



МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ
ФЕДЕРАЦИИ

Федеральное агентство по образованию

Рубцовский индустриальный институт (филиал) ГОУ ВПО

«Алтайский государственный технический университет им. И.И. Ползунова»

А.В. Шашок

ЛАБОРАТОРНЫЙ ПРАКТИКУМ В СИСТЕМЕ «STATISTICA»

Методические указания к выполнению лабораторных работ по дисциплине

"Статистика" для студентов всех форм обучения специальности 080507

"Менеджмент организации"

Рубцовск 2010

Шашок А.В. Лабораторный практикум в системе «STATISTICA»: Методические указания к выполнению лабораторных работ по дисциплине «Статистика» для студентов всех форм обучения специальности 080507 «Менеджмент организации»/Рубцовский индустриальный институт. – Рубцовск, 2010. – 81 с.

Методические указания предназначены для студентов всех форм обучения при выполнении лабораторных работ по дисциплине «Статистика» в программном комплексе «STATISTICA». Методические указания содержат краткое описание программного комплекса, теоретический материал (при необходимости), задания к лабораторным работам, а также последовательность их выполнения. Выполнение лабораторных работ в системе «STATISTICA» позволяет студентам закрепить теоретический материал и приобрести практические навыки автоматизации проведения статистических исследований.

Рассмотрены и одобрены на
заседании кафедры «Менеджмент и
экономика» Рубцовского
индустриального института. Протокол
№ от . 2010 г.

Рецензент:

к.э.н., доцент,

зав. кафедрой «Финансы и кредит»

РИИ АлтГТУ

Е.В. Касаткина

©Рубцовский индустриальный институт, 2010

Содержание

| | |
|--|----|
| 1. Лабораторная работа №1 | 4 |
| 2. Лабораторная работа №2 | 17 |
| 3. Лабораторная работа №3 | 26 |
| 4. Лабораторная работа №4 | 36 |
| 5. Лабораторная работа №5 | 50 |
| 6. Лабораторная работа №6 | 57 |
| 7. Лабораторная работа №7 | 70 |
| Правила оформления отчета по лабораторной работе | 80 |
| Приложение | 81 |

1. Лабораторная работа №1

Тема работы: Основные приемы работы в системе «STATISTICA» (2 часа).

Цель работы: Освоение основных приемов работы с документами системы «STATISTICA».


Содержание работы:

1. Запуск системы STATISTICA и ознакомление с основными элементами программы.
2. Создание, редактирование и сохранение файла с исходными данными.
3. Создание и редактирование графиков.
4. Создание автоотчета.

Выполнение работы.

1. Запуск системы STATISTICA и ознакомление с основными элементами программы.

Запуск системы STATISTICA осуществляется аналогично запуску любого Windows-приложения:

- либо из меню *ПУСК - ПРОГРАММЫ-STATISTICA*;
- либо двойным щелчком мыши на ярлыке программы  на рабочем столе.

Окно системы STATISTICA похоже на окно других прикладных программ Windows, таких как MS Word, MS Excel, и состоит из следующих основных элементов: строка заголовка; строка меню; панель инструментов; рабочая область и строка состояния.

Строка заголовка содержит пиктограмму, название программы «STATISTICA» и три кнопки управления размерами основного окна: кнопка минимизации размеров окна; кнопка восстановления размеров окна; кнопка закрытия окна.

Строка меню занимает вторую строчку основного окна модуля и при наличии открытого файла с данными в рабочей области содержит так называемые выпадающие меню: **File-Файл, Edit-Правка, View-Вид, Insert-Вставка, Format-Формат, Statistics-Статистики, Graphs-Графики, Tools-Инструменты, Data-Данные, Window-Окно, Help-Справка.**

Панель инструментов содержит кнопки для быстрого доступа к чаще всего используемым командам меню.

Рабочая область занимает большую часть основного окна STATISTICA. В нее выводятся различные документы:

- электронная таблица с исходными данными. При первом открытии STATISTICA в рабочей области автоматически открывается новый файл размерности 10×10 (10 столбцов и строк) с названием Spreadsheet.sta;
- стартовое окно используемого модуля статистического анализа;
- электронные таблицы с результатами анализа;
- различные графики, визуализирующие либо исходные данные, либо результаты анализа;
- окно автоотчета.

Строка состояния расположена в самой нижней части окна системы STATISTICA. В зависимости от состояния, в котором находится система, *строка состояния* содержит кнопку быстрого доступа к основным статистическим модулям и пунктам меню, а также отражает различную информацию и позволяет управлять функционированием системы.

При обработке данных и построении графиков *Строка состояния* содержит шкалу прогресса, которая отражает степень завершенности процесса обработки данных и таймер, отражающий время, прошедшее с начала обработки.

2. Создание, редактирование и сохранение файла с исходными данными.

В качестве примера рассмотрим данные (таблица 1.1) о цене одной бутылки *марочного вина* в зависимости от *года его закладки*. Цены указаны

в долларах на одну бутылку. Таким образом, в качестве статистической совокупности в рассматриваемом случае выступает совокупность марочных вин, а статистическими признаками, зарегистрированными при проведении исследований, являются «цена одной бутылки вина» и «год закладки вина».

Таблица 1.1

Цены на марочное вино в зависимости от года закладки

| Год закладки | Цена, S | Год закладки | Цена, S |
|--------------|---------|--------------|---------|
| 1890 | 50.00 | 1941 | 10.00 |
| 1900 | 35.00 | 1944 | 5.99 |
| 1920 | 25.00 | 1948 | 8.98 |
| 1931 | 11.98 | 1950 | 6.98 |
| 1934 | 15.00 | 1952 | 4.99 |
| 1935 | 13.00 | 1955 | 5.98 |
| 1940 | 6.98 | 1960 | 4.98 |

Прежде чем начать создавать файл с данными, в папке «Мои документы» создайте папку под своей фамилией, где будут сохраняться как создаваемые вами файлы с данными, так и файлы отчета о проделанной работе. После этого в пункте меню **File-Файл** системы STATISTICA выберите команду «New-новый». В результате будет открыто окно создания нового файла (рис. 1.1).

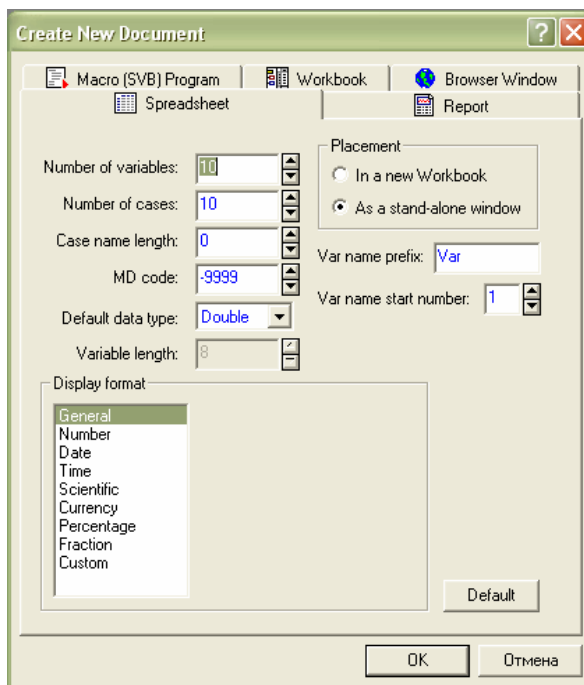
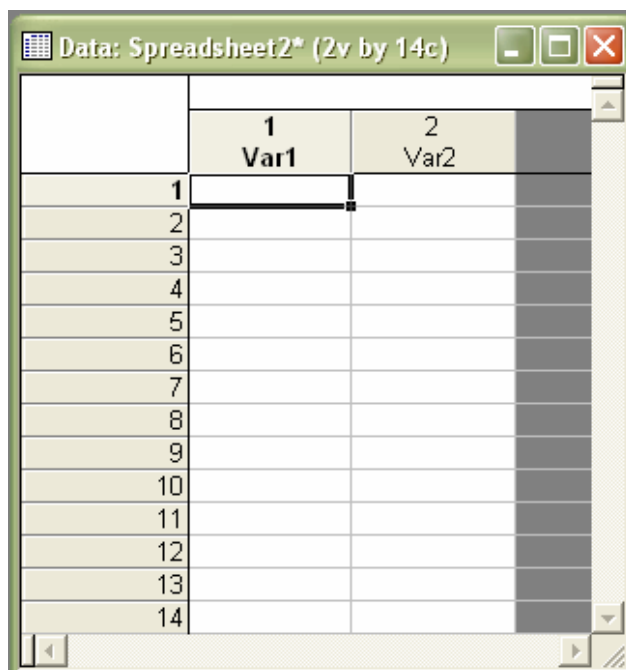


Рис. 1.1. Стартовое окно создания нового файла с данными

Нам необходимо создать электронную таблицу из двух переменных (столбцов) и 14 случаев (строк) и ввести в нее числовые значения. В поле «Number of variables» (количество переменных или количество статистических признаков) введите число 2, а в поле «Number of cases» (количество случаев или количество наблюдений) введите число 14. Остальные поля и пункты оставьте без изменений и нажмите «ОК». В результате в рабочей области будет создана электронная таблица размерности 2×14 (рис. 1.2).



| | 1 Var1 | 2 Var2 |
|----|-----------|-----------|
| 1 | | |
| 2 | | |
| 3 | | |
| 4 | | |
| 5 | | |
| 6 | | |
| 7 | | |
| 8 | | |
| 9 | | |
| 10 | | |
| 11 | | |
| 12 | | |
| 13 | | |
| 14 | | |

Рис. 1.2. Электронная таблица размерности 2×14

Сделайте двойной щелчок на поле «**Var1**». В открывшемся окне заполните поля так, как показано на рисунке 1.3, а. После заполнения нажмите «ОК». Дважды «кликните» на поле «**Var2**» и заполните поля окна спецификаций так, как показано на рисунке 1.3, б. Выполнив эти действия, вы указали: имя переменных (Name), формат данных (Display format) – числовой (Number) и количество знаков после запятой, отражаемых в значениях переменных (Decimals places). После этого можно вводить данные в электронную таблицу в соответствии с тем, как они представлены в таблице 1.1.

Variable 1

Name: Год закладки Type: Double OK

Measurement Type: Auto Length: 8 Cancel

☐ Excluded ☐ Label ☐ Case State MD code: -9999 << >>

Display format

| | |
|------------|-------------------|
| General | Decimal places: 0 |
| Number | 1000; -1000 |
| Date | 1 000; -1 000 |
| Time | 1000; (1000) |
| Scientific | 1 000; (1 000) |
| Currency | |
| Percentage | |
| Fraction | |
| Custom | |

All Specs... Text Labels... Values/Stats... Properties...

Long name (label or formula with Functions): ☒ Function guide

Labels: use any text. Formulas: use variable names or v1, v2, ..., v0 is case #.
Examples: (a) = mean(v1:v3, sqrt(v7), AGE) (b) = v1+v2; comment (after:)

a)

Variable 2

Name: Цена Type: Double OK

Measurement Type: Auto Length: 8 Cancel

☐ Excluded ☐ Label ☐ Case State MD code: -9999 << >>

Display format

| | |
|------------|----------------------|
| General | Decimal places: 2 |
| Number | 1000.00; -1000.00 |
| Date | 1 000.00; -1 000.00 |
| Time | 1000.00; (1000.00) |
| Scientific | 1 000.00; (1 000.00) |
| Currency | |
| Percentage | |
| Fraction | |
| Custom | |

All Specs... Text Labels... Values/Stats... Properties...

Long name (label or formula with Functions): ☒ Function guide

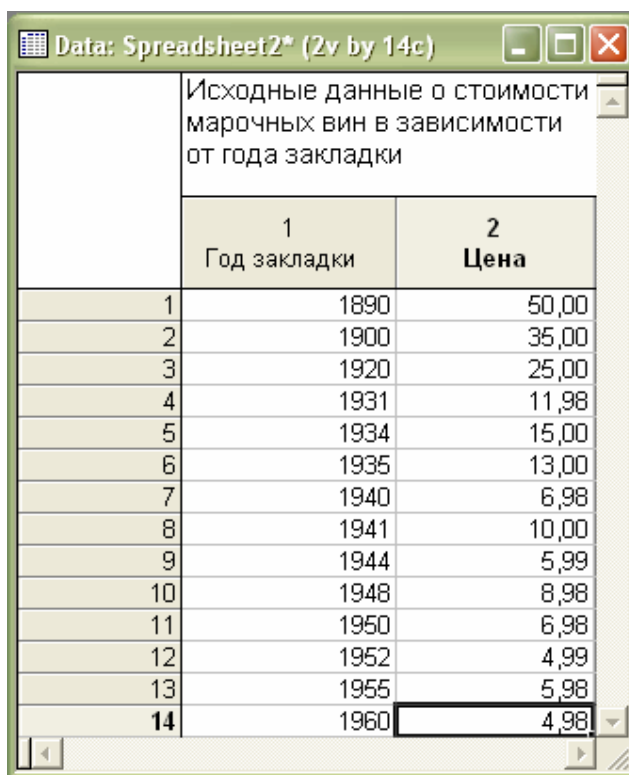
Labels: use any text. Formulas: use variable names or v1, v2, ..., v0 is case #.
Examples: (a) = mean(v1:v3, sqrt(v7), AGE) (b) = v1+v2; comment (after:)

б)

Рис. 1.3. Окно задания спецификаций переменных «Год закладки» - а), «Цена» - б)


После ввода данных в поля таблицы подведите курсор к белому полю над переменными «Год закладки» и «Цена», два раза кликните мышкой по пустой строке и введите текст следующего содержания: «Исходные данные о стоимости марочных вин в зависимости от года закладки».

После выполнения указанных действий электронная таблица будет выглядеть так, как показано на рисунке 1.4.



| | 1 | 2 |
|----|--------------|-------|
| | Год закладки | Цена |
| 1 | 1890 | 50,00 |
| 2 | 1900 | 35,00 |
| 3 | 1920 | 25,00 |
| 4 | 1931 | 11,98 |
| 5 | 1934 | 15,00 |
| 6 | 1935 | 13,00 |
| 7 | 1940 | 6,98 |
| 8 | 1941 | 10,00 |
| 9 | 1944 | 5,99 |
| 10 | 1948 | 8,98 |
| 11 | 1950 | 6,98 |
| 12 | 1952 | 4,99 |
| 13 | 1955 | 5,98 |
| 14 | 1960 | 4,98 |

Рис. 1.4. Электронная таблица с введенными исходными данными

В пункте меню **File-Файл** выберите команду **Save** и сохраните файл в созданную папку под любым подходящим именем (например, Vine.sta). После сохранения файла изменится заголовок электронной таблицы с данными и примет следующий вид: .

В системе STATISTICA структуру электронной таблицы с данными можно менять, выполняя следующие операции:

- добавляя или убавляя переменные и случаи наблюдений;
- транспонируя (меняя местами) строки и столбцы;

- сортируя данные по возрастанию или убыванию
- и.т.д.

Указанные действия (и еще множество других операций) можно выполнить, выбрав в меню **Data-данные** соответствующий пункт. Однако наиболее часто выполняемые операции – это добавление и удаление переменных (признаков) и случаев (наблюдений). Выполним, например, добавление переменных к таблице с данными. В меню **Data-данные** выберите Variables – Add (добавить). В результате будет открыто окно добавления переменных (рис. 1.5).

