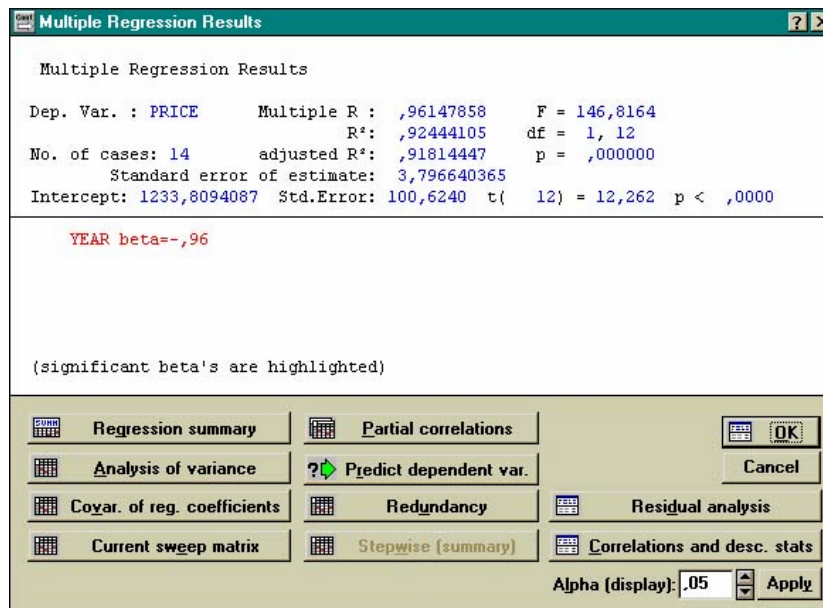


Рис. 9 Окно с результатами анализа

Для начала проведения статистического анализа вызывается "Стартовая панель" модуля, в котором задаются различные опции анализа. Выбирается зависимая переменная "Цена" и независимая переменная "Возраст". Для их задания используют кнопку *Variables – Переменные*, при нажатии на которую открывается окно *Select dependent and independent variable list – Выбор зависимой переменной и списка независимых переменных*, в котором выбираются соответствующие переменные (Цена, возраст). Нажатием на кнопку ОК возвращаемся в "Стартовую панель" модуля.

Для вывода результатов и их анализа (рис. 9) на стартовой панели необходимо нажать на кнопку ОК. Система производит вычисления, и через секунду окно результатов появится на экране. Окно результатов анализа имеет следующую простую структуру: верхняя часть окна – информационная, нижняя содержит функциональные кнопки, позволяющие всесторонне просмотреть результаты анализа.

В информационной части окна содержится краткая информация о проведенном анализе, а именно:

- Dep. Var. – имя зависимой переменной: Цена.
- No. Of Cases – число случаев, по которым построена регрессия: 14.
- Multiple R – коэффициент множественной корреляции.
- R – square – RI – квадрат коэффициента множественной корреляции.
- Adjusted R – square: adjusted RI – Скорректированный коэффициент корреляции, определяемый как

$$\text{Adjusted } R - \text{square} = 1 - (1 - R - \text{Square}) (n / (n - p)),$$

где n – число наблюдений в модели; p – число параметров модели.

- Std. Error of estimate – Стандартная ошибка оценки. Эта статистика является мерой рассеяния наблюдаемых значений относительно регрессии прямой.

– Intercept – Оценка свободного члена регрессии. Значение коэффициента b_0 в уравнении регрессии.

– Std. Error – Стандартная ошибка оценки свободного члена. Стандартная ошибка коэффициента b_0 в уравнении регрессии.

– $t(df)$ and p -value – Значение t -критерия и уровня p . t -критерий используется для проверки гипотезы о равенстве нулю свободного члена регрессии.

– F – значения F -критерия.

– df – Число степеней свободы F -критерия.

– p – уровень значимости.

В рассматриваемом примере значение коэффициента детерминации $RI = 0,929$, что значит, что построенная регрессия объясняет 92,9 % разброса значений ЦЕНА относительно среднего.

Для проверки значимости регрессии используют F -критерий. В данном случае этот критерий используется для проверки гипотезы, утверждающей, что между зависимой переменной ЦЕНА и независимой переменной ВОЗРАСТ нет линейной зависимости, т.е. $b_1 = 0$ против альтернативы $b_1 \neq 0$. Большое значение F -критерия: $F = 157,0486$ и уровень значимости $p = 0,0000$ показывает, что построенная регрессия высоко значима.