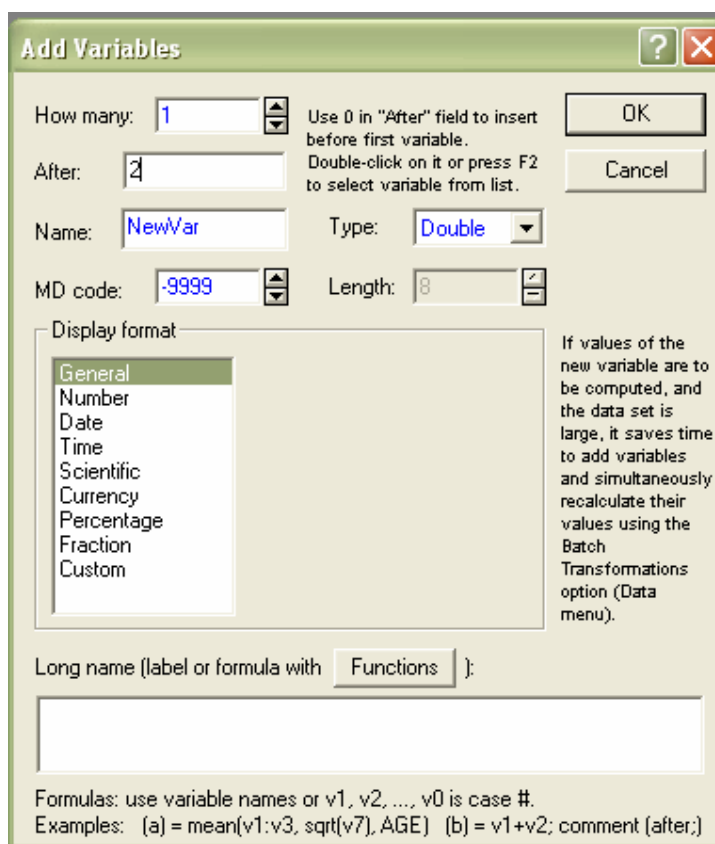


Рис. 1.5. Окно добавления переменных

В этом окне в поле «How many» (сколько) укажите 1, а в поле After (После) укажите 2. Все остальное оставьте без изменений и нажмите «OK». Вы увидите, что в электронной таблице добавлена одна колонка с именем переменной «NewVar» (Новая переменная). Для удаления переменной в

меню **Data-данные** выберите Variables – Delete (удалить). В открывшемся окне установите значения полей так, как показано на рисунке 1.6.

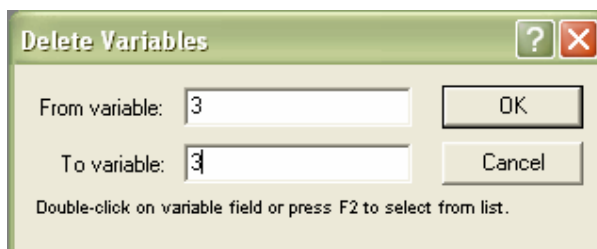



Рис. 1.6. Окно удаления переменных

Выполнив указанные действия, сообщаем системе, что необходимо удалить переменные, начиная с номера 3 и заканчивая номером 3. Нажмите «ОК». Вы увидите, что переменная с номером 3 удалена из электронной таблицы. Аналогичным образом можно добавлять и удалять случаи (наблюдения). Попробуйте сделать это самостоятельно.

3. Создание и редактирование графиков. В системе STATISTICA могут быть построены самые разнообразные статистические графики: гистограммы, точечные и линейные графики, двумерные и трехмерные и т.д.

Построим, например, точечный график, отражающий зависимость между годом закладки вина и его ценой. Для этого в меню **Graphs-графики** выберите Scatterplots (точечные графики). Откроется окно установки параметров точечного графика (рис. 1.7).

Нажмите на кнопку  **Variables:** и установите в качестве переменной x – год закладки вина, а в качестве переменной y - его цену. Снимите флажок с поля *Fit type* и нажмите «ОК». В результате будет построен график так, как показано на рисунке 1.8.

Отредактируем построенный график. «Кликните» дважды на названии графика и введите название графика так, как показано на рисунке 1.9. Точно так же отредактируйте названия осей.

Аналогичным образом могут быть построены и отредактированы все виды статистических графиков.

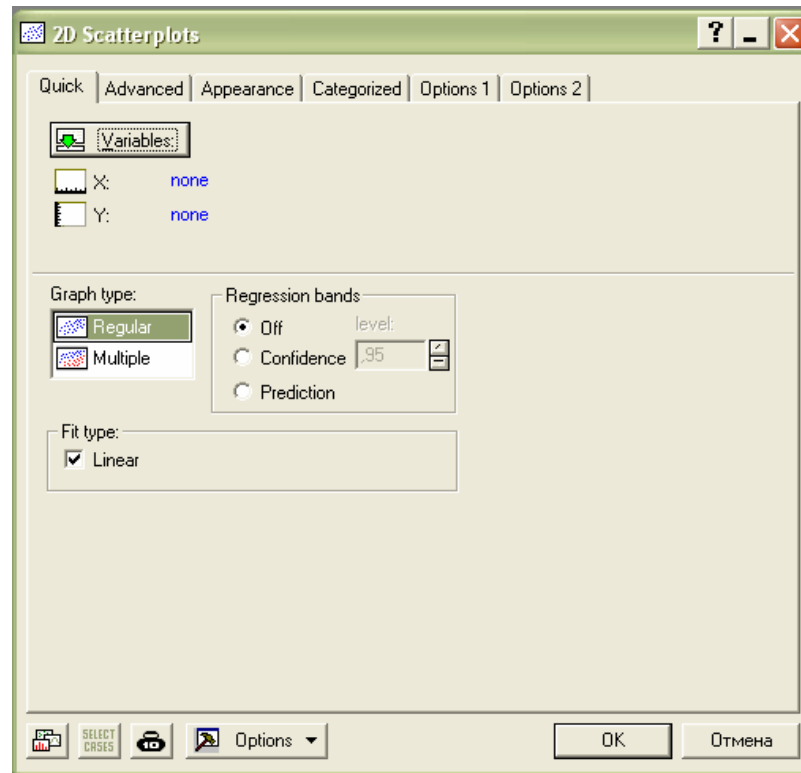


Рис. 1.7. Окно установки параметров построения точечного графика

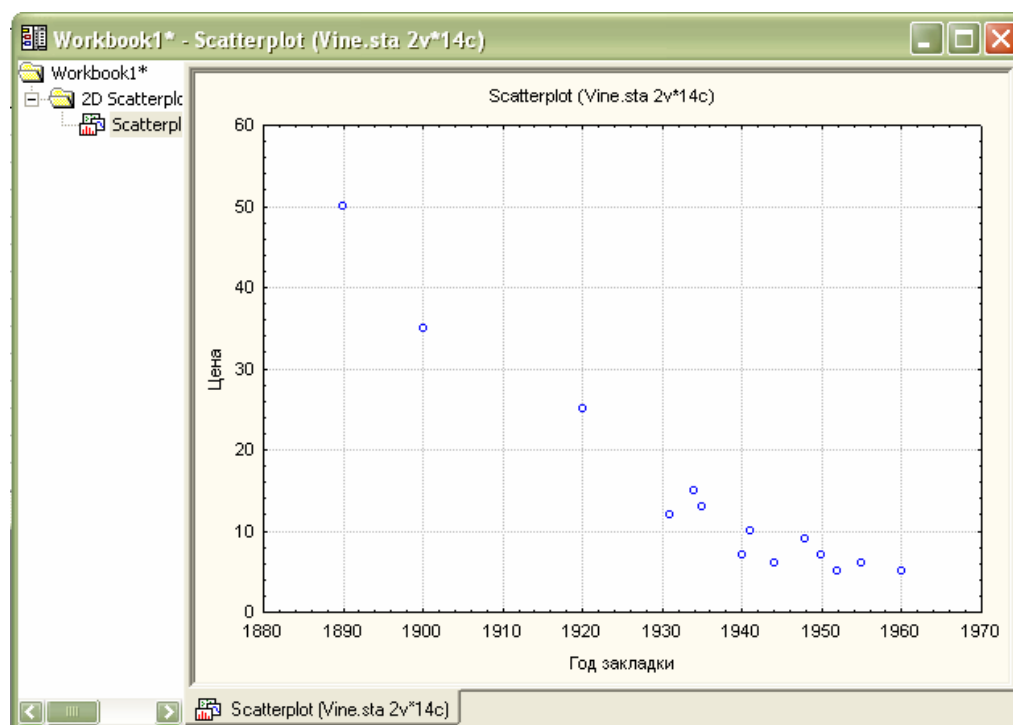


Рис. 1.8. Точечный статистический график

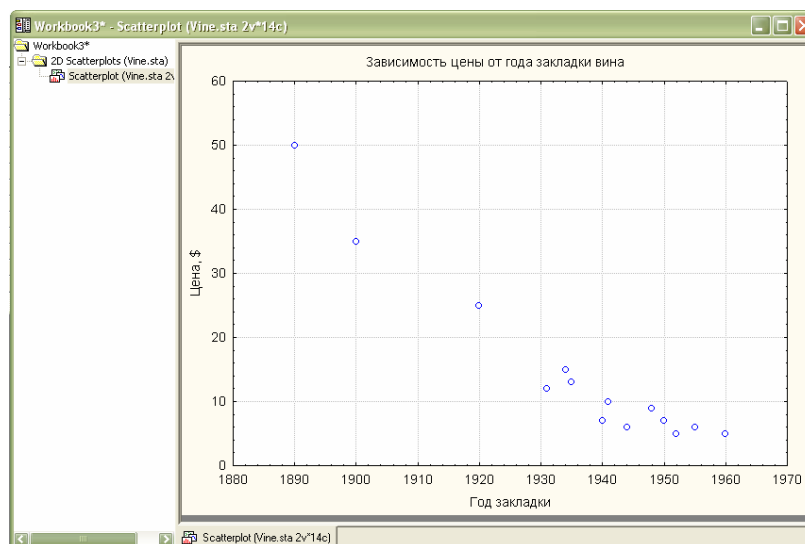


Рис. 1.9. Отредактированный точечный статистический график

3. Создание автоотчета. Результаты всех выполняемых в системе STATISTICA процедур создания электронной таблицы с данными статистического анализа, построения графиков могут быть перенесены в так называемый автоотчет. Для создания файла отчета необходимо на панели инструментов найти пункт **Add to Report**, выбрать **New Report (новый отчет)**. В результате будет открыто окно с файлом отчета, в который автоматически добавится электронная таблица с исходными данными (рис. 1.10).

Report1*

Contents

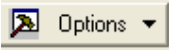
Vine.sta

Исходные данные о стоимости марочных вин в зависимости от года закладки

| | 1 Год закладки | 2 Цена |
|----|-------------------|-----------|
| 1 | 1890 | 50,00 |
| 2 | 1900 | 35,00 |
| 3 | 1920 | 25,00 |
| 4 | 1931 | 11,98 |
| 5 | 1934 | 15,00 |
| 6 | 1935 | 13,00 |
| 7 | 1940 | 6,98 |
| 8 | 1941 | 10,00 |
| 9 | 1944 | 5,99 |
| 10 | 1948 | 8,98 |
| 11 | 1950 | 6,98 |
| 12 | 1952 | 4,99 |
| 13 | 1955 | 5,98 |
| 14 | 1960 | 4,98 |

Рис. 1.10. Окно файла отчета

Файл отчета является обычным текстовым файлом, поэтому он может редактироваться теми же способами, что и другие текстовые файлы. По умолчанию файл отчета называется Report1. После создания сохраните файл отчета под любым удобным вам названием в созданной ранее папке.

Для автоматического добавления результатов работы в файл отчета необходимо выполнять следующую последовательность действий. При открытии начального окна статистического анализа или окна построения графиков необходимо найти и нажать на этом окне кнопку  и выбрать Output (Вывод). В результате будет открыто окно опций вывода результатов (рис. 1.11).

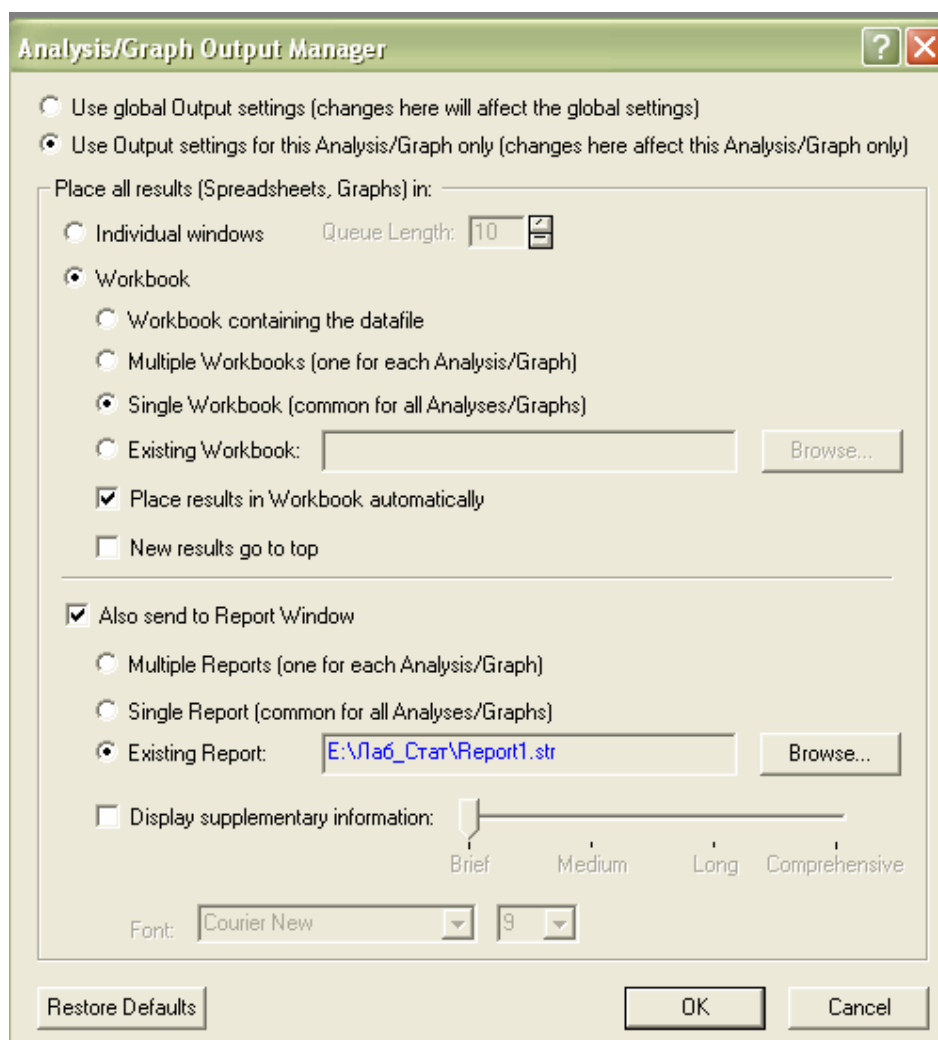


Рис. 1.11. Окно установки параметров вывода результатов

Установите опции вывода так, как показано на рисунке 1.11. (Примечание. В поле Existing Report (Существующий отчет) необходимо указать путь к созданному вами файлу отчета). После установки указанных опций все получаемые результаты, построенные графики будут автоматически добавляться в файл отчета.

Задание для самостоятельной работы. Предположим, что производится опрос группы населения, в результате которого фиксируются следующие статистические признаки: *Пол, Возраст, Образование, Доход*. Результаты опроса конкретной группы представлены в таблице 1.2.

Таблица 1.2

Результаты статистического опроса

| <i>Имена опрошенных (Названия случаев)</i> | <i>Пол</i> | <i>Возраст, лет</i> | <i>Образование</i> | <i>Доход, \$</i> |
|--|------------|---------------------|--------------------|------------------|
| Петров П. | Муж | 21 | Среднее | 125 |
| Комарова И. | Жен | 23 | Высшее | 345 |
| Статистов М. | Муж | 40 | Нет | 50 |
| Файлова Р. | Жен | 35 | Среднее | 399 |
| Ивлева Т. | Жен | 18 | Среднее | 136 |
| Часов И. | Муж | 20 | Высшее | 120 |
| Сидоров С. | Муж | 45 | Среднее | 200 |
| Медведев Д. | Муж | 38 | Высше | 300 |
| Сухнова Г. | Жен | 23 | Среднее | 125 |
| Лапаева О. | Жен | 55 | Высшее | 400 |

Создайте и заполните электронную таблицу, структура и содержание которой соответствует таблице 2. В поле ввода комментария к файлу введите «Социологический опрос. Данные собраны в 20.. году в г. Рубцовске Алтайского края. Файл создал». После заполнения таблицы сохраните файл под именем «Socio.sta». Примечание: для задания имен случаев достаточно дважды щелкнуть на номере случая в электронной таблице и ввести требуемые данные (в нашем случае фамилию и имя опрашиваемого).

Добавьте заполненную электронную таблицу в файл отчета. Сохраните файл отчета. Постройте точечный график, отражающий связь между возрастом (переменная x) и уровнем дохода (переменная y). Отредактируйте и добавьте график в файл отчета.

Оформление отчета по лабораторной работе.

Отчет по лабораторной работе №1 должен содержать:

- электронную таблицу с исходными данными из таблицы 1.1 и точечный график (рисунок 1.9);
- электронную таблицу с исходными данными из таблицы 1.2 и точечный график, отражающий связь между возрастом и уровнем дохода респондентов.

2. Лабораторная работа №2

Тема работы: Ряды распределения (4 часа).

Цель работы: Освоение основных приемов анализа рядов распределений.

Содержание работы:

1. Генерация данных.
2. Визуализация распределения (построение гистограммы частот и графика накопленных частот).
3. Расчет описательных статистик распределения.
4. Оформление отчета по лабораторной работе.

Выполнение работы.

1. Генерация данных.

Запустите систему “STATISTICA”. Создайте файл с данными размерности 5×100 (5 переменных 100 наблюдений). Переменным присвойте имена и характеристики в соответствии с таблицей 2.1.

Таблица 2.1

Имена и характеристики переменных

| Переменная | Имя | Формат отображения | Тип | Количество десятичных знаков | Вводимая строка в поле «Длинное имя или формула» |
|------------|-----------------------------|--------------------|-------------------------|------------------------------|--|
| Var1 | Равномерное распределение | Числовой | Целочисленный (Integer) | 0 | =Rnd(5) |
| Var2 | Нормальное распределение | Числовой | Double | 1 | =VNormal(rnd(1);2;0,5) |
| Var3 | Показательное распределение | Числовой | Double | 1 | =VExpon(rnd(1); 0,2) |
| Var4 | Распределение Релея | Числовой | Double | 1 | =VRayl(rnd(1);0,5) |
| Var5 | Распределение Пуассона | Числовой | Целочисленный (Integer) | 0 | =RndPoisson(5) |

После нажатия кнопки «ОК» введенные характеристики позволяют сгенерировать данные, распределенные по некоторым теоретическим законам распределения случайных величин (в частности, по равномерному,

нормальному, показательному законам, а также по законам Релея и Пуассона).

2. Визуализация распределения (построение гистограммы частот и графика накопленных частот)

Дальнейшей задачей является визуализация полученных данных, т.е. их изображение в виде гистограммы частот и гистограммы накопленных частот.

В пункте меню «Graphs (Графы)» выберите пункт «Гистограммы». В результате выбора откроется окно выбора параметров построения гистограммы, представленное на рисунке 2.1.

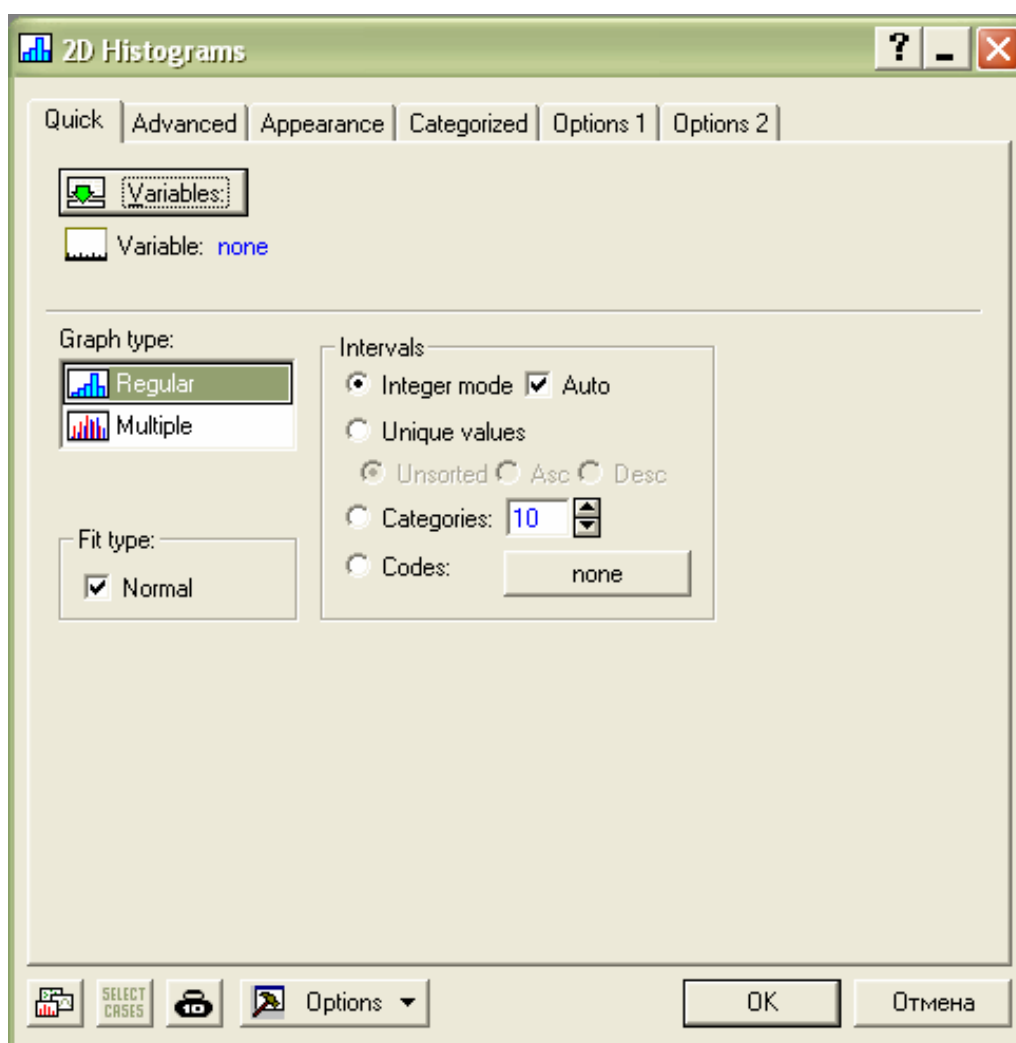


Рис. 2.1. Окно выбора параметров построения гистограммы (вкладка «Quick»)

После нажатия кнопки «Variables (Переменная)» в окне выбора переменных выберите переменную «Равномерное распределение» и после нажатия кнопки «ОК» установите следующие параметры на вкладке «Quick (Быстрое)» так же, как показано на рисунке 2.1. Снимите флажок с пункта «Fit type (подбираемый тип)». Перейдите к вкладке «Advanced (Расширенное)». На этой вкладке установите параметры так, как показано на рисунке 2.2.

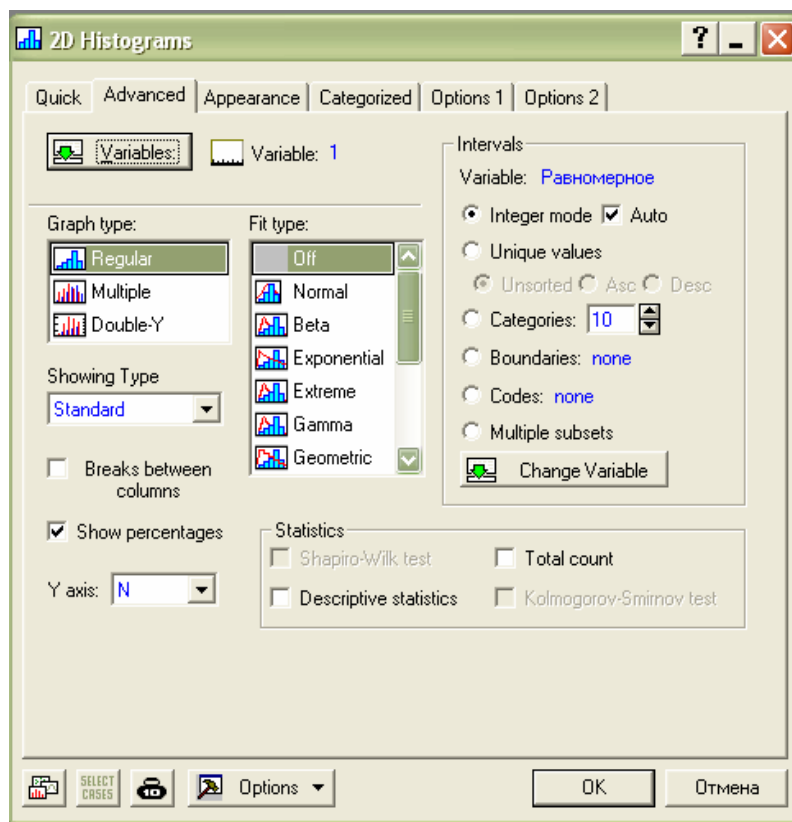
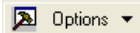


Рис. 2.2. Окно выбора параметров построения гистограммы (вкладка «Advanced» (Расширенное))

После установления параметров на вкладках «Quick (Быстрое)» и «Advanced (Расширенное)» нажмите кнопку «ОК». В результате расчетов будет произведена автоматическая группировка данных и построена гистограмма частот (см. рис. 2.3). **Не забывайте все построения добавлять в отчет по лабораторной работе путем выбора соответствующих параметров после нажатия кнопки**  **!**

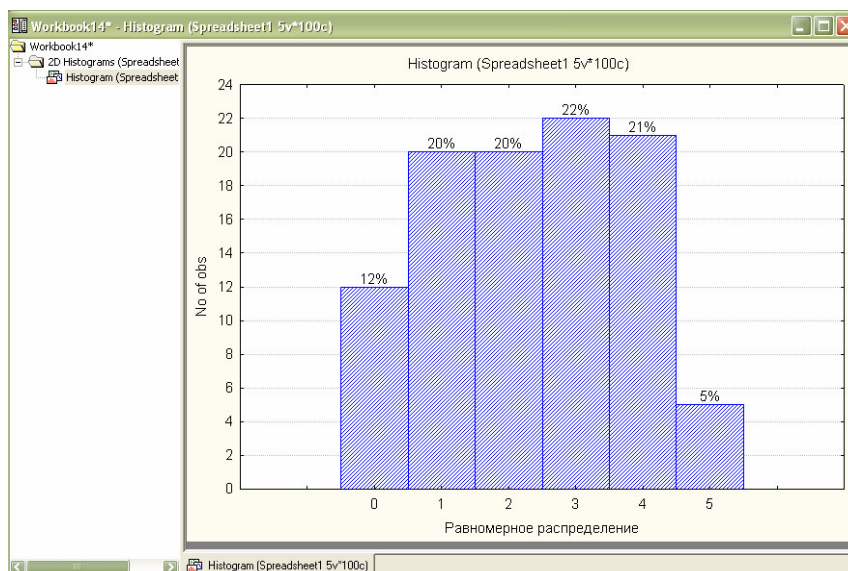


Рис. 2.3. Гистограмма распределения признака с равномерным законом распределения

Измените на вкладке «Advanced (Расширенное)» вид отображения в окне «Showing type» так, как показано на рисунке 2.4. В окне должно отображаться «Cumulative (Накопленное)».

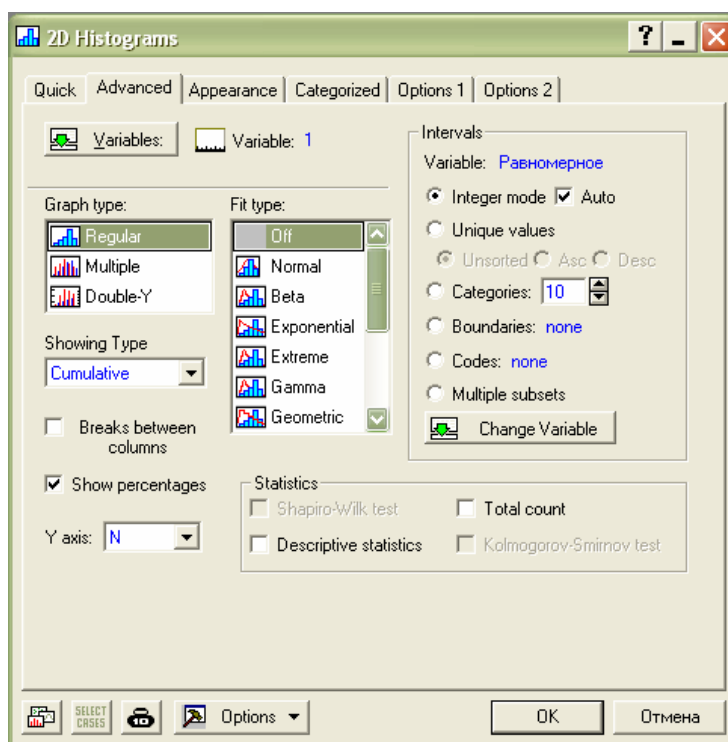


Рис. 2.4. Окно выбора параметров построения гистограммы (вкладка «Advanced» (Расширенное))

После изменения типа отображения нажмите «ОК». В результате будет построен график накопленных частот так, как показано на рисунке 2.5.

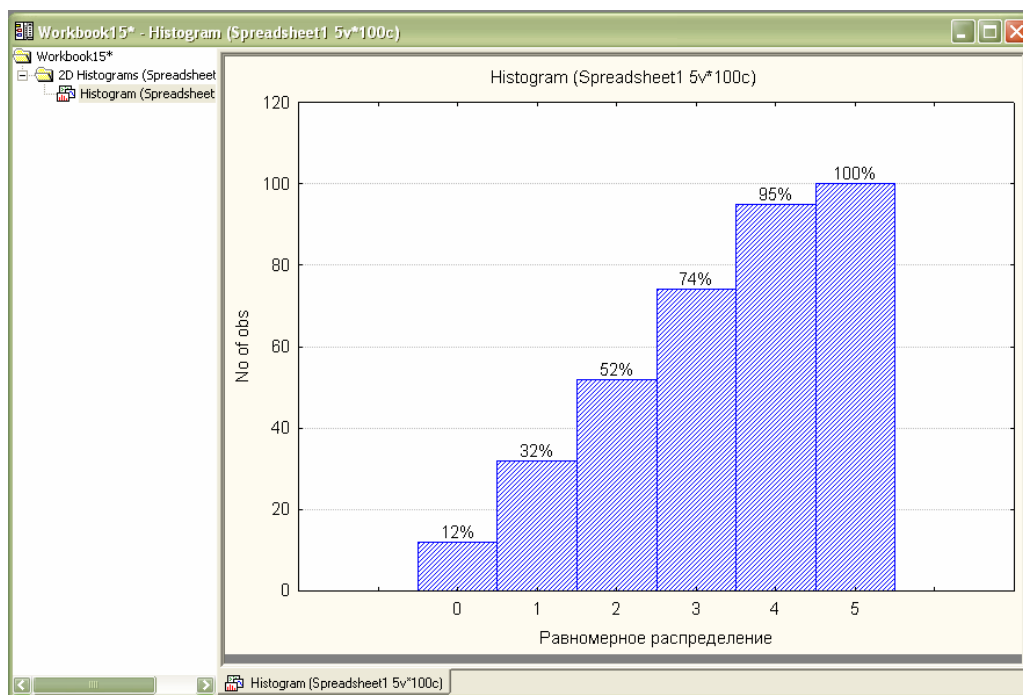


Рис. 2.5. График накопленных частот распределения признака с равномерным законом распределения

3. Расчет описательных статистик распределения.

После визуализации данных необходимо осуществить расчет основных числовых характеристик полученных распределений. Для решения данной задачи закройте окно построения гистограмм. В меню «Statistics» выберите пункт «Basic statistics/Tables (Основные статистики/Таблицы)». В открывшемся окне (рис. 2.6) выберите пункт «Descriptive statistics» (Описательные статистики). В результате будет открыто окно расчета комплекса описательных статистик (рис. 2.7).

Перейдите от вкладки «Quick (быстрое)» к вкладке «Advanced» (расширенное). На данной вкладке отметьте рассчитываемые показатели так, как показано на рисунке 2.8.