При нажатии на кнопку *Regression summary – Краткие результаты регрессии* на экране появляется электронная таблица с результатами анализа (рис. 10).

В третьем столбце этой таблицы расположены оценки неизвестных параметров модели:

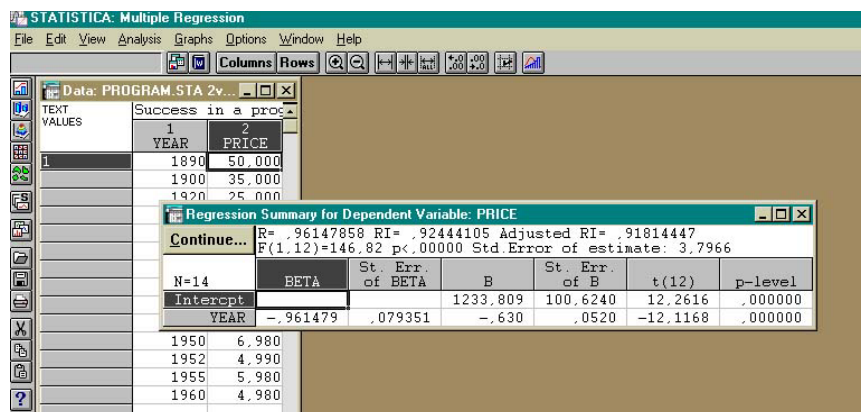


Рис. 10 Таблица с результатами анализа

В третьем столбце этой таблицы расположены оценки неизвестных параметров модели:

$$b_0 = 1,143891,$$

$$b_1 = 0,034652,$$

т.е. модель имеет вид

$$\text{ЦЕНА_ЛОГ} = 0,034652 \cdot \text{ВОЗРАСТ} + 1,143891.$$

Требуемая регрессия построена, в исходных данных модель имеет вид

$$\text{ЦЕНА} = \exp(0,034652 \cdot \text{ВОЗРАСТ} + 1,143891).$$

Система STATISTICA помимо вывода результатов анализа в виде электронных таблиц и графиков позволяет создать специальный файл с отчетом. Для него на рабочем пространстве системы открывается специальное окно. В котором можно "распечатать" любой документ.

Множественная регрессия

В этом случае модель содержит несколько независимых переменных и записывается как $Y = Xb$, где Y – вектор наблюдений, состоящий из n переменных, X – матрица независимых переменных, размером $m \times n$, m – число независимых переменных, b – вектор неизвестных параметров, которые необходимо оценить.

Также. Как и в случае линейной регрессии прежде всего создается электронная таблица с исходными данными, которые в случае необходимости могут быть преобразованы к желаемому виду. С целью полного исследования исходных данных строятся специализированные статистические графики.

Для начала статистического анализа вызывается "Стартовая панель" модуля, в которой задаются различные оценки анализа. После проведения вычислений появляется окно результатов со следующей простой структурой: верхняя часть окна – информационная, нижняя содержит функциональные кнопки, позволяющие всесторонне просмотреть результаты анализа.

ПОРЯДОК ПРОВЕДЕНИЯ ЗАНЯТИЯ

- 1 В соответствии с номером задания (табл. 1) построить уравнение простой линейной регрессии.
- 2 Создать электронную таблицу с исходными данными в соответствии со своим вариантом.
- 3 Построить двумерные диаграммы рассеяния.
- 4 Выбрать зависимую и независимую переменные.
- 5 Провести вычисления.
- 6 Провести всесторонний анализ полученных результатов.
- 7 Создать отчет.

Контрольные вопросы

- 1 Назовите порядок операций, которые необходимо выполнить для построения уравнения простой регрессии.
- 2 Что представляет собой окно результатов анализа?
- 3 Какую информацию содержит информационная часть окна?
- 4 С помощью каких статистических критериев оценивается значимость коэффициентов уравнения регрессии?
- 5 Какую информацию содержит отчет?

Литература: [1], [2].

СПИСОК ЛИТЕРАТУРЫ

- 1 Боровиков В. П., Боровиков И. П. STATISTICA: Статистический анализ и обработка данных в среде Windows. – М.: Информационно-издательский дом "Филинь", 1998. – 608 с.
- 2 Дубров А. М. и др. Многомерные статистические методы. – М.: Финансы и статистика, 1998. – 352 с.
- 3 Герасимович А. И. Математическая статистика. – Минск: Высш. шк., 1983. – 271 с.