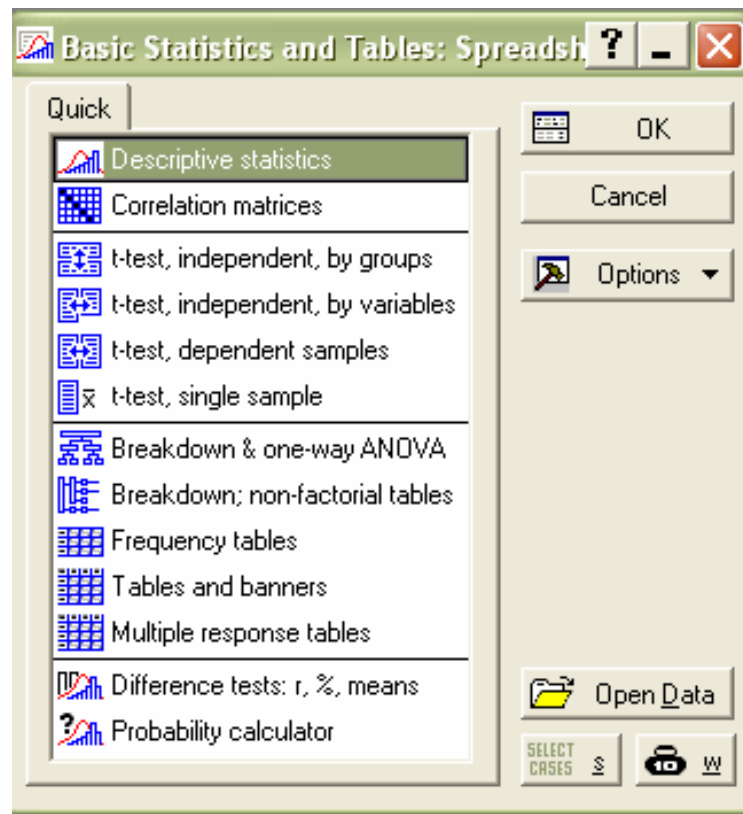


Рис. 2.6. Начальное окно расчета описательных статистик

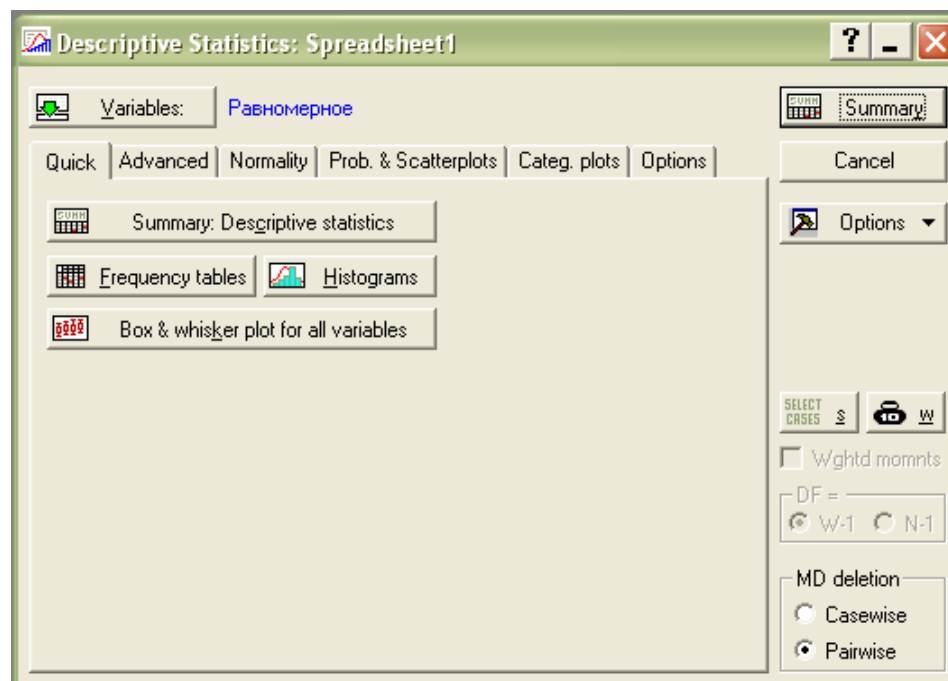


Рис. 2.7. Окно расчета описательных статистик. Вкладка «Quick (быстрое)»

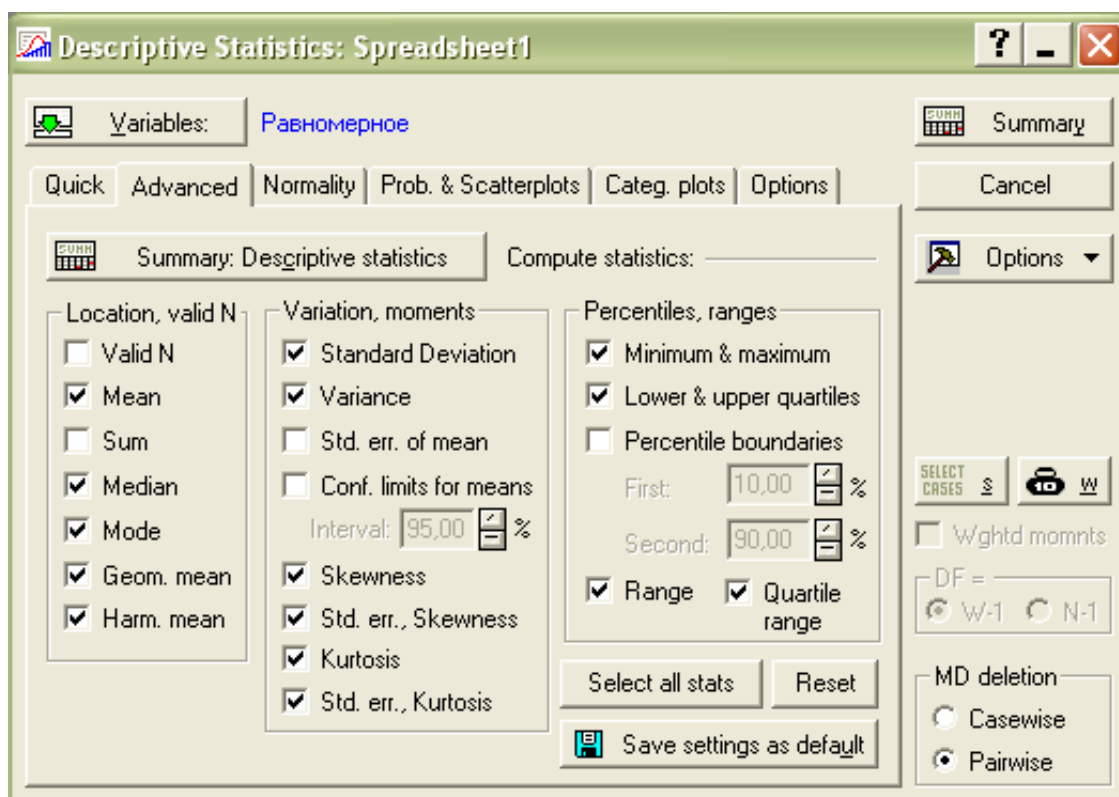


Рис. 2.8. Окно расчета описательных статистик. Вкладка «Advanced (расширенное)»

Установив флажки так, как показано на рисунке 2.8, вы выбираете расчет следующих описательных статистик: Mean – среднее арифметическое; Median - медиана; Mode - мода; Geom. mean – геометрическое среднее; Harm. mean – гармоническое среднее; Standard Deviation – среднее квадратическое отклонение; Variance - дисперсия; Skewness - асимметрия; Std. err. Skewness – стандартная ошибка асимметрии; Kurtosis - эксцесс; Std. err. Kurtosis – стандартная ошибка эксцесса; Minimum – минимальное значение; Maximum – максимальное значение; Lower & upper quartiles – первая и третья квартиль соответственно; Range – размах вариации; Quartile range – квартильная вариация.

Перейдите на вкладку «Normality (нормализация)» (рис. 2.9) и установите опции так, как показано на рисунке 2.9.

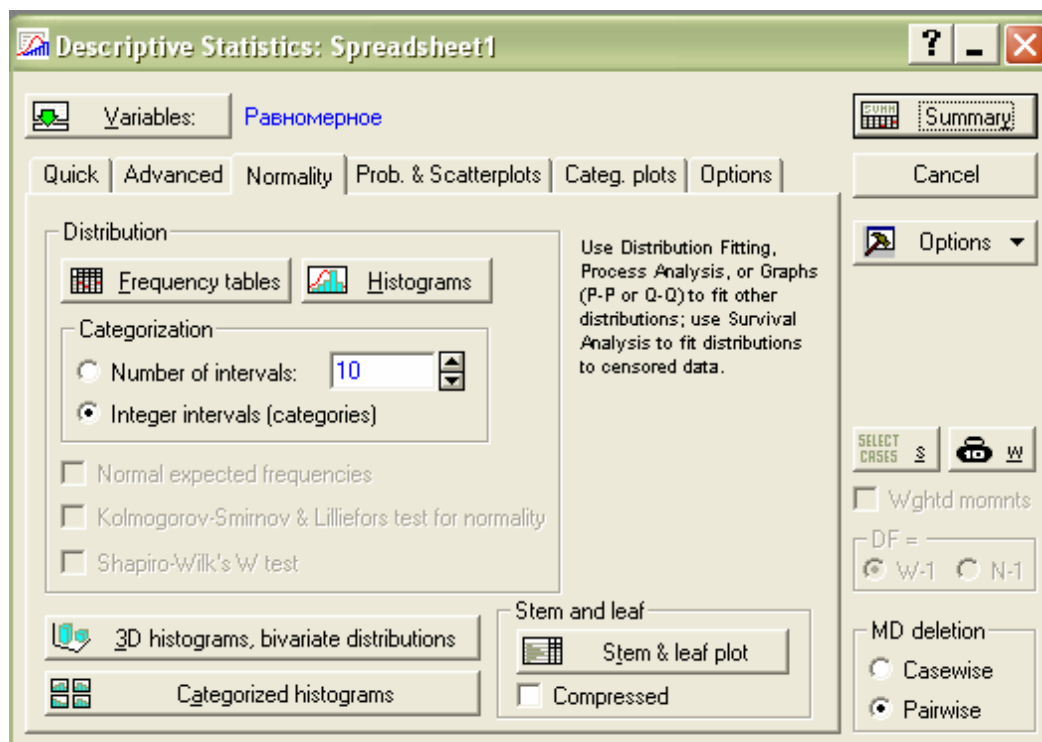





Рис. 2.9. Окно расчета описательных статистик. Вкладка «Normality (нормализация)»

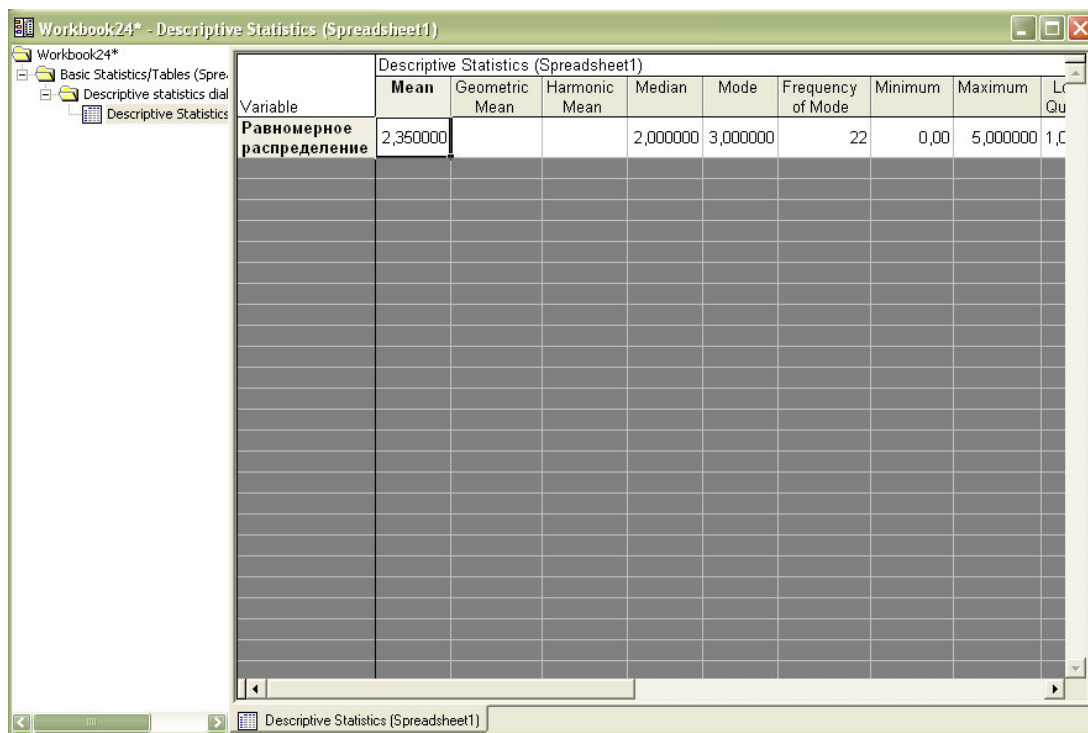
Установленные таким образом опции позволяют еще раз посмотреть ряд распределения, нажав на кнопку  **Frequency tables**, и гистограмму распределения, нажав на кнопку  **Histograms**. В частности, при просмотре ряда распределения будет построена таблица, представленная на рисунке 2.10.

Workbook21* - Frequency table: Равномерное распределение: =Rnd(5) (Spreadsheet1)

| Category | Count | Cumulative Count | Percent of Valid | Cumul % of Valid | % of all Cases | Cumulative % of All |
|----------|-------|------------------|------------------|------------------|----------------|---------------------|
| 0 | 12 | 12 | 12,00000 | 12,0000 | 12,00000 | 12,0000 |
| 1 | 20 | 32 | 20,00000 | 32,0000 | 20,00000 | 32,0000 |
| 2 | 20 | 52 | 20,00000 | 52,0000 | 20,00000 | 52,0000 |
| 3 | 22 | 74 | 22,00000 | 74,0000 | 22,00000 | 74,0000 |
| 4 | 21 | 95 | 21,00000 | 95,0000 | 21,00000 | 95,0000 |
| 5 | 5 | 100 | 5,00000 | 100,0000 | 5,00000 | 100,0000 |
| Missing | 0 | 100 | 0,00000 | | 0,00000 | 100,0000 |

Рис. 2.10. Ряд распределения признака с равномерным законом распределения

После просмотра ряда распределения нажмите кнопку . Система осуществит расчет отмеченных показателей и представит результаты в виде таблицы (рис. 2.11).



| Variable | Mean | Geometric Mean | Harmonic Mean | Median | Mode | Frequency of Mode | Minimum | Maximum | L | Qu |
|---------------------------|----------|----------------|---------------|----------|----------|-------------------|---------|----------|-----|----|
| Равномерное распределение | 2,350000 | | | 2,000000 | 3,000000 | 22 | 0,00 | 5,000000 | 1,0 | |

Рис. 2.11. Окно результатов расчета описательных статистик

Просмотреть все рассчитанные статистики можно, управляя полосой прокрутки внизу окна результатов.

4. Оформление отчета по лабораторной работе.

Отчет по лабораторной работе №2 должен содержать:

- постановку задачи;
- результаты построения гистограммы распределения, график накопленных частот, таблицу с построенным рядом распределения и таблицу, содержащую результаты расчета основных описательных статистик.

3. Лабораторная работа №3

Тема работы: Законы распределения случайных величин (2 часа).

Цель работы: Освоение приемов работы с калькулятором распределений.

Содержание работы:

1. Исследование функций плотности распределения и функций распределения законов: Гаусса (нормального), экспоненциального (показательного), Релея, Стьюдента, Фишера и χ^2 (хи-квадрат). Вычисление односторонних и двусторонних критических границ распределений (квантилей распределений), соответствующих заданному значению вероятности.

2. Оформление отчета по лабораторной работе.

Краткая теоретическая справка.

В теории вероятностей и теории статистики при решении ряда задач, связанных с проверкой статистических гипотез, дисперсионным, корреляционным, регрессионным и другими видами статистического анализа, приходится иметь дело с различными законами распределения случайных величин (СВ), а также уметь вычислять критические границы (квантили) распределений, соответствующие определенному уровню вероятности.

Напомним, что в теории вероятностей известно большое количество законов распределений (некоторые из них упомянуты в п.1 содержания работы) случайных величин. В общем случае закон распределения непрерывной СВ может быть задан функцией плотности вероятности, определенной либо на всей числовой оси (например, нормальный закон распределения или закон распределения Релея и т.д.), либо на некоторой ее части (закон равной вероятности, закон Симпсона и т.д.).

Пусть, например, непрерывная СВ X определена на всей числовой прямой. Тогда для функции плотности вероятности (обозначается $\varphi(x)$) и

интегральной функции распределения (обозначается $F(x)$) справедливы соотношения:

$$\int_{-\infty}^{+\infty} \varphi(x) dx = 1, \quad (3.1)$$

$$F(x) = \int_{-\infty}^x \varphi(x) dx. \quad (3.2)$$

Соотношения (3.2) позволяют ввести такое понятие, как односторонние или двусторонние критические границы (квантили) распределений. Дадим определение данных понятий. Левосторонней критической границей \underline{K}_α , отвечающей вероятности α , называется такое значение случайной величины X , подчиняющейся закону распределения с функцией плотности вероятности $\varphi(x)$, для которого:

$$P(X \leq \underline{K}_\alpha) = F(\underline{K}_\alpha) = \int_{-\infty}^{\underline{K}_\alpha} \varphi(x) dx = \alpha. \quad (3.3)$$

Графическое изображение левосторонней критической границы (на примере стандартного нормального закона) показано на рисунке 3.1.

Правосторонней критической границей \overline{K}_α , отвечающей вероятности α , называется такое значение аргумента функции плотности вероятности $\varphi(x)$, для которого:

$$P(X > \overline{K}_\alpha) = 1 - F(\overline{K}_\alpha) = \int_{\overline{K}_\alpha}^{+\infty} \varphi(x) dx = \alpha. \quad (3.4)$$

Графическое изображение правосторонней критической границы (на примере стандартного нормального закона) показано на рисунке 3.2.

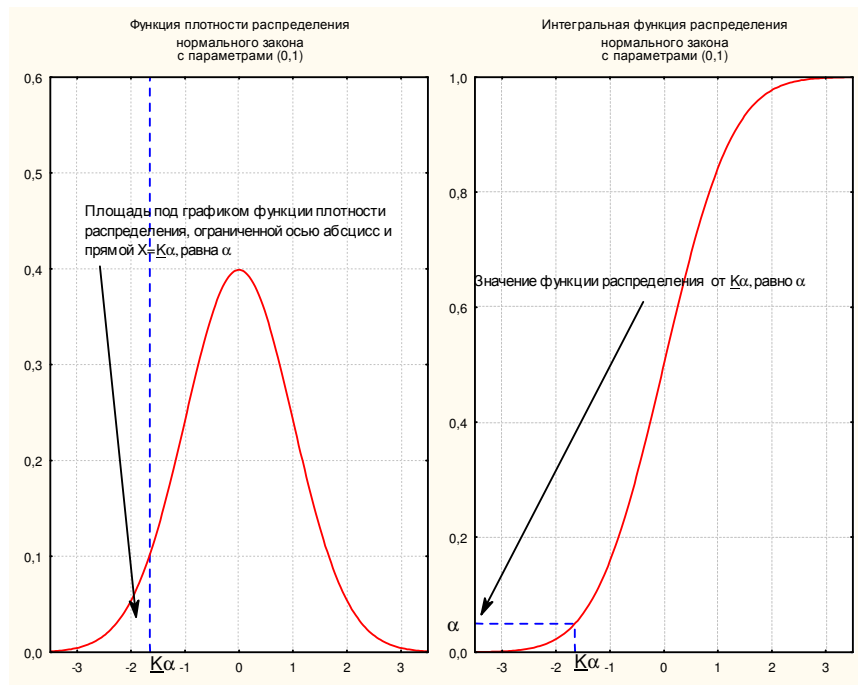


Рис. 3.1. Графическое изображение левосторонней критической границы (квантили), отвечающей вероятности α , на графике плотности вероятности и интегральной функции распределения

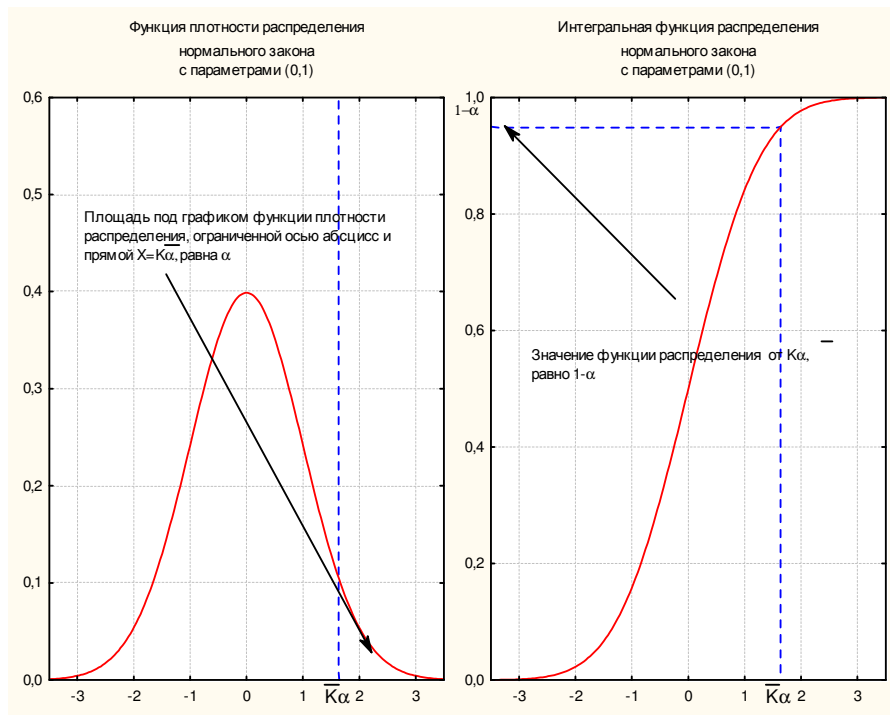


Рис. 3.2. Графическое изображение правосторонней критической границы (квантили), отвечающей вероятности α , на графике плотности вероятности и интегральной функции распределения

Двусторонними критическими границами \underline{B}_α и \bar{B}_α , отвечающими вероятности α , называются такие значения аргумента функции плотности вероятности $\varphi(x)$, для которых:

$$P(\underline{B}_\alpha \leq X < \bar{B}_\alpha) = \int_{\underline{B}_\alpha}^{\bar{B}_\alpha} \varphi(x) dx = 1 - \alpha. \quad (3.5)$$

Графическое изображение двусторонних критических границ и соответствующих им вероятностей приведено на рисунке 3.3.

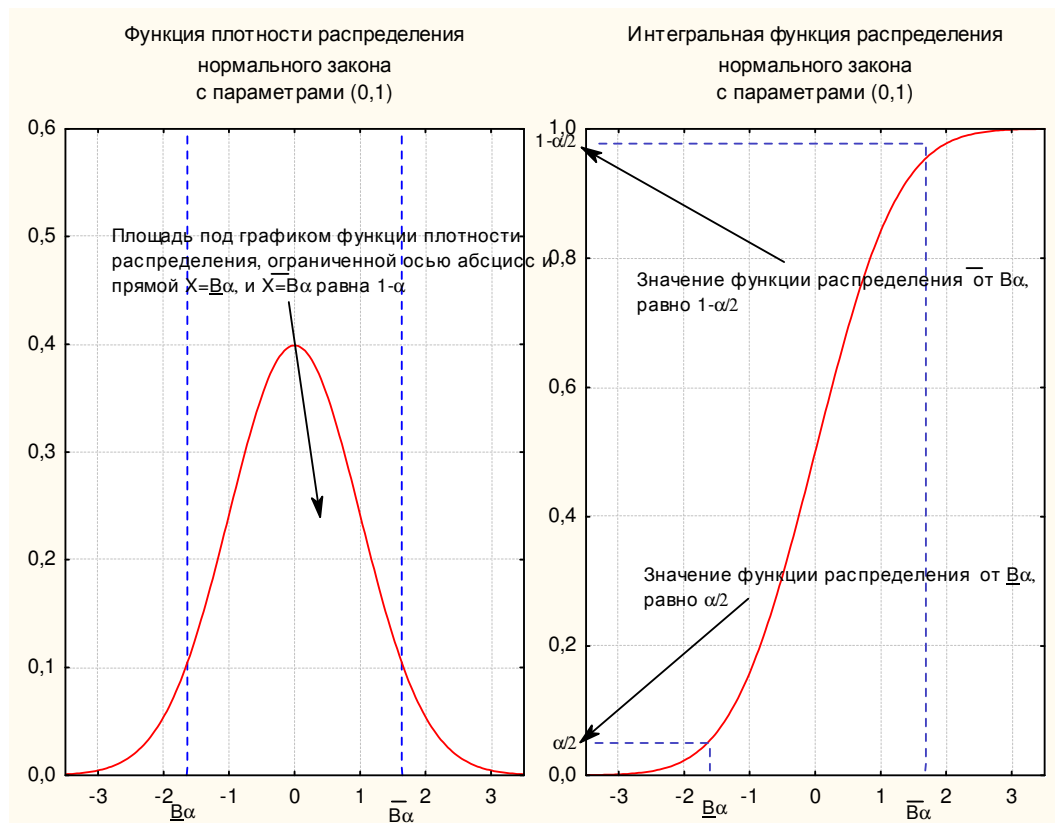


Рис. 3.3. Графическое изображение двусторонних критических границ (квантилей), отвечающих вероятности α , на графике плотности вероятности и интегральной функции распределения

Отметим, что двусторонние критические границы в зависимости от распределения могут быть как симметричными, так и несимметричными. Их значения будут зависеть не только от вида закона распределения, но и от его параметров.

В частности, для стандартного нормального распределения двусторонние критические границы симметричны относительно центра распределения и имеют специальное обозначение $\pm U_\alpha$. При этом значение двусторонних критических границ нормального распределения определяется из решения уравнения:

$$\Phi(U_\alpha) = \frac{1-\alpha}{2}. \quad (3.6)$$

Характер поведения законов распределения при изменении их параметров, а также нахождение правосторонних, левосторонних и двусторонних критических границ, соответствующих определенной вероятности для различных видов законов распределений, и будет составлять суть третьей лабораторной работы по курсу «Статистика».

Выполнение работы.

1. Исследование функций плотности распределения и функций распределения законов: Гаусса (нормального), экспоненциального (показательного), Релея, Стюдента, Фишера и χ^2 (хи-квадрат). Вычисление односторонних и двусторонних критических границ распределений (квантилей распределений), соответствующих заданному значению вероятности.

Запустить систему STATISTICA. В пункте меню Statistics (статистики) откройте пункт меню Probability Calculator (калькулятор распределений), затем пункт Distributions (распределения), как показано на рисунке 3.4.

В результате откроется окно калькулятора распределений (см. рис. 3.5), позволяющего решать задачи построения графиков функций плотности и интегральных функций для различных законов распределений СВ, а также находить правосторонние, левосторонние и двусторонние критические границы, отвечающие определенной вероятности α .

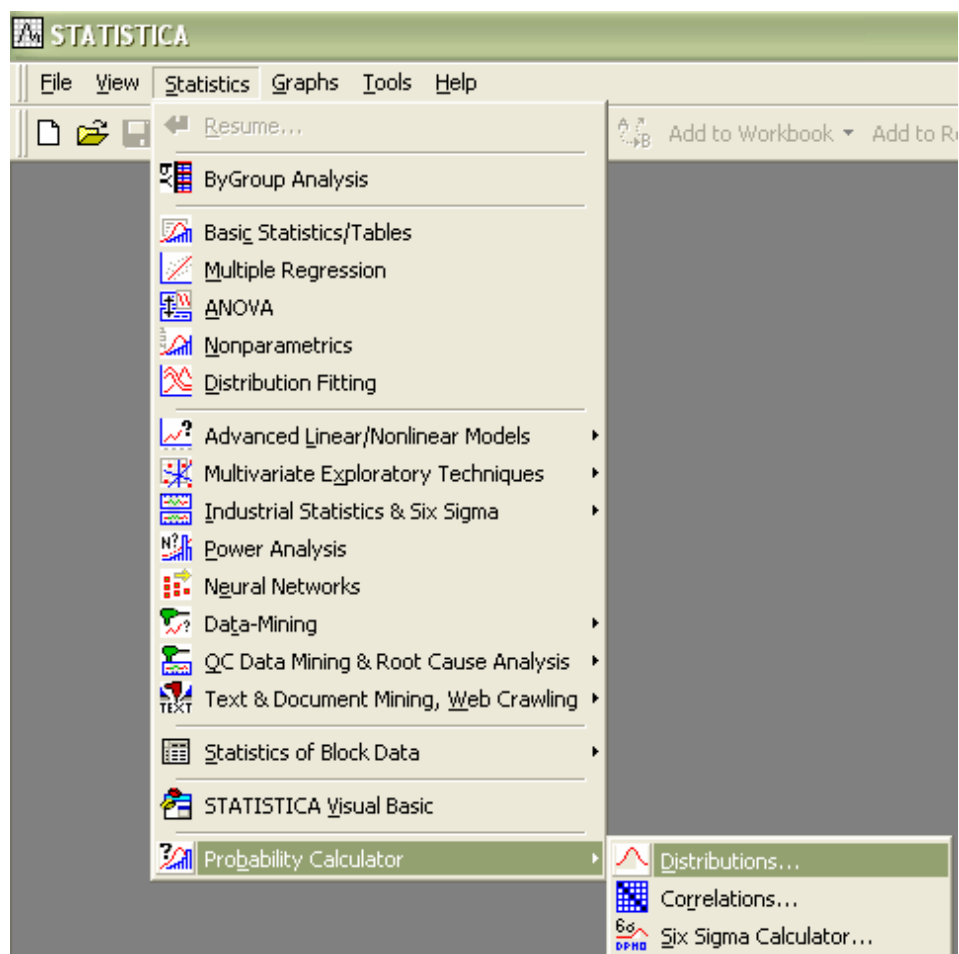


Рис. 3.4. Порядок открытия калькулятора распределений

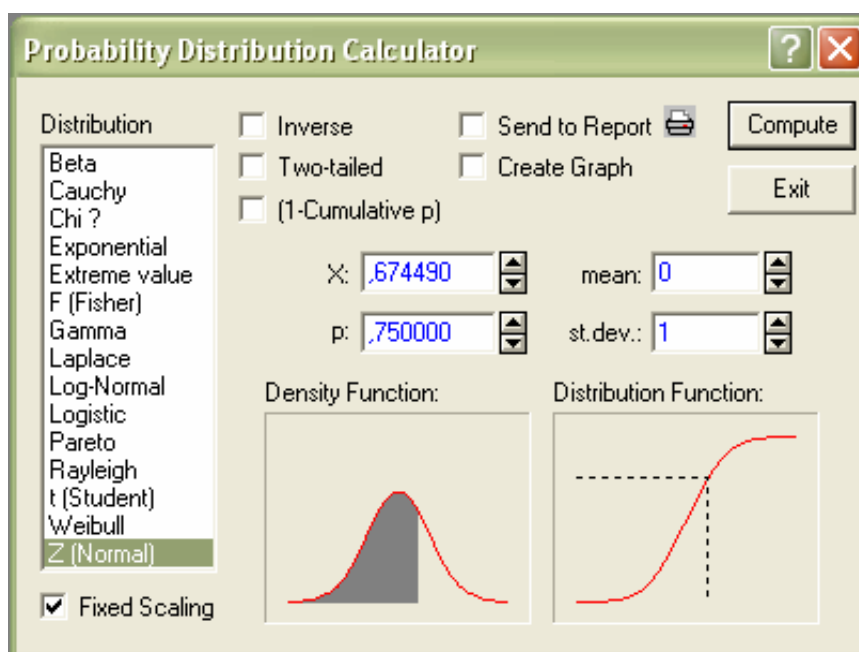


Рис. 3.5. Калькулятор распределений системы «STATISTICA»

В качестве примера рассмотрим построение функции распределения и функции плотности распределения стандартного нормального закона при значениях параметров: «mean» (среднее) равно нулю; «st.dev.» (среднеквадратическое отклонение) равно 1. Установите в окне калькулятора значения так, как показано на рисунке 3.6.

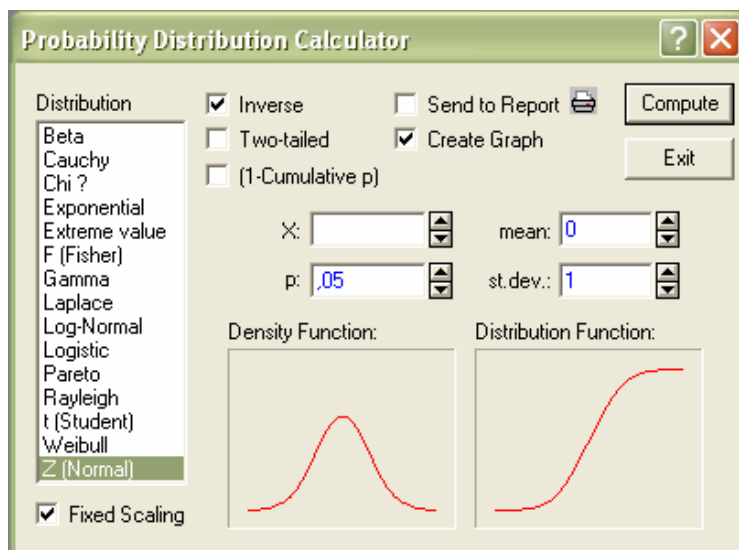


Рис. 3.6. Окно выбора закона распределения и установки его параметров

Таким образом установленные параметры позволят построить графики функции плотности и интегральной функции стандартного нормального распределения и вычислить значение левосторонней критической границы \underline{K}_α , соответствующей вероятности $\alpha = 0,05$ (в системе STATISTICA вместо обозначения α принято обозначение p). После установки нажмите «Compute» (вычислить). В результате будет открыто окно с требуемыми графиками и рассчитанным значением критической границы (рис. 3.7). Определим для стандартного нормального закона значение правосторонней критической границы \overline{K}_α , соответствующей вероятности $\alpha = 0,05$. Учитывая (3.4), можно задать значение $p = 1 - \alpha = 0,95$. После установки нажмите «Compute» (вычислить). В результате будет открыто окно с требуемыми графиками и рассчитанным значением критической границы (рис. 3.8).

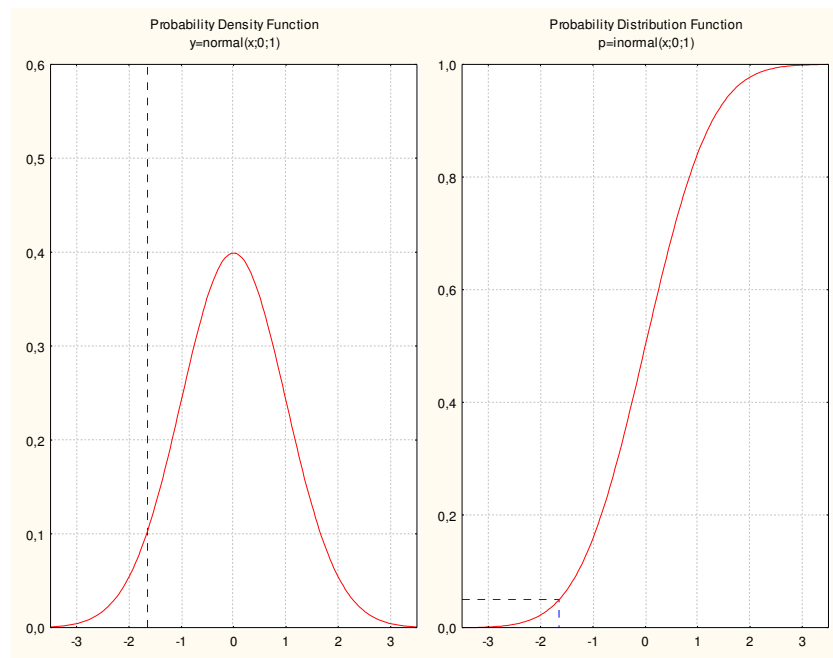


Рис. 3.7. Положение левосторонней критической границы, соответствующей вероятности $\alpha = 0,05$, на функции плотности (слева) и интегральной функции (справа) распределения стандартного нормального закона

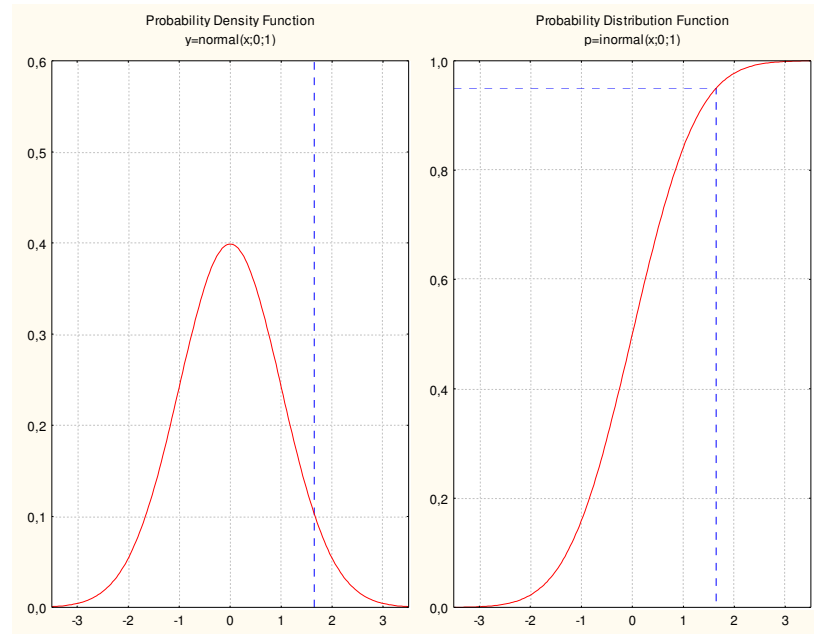


Рис. 3.8. Положение правосторонней критической границы, соответствующей вероятности $\alpha = 0,05$, на функции плотности (слева) и интегральной функции (справа) распределения стандартного нормального закона

Определим для нормального закона двусторонние критические границы \bar{B}_α и \underline{B}_α (или, что то же самое, $\pm U_\alpha$) при значении $\alpha = 0,05$. Учитывая (3.5), установим значение $p = 1 - \alpha = 0,95$, кроме того поставим «флажок» в окне «two-tailed» (двусторонние) и нажмем «Compute» (вычислить). Результат должен иметь вид, представленный на рисунке 3.9.

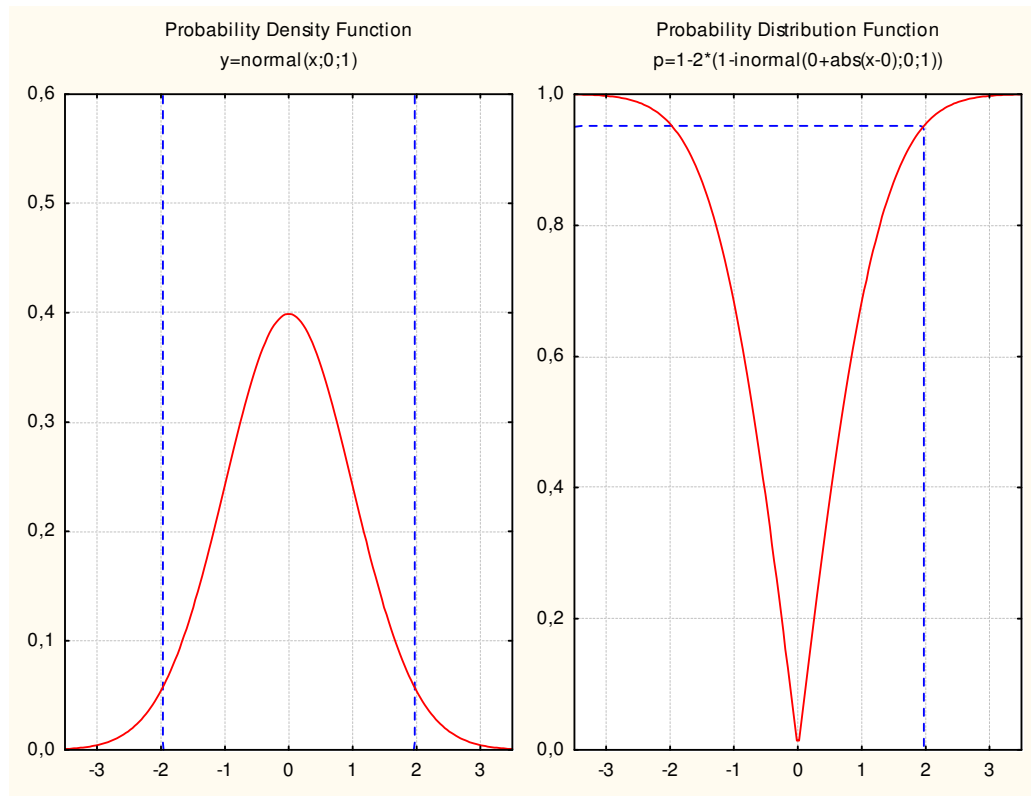


Рис. 3.9. Положение двусторонних критических границ, соответствующих вероятности $\alpha = 0,05$, на функции плотности (слева) и интегральной функции (справа) распределения стандартного нормального закона

Самостоятельно постройте функции плотности и интегральной функции, а также определите правостороннюю, левостороннюю и двусторонние критические границы, соответствующие вероятности $\alpha = 0,1$, для СВ X при значениях параметров $a = 1$ («mean»), $\sigma = 1$ («st.dev.»).

Задание для самостоятельной работы.

Для представленных в таблице 3.1 законов распределения построить интегральную функцию и функцию плотности распределения и определить значения правосторонних, левосторонних и двусторонних критических границ, соответствующих значениям α , равным 0,01, 0,05 и 0,1.

Таблица 3.1

Законы распределений, для построения графиков функций плотности распределения и определения значений критических границ

| Закон распределения | Параметры закона |
|-----------------------|---|
| Нормальный | $\mu = 1, \sigma = 1$ |
| Экспоненциальный | $\lambda = 2$ |
| Релея | $b = 0,5$ («scale») |
| Стьюдента | Число степеней свободы - $k = 10$ («df») |
| Фишера | Числа степеней свободы- $k_1 = 10$ и $k_2 = 10$ («df1» и «df2») |
| χ^2 (Хи-квадрат) | Число степеней свободы - $k = 10$ («df») |

Примечание: при определении двусторонних критических границ для некоторых законов (не являющихся симметричными, относительно центра распределения) не удастся установить опцию «two-tailed» (двусторонние). Поэтому для определения двусторонних границ, соответствующих уровню

α , необходимо учитывать тот факт, что $\int_{-\infty}^{B_\alpha} \varphi(x) dx = \frac{\alpha}{2}$ (см. рис. 3.3).

2. Оформление отчета по лабораторной работе.

Отчет по лабораторной работе №3 должен содержать:

- постановку задачи;
- графики функций плотности и интегральных функций распределения с нанесенными на них значениями правосторонних, левосторонних и двусторонних критических границ, соответствующих заданным значениям вероятности α .

4. Лабораторная работа №4

Тема работы: Проверка статистических гипотез (2 часа).

Цель работы: Освоение методов проверки статистических гипотез.

Содержание работы:

1. Проверка параметрических гипотез относительно равенства генеральной средней нормально распределенной генеральной совокупности установленному значению.

2. Проверка параметрической гипотезы относительно равенства двух генеральных средних двух нормально распределенных генеральных совокупностей.

3. Оформление отчета по лабораторной работе.

Краткая теоретическая справка.

Проверка статистических гипотез является одной из наиболее важных задач выборочного метода в статистике. Вообще под статистической гипотезой понимается некоторое предположение относительно генеральной совокупности, которое проверяется с использованием выборочных данных.

В зависимости от целей исследования может быть сформулировано большое количество различных гипотез. Ниже представлены статистические гипотезы, наиболее часто используемые в рамках выборочного метода:

- нормально распределенная генеральная совокупность имеет генеральную среднюю a , равную некоторому числу a_0 (т.е. $a = a_0$);
- нормально распределенная генеральная совокупность имеет генеральную дисперсию $\sigma^2 = \sigma_0^2$;
- выборки $X^1(x_1^1, x_2^1, \dots, x_n^1)$ и $X^2(x_1^2, x_2^2, \dots, x_k^2)$, взятые из нормально распределенных генеральных совокупностей, имеют равные генеральные средние $a_1 = a_2$;

– выборка $X(x_1, x_2, \dots, x_n)$ взята из генеральной совокупности с законом распределения $F(x)$ (иногда говорят: выборка хорошо согласуется с законом распределения $F(x)$).

Очевидно, что могут быть сформулированы и другие виды статистических гипотез.

При проверке какой-либо статистической гипотезы принято формулировать основную или нуль-гипотезу (обозначается символом H_0 ;) и альтернативную (конкурирующую) гипотезу (обозначаемую символом H_1 ;), которая является логическим отрицанием основной гипотезы. Например, при гипотезе $H_0: a = a_0$ альтернативной может являться одна из трех гипотез $H_1: a \neq a_0$, $H_1: a > a_0$ или $H_1: a < a_0$.

Правило, по которому решают принять нулевую гипотезу или отклонить ее в пользу альтернативной гипотезы, называется критерием. В общем случае алгоритм построения критерия состоит из следующих этапов:

1. По данным выборочного наблюдения рассчитывается некоторая величина - θ , которую принято называть «статистикой критерия» или просто «статистикой». В зависимости от выдвигаемой гипотезы формулы расчета θ , очевидно, будут разными.

2. Устанавливается закон распределения «статистики» $F(\theta)$.

3. Область значений «статистики» разбивается на две взаимодополняющие подобласти: область принятия гипотезы H_0 (назовем ее область S) и область отвержения гипотезы H_0 (назовем ее область \bar{S}). Очень часто область \bar{S} называют «критической» областью.

4. Если «статистика критерия» попадает в область S , то принимается нуль-гипотеза. В противном случае нуль-гипотеза отвергается в пользу альтернативной гипотезы.

Примечание: характер расположения областей S и \bar{S} в области значений статистики θ определяется двумя факторами. Во-первых,

характером альтернативной гипотезы H_1 , а во-вторых, уровнем вероятности попадания статистики в «критическую область». Данную вероятность принято называть уровнем значимости проверки нуль-гипотезы и обозначать символом α . Уровень значимости при проверке гипотезы задается заранее и принимается достаточно малым (например, 0,01; 0,05; 0,1).

При заданном значении α , область S принятия нуль-гипотезы будет определяться вероятностью принадлежности статистики θ этой области, равной $1 - \alpha$.

Очевидно, что при известном законе распределения статистики θ и сформированной альтернативной гипотезе критическая область будет определяться:

- либо левосторонней критической границей $\underline{\theta}_\alpha$ распределения величины θ , соответствующей вероятности $\alpha - \underline{\theta}_\alpha$ (в этом случае «критическую область» называют левосторонней);
- либо правосторонней критической границей распределения величины θ , соответствующей вероятности $\alpha - \bar{\theta}_\alpha$ (в этом случае «критическую область» называют правосторонней);
- либо двумя двусторонними критическими границами распределения величины θ , соответствующими вероятности $\alpha - \underline{\theta}_{\alpha/2}$ и $\bar{\theta}_{\alpha/2}$ (в этом случае «критическую область» называют двусторонней).

Рассмотрим следующий пример. Пусть *утверждается*, что в некоторой статистической совокупности студентов их средний рост равен 175 см. Пусть по данным выборочного наблюдения (объем выборки $n = 100$ наблюдений) собраны данные о росте студентов. Пусть по этим данным определены: выборочное среднее арифметическое значение роста $\bar{x} = 178$ см, а выборочное среднеквадратическое отклонение роста в генеральной совокупности известно и равно $\sigma = 10$ см. Пусть также известно, что рост студентов в их генеральной совокупности является нормально распределенной случайной величиной. Возникает вопрос, можно ли

согласиться с утверждением о том, что полученные выборочные данные не противоречат сделанному утверждению.

Поставленная задача – типичный пример необходимости проверки статистической гипотезы. В этой задаче $H_0: a = a_0 = 175$ см. Поскольку выборочное среднее оказалось больше $a_0 = 175$ см, то в качестве альтернативной гипотезы можно выбрать $H_1: a > 175$ см (т.е. допустить, что среднее значение роста может отклоняться от a_0 только в область больших значений).

Известно, что статистика для проверки такой гипотезы вычисляется по формуле:

$$Z = \frac{\bar{x} - a_0}{\sigma / \sqrt{n}}.$$

Кроме того, известно, что величина Z является стандартной нормально распределенной случайной величиной. Выберем уровень значимости при проверке нуль-гипотезы $\alpha = 0,05$. Так как в качестве альтернативной гипотезы выбрано утверждение, что средний рост в генеральной совокупности больше, чем $a_0 = 175$ см, то условие отклонения гипотезы H_0 будет определяться значением правосторонней критической границы стандартного нормального распределения, соответствующей вероятности $\alpha = 0,05$ - \bar{Z}_α , которую можно найти из условия:

$$\Phi(\bar{Z}_\alpha) = 0,5 - \alpha.$$

Рассчитаем значение статистики:

$$Z = \frac{178 - 175}{10 / \sqrt{100}} = 3.$$

Определим значение правосторонней критической границы, соответствующей выбранному уровню значимости (например, по таблицам значений функции Лапласа или с использованием калькулятора распределений системы «STATISTICA»): $\bar{Z}_{0,05} = 1,64$. Поскольку расчетное значение статистики $Z = 3$ существенно больше, чем правосторонняя

критическая граница (иными словами, значение статистики Z принадлежит области отвержения гипотезы \bar{S}), то нуль-гипотезу мы должны отвергнуть в пользу альтернативной и считать, что среднее значение роста в генеральной совокупности студентов больше, чем $a_0 = 175$ см.

Примечание: в рассмотренной задаче в качестве альтернативной гипотезы можно было принять гипотезу $H_1: a \neq a_0$. В этом случае «критическая область» определялась бы двусторонними критическими границами стандартного нормального распределения (т.е. «критическая область» была бы двусторонней, так как в альтернативной гипотезе допускается отклонение среднего роста от значения $a_0 = 175$ см как в область больших значений, так и в область меньших значений). В этом случае гипотеза H_0 принималась бы при выполнении условия: $|Z| < U_\alpha$, где U_α - значение критической границы стандартного нормального распределения, определяемое из условия:

$$\Phi(U_\alpha) = \frac{1-\alpha}{2} = 0,5 - \frac{\alpha}{2}.$$

Выполнение работы.

1. Проверка параметрических гипотез относительно равенства генеральной средней нормально распределенной генеральной совокупности установленному значению.

В системе «Statistica» создайте файл размерности 2×30 (2 переменные, 30 наблюдений). Переменным присвойте имена и характеристики в соответствии с таблицей 4.1.

Выполненные действия позволят сгенерировать две выборки, извлеченные из нормально распределенных генеральных совокупностей с различными генеральными средними значениями ($a_1 = 5$ и $a_2 = 4$) и одинаковым генеральным среднеквадратическим отклонением ($\sigma = 0,5$). Сохраните файл.

Таблица 4.1

Имена и характеристики переменных

| Переменная | Имя | Формат отображения | Тип | Количество десятичных знаков | Вводимая строка в поле «Длинное имя или формула» |
|------------|----------|-----------------------|--------|------------------------------------|--|
| Var1 | Выборка1 | Числовой | Double | 1 | =VNormal(rnd(1);5;0,5) |
| Var2 | Выборка2 | Числовой | Double | 1 | = VNormal(rnd(1);4;0,5) |

В пункте меню «Statistics» выберите пункт «Basic statistics/Tables» (Основные статистики/Таблицы). В открывшемся окне выберите пункт «Descriptive statistics» (Описательные статистики) и откройте окно расчета описательных статистик. Перейдите от вкладки «Quick» (Быстрый) к вкладке «Advanced» (Расширенный). На данной вкладке отметьте рассчитываемые показатели так, как показано на рисунке 4.1.

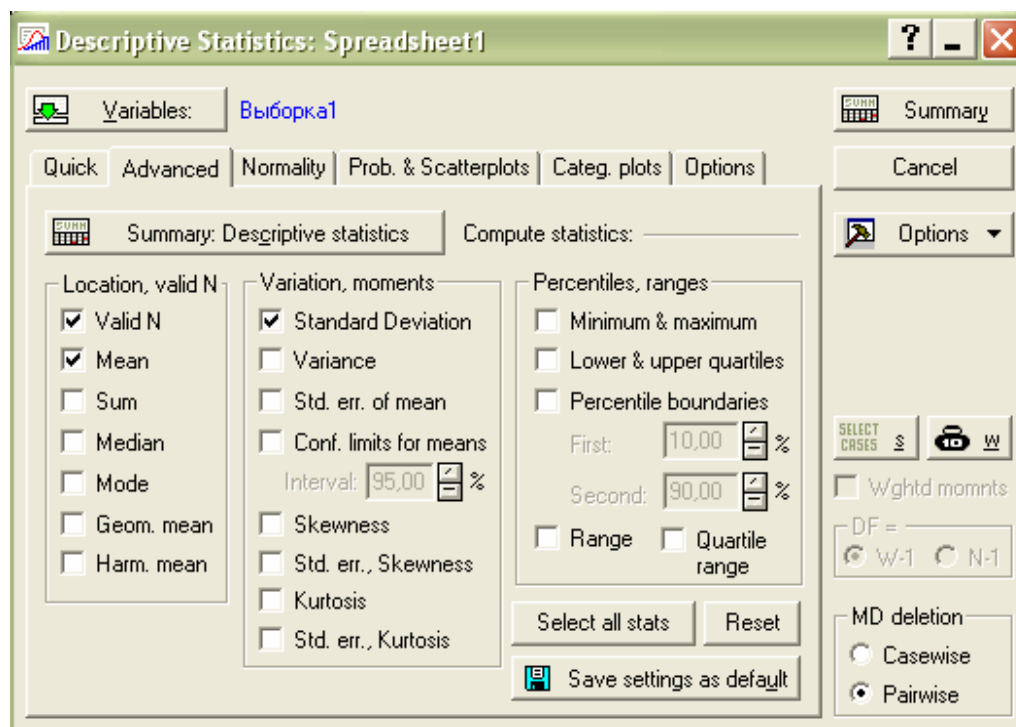


Рис. 4.1. Окно расчета описательных статистик. Вкладка «Advanced (расширенное)»

Установив флажки так, как показано на рисунке 4.1, система произведет расчет количества наблюдений (Valid N), выборочного среднего арифметического (Mean) и выборочного среднеквадратического отклонения (Standard Deviation). Выполните расчет для обеих выборок и убедитесь, что значения рассчитанных параметров отличаются от тех, которые вы задавали при генерации данных.

Для проверки гипотез о равенстве генеральных средних значениям $H_0 : a_1 = a_0 = 5$ и $H_0 : a_2 = a_0 = 4$ соответственно закройте окно описательных статистик и в окне «Basic statistics/Tables (Основные статистики/Таблицы)» выберите пункт «t-test, single sample» (рис. 4.2).

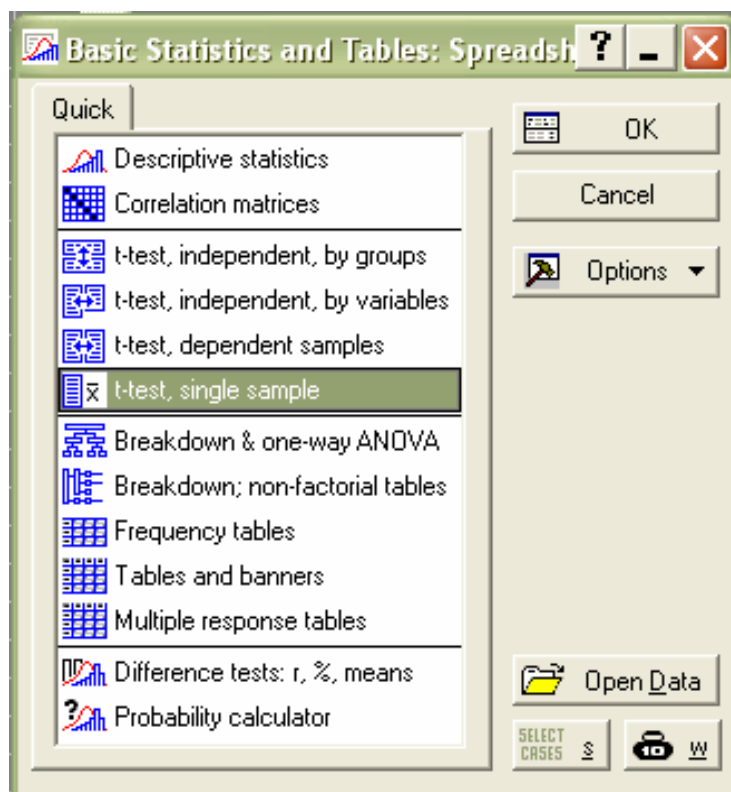


Рис. 4.2. Начальное окно расчета описательных статистик

В результате будет открыто окно проверки простых параметрических гипотез относительно генеральных средних (рис. 4.3).

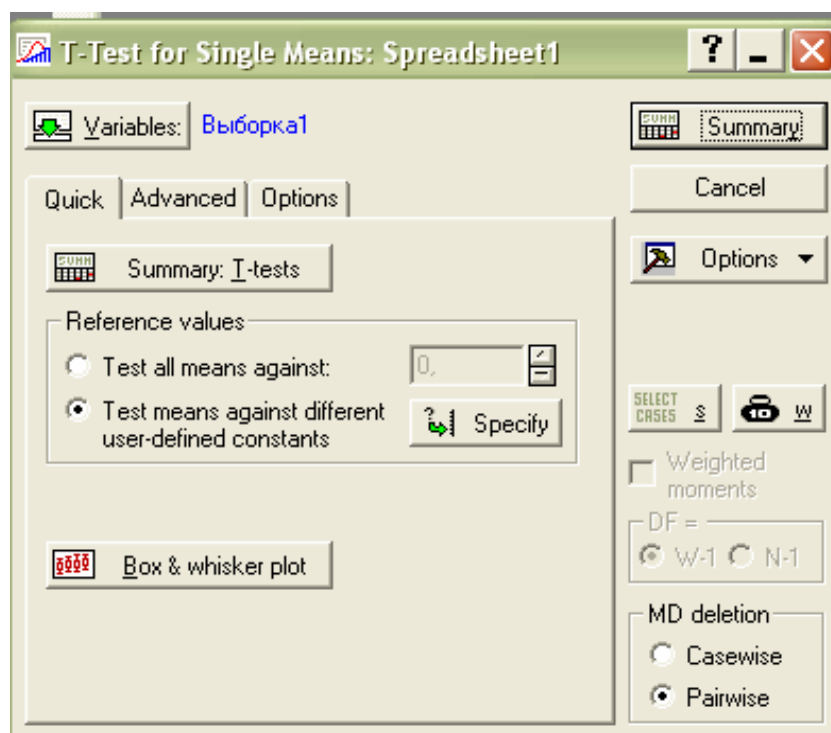



Рис. 4.3. Окно проверки параметрических гипотез относительно среднего значения генеральной совокупности

Выберите переменную «Выборка 1» (нажав на кнопку  Variables:) и отметьте пункт «test means against different user-defined constant» (что можно перевести как «гипотеза для среднего против *неравенства* установленной пользователем константе»). В этом случае мы проверяем гипотезу $H_0 : a_1 = a_0$ при $H_1 : a_1 \neq a_0$.

Перейдите на вкладку «Advanced» («Расширенное») и убедитесь, что в поле «p-level for highlighting» (см. рис. 4.4) установлено значение уровня значимости при проверке гипотезы равное 0,05. При такой установке значения t-статистики, рассчитываемой по формуле $t = \frac{\bar{x} - a_0}{S/\sqrt{n-1}}$ (статистики Стьюдента) и выходящие за *двусторонние* критические границы распределения Стьюдента с $n-1$ степенью свободы, соответствующие установленному уровню значимости, будут выделены красным цветом.

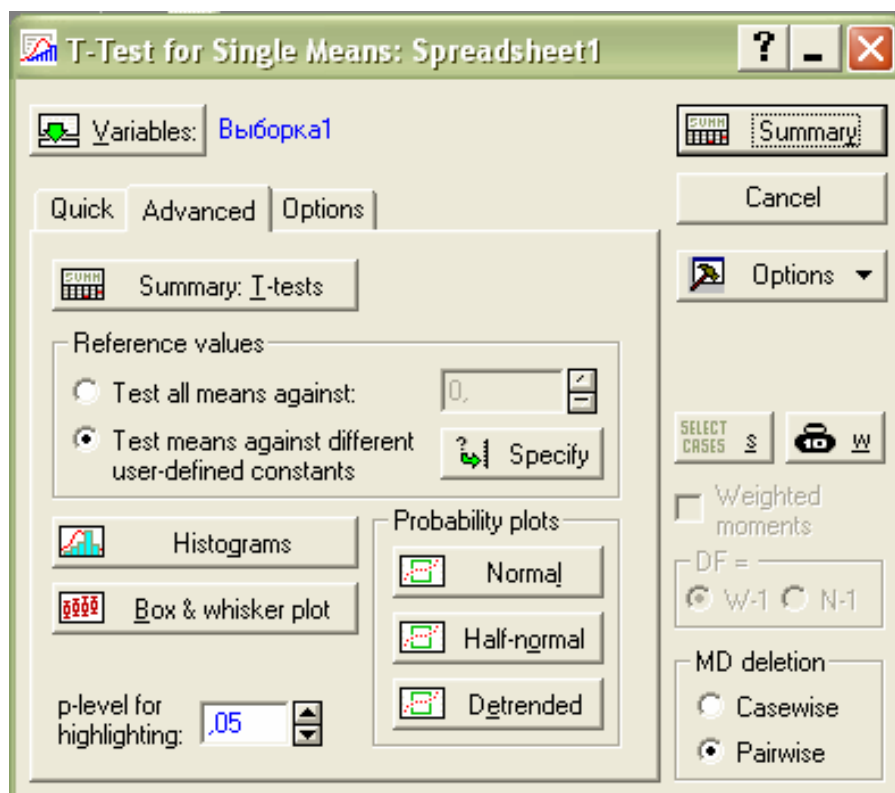



Рис. 4.4. Окно установки дополнительных параметров при проверке гипотезы относительно среднего значения

Нажмите кнопку  Specify и установите в открывшемся окне значение генеральной средней в первой совокупности « $a_0 = 5$ » (рис. 4.5).

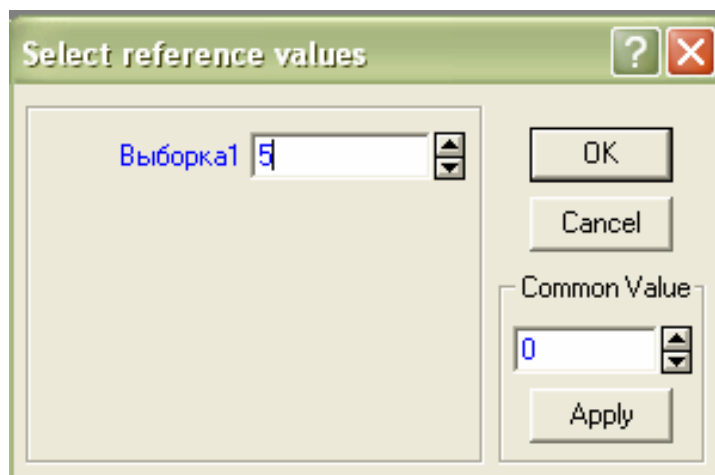

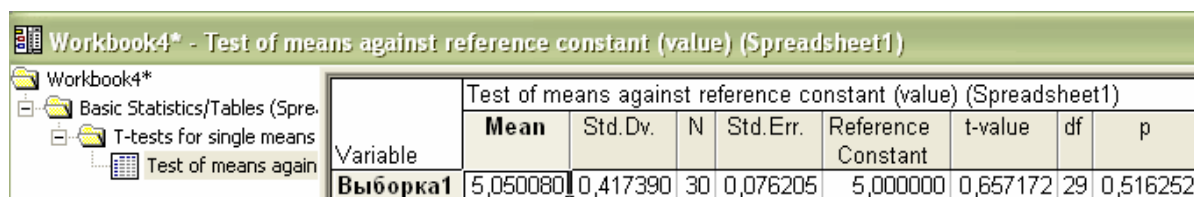


Рис. 4.5. Окно установки значения параметра a_0

После нажатия кнопки «ОК» нажмите кнопку . Система произведет расчет необходимых значений и представит результаты так, как показано на рисунке 4.6 (обратите внимание, что результаты, получившиеся у вас, могут отличаться от представленных).



| Variable | Mean | Std.Dv. | N | Std.Err. | Reference Constant | t-value | df | p |
|----------|----------|----------|----|----------|--------------------|----------|----|----------|
| Выборка1 | 5,050080 | 0,417390 | 30 | 0,076205 | 5,000000 | 0,657172 | 29 | 0,516252 |

Рис. 4.6. Результаты проверки статистической гипотезы относительно равенства генеральной средней, проверяемой по «Выборке 1», значению

$$a_0 = 5$$

Проанализируем результаты, представленные в таблице на рисунке 4.6. В таблице представлены:

- выборочное среднее значение (Mean) - 5,050080;
- выборочное среднее квадратическое отклонение (Std.dev.) - 0,417390;
- количество наблюдений (N) – 30;
- среднее квадратическое отклонение средней арифметической (Std.err) – 0,076205;
- значение константы, относительно которой проверяется гипотеза (Reference Constant) – 5;
- расчетное значение статистики Стьюдента (t-value) – 0,657172;
- число степеней свободы статистики Стьюдента (df = N-1) – 29;
- вероятность того, что значения статистики (по модулю) могут принять такие же или еще большие значения при справедливости нуль-гипотезы (p) – 0,516252.

Обратите внимание на то, что значение вероятности p, представленное в последней колонке таблицы, существенно больше установленного нами

уровня значимости $\alpha = 0,05$. Последнее означает, что при справедливости нуль-гипотезы такое значение статистики может быть обусловлено случайными причинами. А это, в свою очередь, говорит о том, что у нас **нет оснований не принять** нуль-гипотезу.

Задание для самостоятельной работы.


Проверьте по данным «Выборки 1» и данным «Выборки 2» следующие статистические гипотезы:

- $H_0 : a_1 = a_0 = 4$ при $H_1 : a_1 \neq a_0 = 4$;
- $H_0 : a_2 = a_0 = 4$ при $H_1 : a_2 \neq a_0 = 4$;
- $H_0 : a_2 = a_0 = 5$ при $H_1 : a_2 \neq a_0 = 5$.

Проанализируйте полученные результаты. Представьте соответствующие выводы.

2. Проверка параметрической гипотезы относительно равенства двух генеральных средних двух нормально распределенных генеральных совокупностей.

В этом случае мы пытаемся проверить статистическую гипотезу следующего содержания: $H_0 : a_1 = a_2$ при $H_1 : a_1 \neq a_2$. Указанная гипотеза проверяется в предположении, что в обеих генеральных совокупностях дисперсии являются равными. Иными словами, исходная гипотеза должна проверяться совместно с гипотезой $H_0 : \sigma_1^2 = \sigma_2^2$ при $H_1 : \sigma_1^2 > \sigma_2^2$.

Для проверки этих гипотез вернитесь к окну «Basic statistics/Tables (Основные статистики/Таблицы)», выберите пункт «t-test, independent sample» (см. рис. 4.2). Нажмите . В открывшемся окне установите параметры так, как показано на рисунке 4.7. Убедитесь, что на вкладке «Options» в поле «p-level for highlighting» установлено значение 0,05.

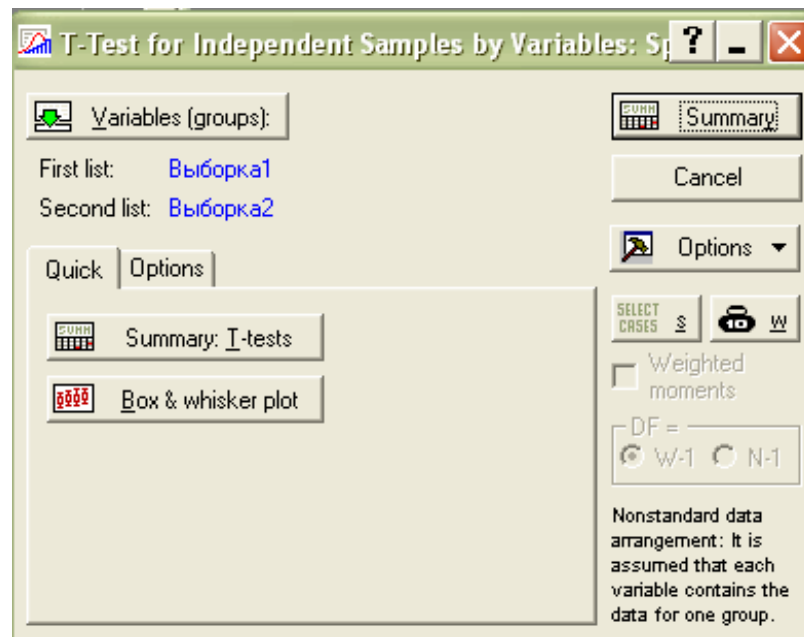



Рис. 4.7. Окно проверки статистической гипотезы о равенстве двух генеральных средних двух нормально распределенных генеральных совокупностей

Нажмите кнопку  Summary: I-tests. Система осуществит расчет t-статистики (статистики Стьюдента), F-статистики (статистики Фишера) по формулам:

$$t = \frac{(\bar{x}_1 - \bar{x}_2)}{S^* \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}; S^{*2} = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{(n_1 + n_2 - 2)}; F = \frac{\sigma_2^2}{\sigma_1^2},$$

сделает ряд дополнительных вычислений и представит результаты, как показано на рисунке 4.8. (обратите внимание, что результаты, получившиеся у вас, могут отличаться от представленных).

Workbook24* - T-test for Independent Samples (Spreadsheet1)

Basic Statistics/Tables (Spreadsheet1)

T-test for independent s

T-test for Independent

| Group 1 vs. Group 2 | | Mean | Mean | t-value | df | p | Valid N | Valid N | Std. Dev. | Std. Dev. | F-ratio | p |
|---------------------|----------|----------|----------|----------|----|----------|---------|---------|-----------|-----------|-----------|-----------|
| Group 1 | Group 2 | Group 1 | Group 2 | | | | Group 1 | Group 2 | Group 1 | Group 2 | Variances | Variances |
| Выборка1 | Выборка2 | 5,050080 | 4,099263 | 8,059059 | 58 | 0,000000 | 30 | 30 | 0,417390 | 0,493327 | 1,396964 | 0,373272 |

Рис. 4.8. Результаты проверки гипотезы о равенстве генеральных средних двух нормально распределенных генеральных совокупностей

Проанализируем полученные результаты. В таблице рассчитаны:

- выборочные средние значения (Mean Group 1 и Mean Group 2) в первой и второй выборках - 5,050080, 4,099263 соответственно;
- значения среднеквадратических отклонений (Std. Dev. Group 1 и Std.Dev. Group 2) в первой и второй выборках – 0,417390, 0,493327 соответственно;
- количество наблюдений в обеих выборках (Valid N Group 1 и Valid N Group 2) – 30;
- значение t-статистики (статистики Стьюдента - t) – 8,059059;
- число степеней свободы t-статистики ($df = n_1 + n_2 - 2$) – 58;
- вероятность того, что при справедливости нуль-гипотезы относительно средних, значения t-статистики (по модулю) могут принять такие же или еще большие значения (p) – 0,00000;
- значение F-статистики (статистики, проверяющей гипотезу о равенстве дисперсий в первой и во второй выборке) - 1,396964;
- вероятность того, что при справедливости нуль-гипотезы относительно дисперсий, значения F-статистики могут принять такие же или еще большие значения (p) – 0,373272.

Обратите внимание на то, что значение вероятности p, представленное в последней колонке таблицы, больше установленного нами уровня значимости $\alpha = 0,05$. Это говорит о том, что при справедливости нуль-гипотезы $H_0 : \sigma_1^2 = \sigma_2^2$ вероятность получить такие же или большие значения F-статистики составляет 0,373272. Эта вероятность достаточно велика, и, следовательно, полученное расхождение **может быть** объяснено случайными причинами. А это, в свою очередь, говорит о том, что у нас **нет оснований отвергнуть нуль-гипотезу** о равенстве дисперсий в двух генеральных совокупностях.

Теперь обратите внимание, что значение вероятности p, представленное в пятой колонке таблицы, существенно **меньше**

установленного нами уровня значимости $\alpha = 0,05$. Это говорит о том, что при справедливости нуль-гипотезы $H_0 : a_1 = a_2$ вероятность получить такие же или большие значения t-статистики практически равна нулю, т.е. полученное расхождение не может быть объяснено случайными причинами. А это, в свою очередь, говорит о том, что **нуль-гипотезу мы должны отклонить** и считать, что генеральные средние в первой и второй генеральных совокупностях различаются значимо.

Задание для самостоятельной работы.

Закройте окно анализа. В окне задания характеристик переменной «Выборка 2» в строке «Длинное имя или формула» наберите $= VNormal(rnd(1);5;0,5)$. После пересчета значений переменной еще раз проверьте гипотезу $H_0 : a_1 = a_2$ при $H_1 : a_1 \neq a_2$. Сделайте соответствующие выводы.

3. Оформление отчета по лабораторной работе.

Отчет по лабораторной работе №4 должен содержать:

- постановку задачи;
- таблицы с результатами проверки параметрических гипотез, указанных в пунктах 1 и 2;
- выводы, сделанные по результатам проверки гипотез.

5. Лабораторная работа №5

Тема работы: Проверка непараметрических гипотез относительно законов распределения по критерию согласия Пирсона χ^2 (2 часа).

Цель работы: Освоение методов проверки непараметрических гипотез.


Содержание работы:

1. Проверка непараметрических гипотез относительно законов распределения по критерию согласия Пирсона χ^2 .
2. Оформление отчета по лабораторной работе.

Выполнение работы.

1. Проверка непараметрических гипотез относительно законов распределения по критерию согласия Пирсона χ^2 .

Откройте файл с данными, созданный при выполнении лабораторной работы №2. Файл должен иметь структуру 5×100 , т.е. содержать 5 выборок по 100 наблюдений, «извлеченных» из генеральных совокупностей, подчиняющихся различным законам распределений случайных величин.

В меню «Статистика» выберите пункт  Distribution Fitting (подгонка распределений). В результате выбора откроется начальная панель проверки гипотез относительно законов распределения с использованием критериев согласия (рис. 5.1). В этом окне необходимо выбрать тип распределения для анализируемой переменной.

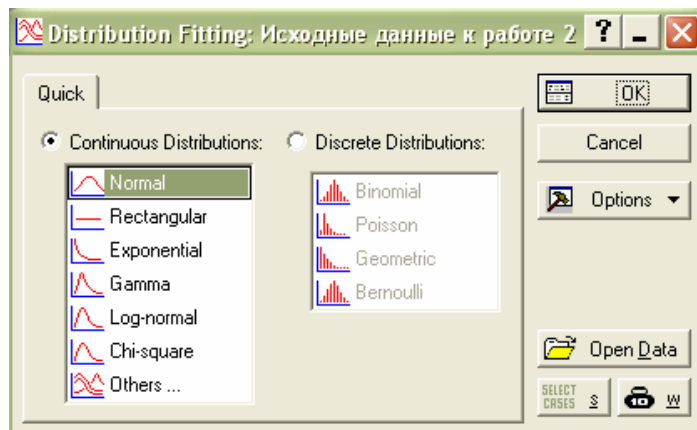



Рис. 5.1. Панель выбора распределения при проверке гипотез относительно закона распределения

Необходимо вспомнить, что в файле исходных данных к лабораторной работе №2 переменные формировались путем генерации случайных величин, подчиняющихся закону равной вероятности, нормальному закону (Гаусса), экспоненциальному (показательному) закону, а также законам распределения Релея и Пуассона.

Т.к. первая переменная файла исходных данных должна подчиняться закону равной вероятности, то в начальном окне необходимо выбрать тип распределения  Rectangular. Иными словами, мы хотим проверить гипотезу H_0 : *выборка извлечена из генеральной совокупности, подчиняющейся закону распределения равной вероятности*. Нажмите «ОК». В результате откроется окно выбора переменных и условий проверки гипотезы относительно закона распределения выборки по критериям согласия.

Выберите переменную «Равномерное распределение» и перейдите на вкладку «Параметры». На этой вкладке задаются условия проверки гипотезы H_0 (рис. 5.2).

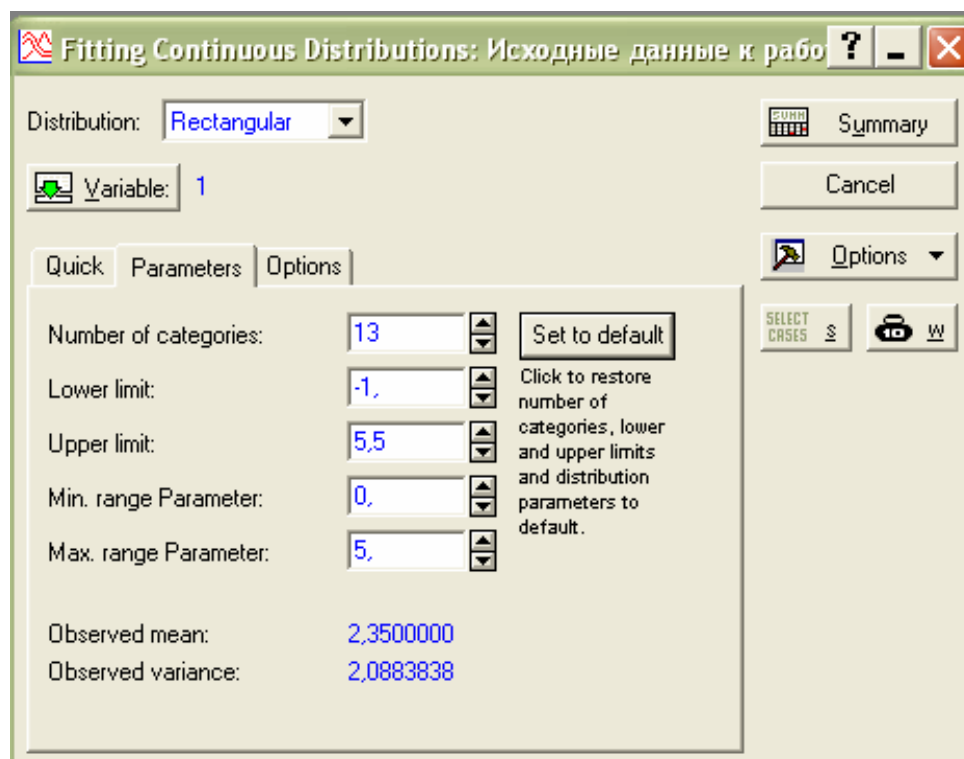


Рис. 5.2. Окно задания параметров проверки гипотезы H_0

Изначально система STATISTICA самостоятельно установит: число интервалов (категорий); нижнюю границу первого и верхнюю границу последнего интервала; параметры x_1 и x_2 закона распределения, а также рассчитает среднее арифметическое и дисперсию значений переменной, выбранной для анализа.

Измените значения параметров анализа так, как показано на рисунке 5.3, и перейдите назад на вкладку «Быстрое (Quick)».

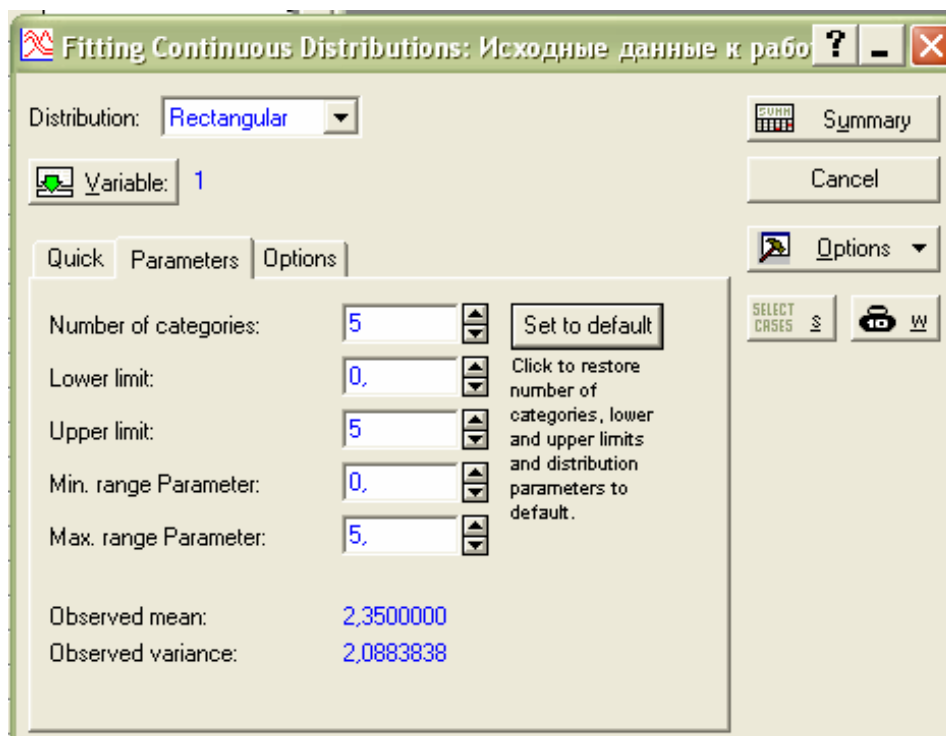



Рис. 5.3. Установленные параметры проверки гипотезы H_0

Нажмите на кнопку  Summary: Observed and expected distribution (итог, наблюдаемые и рассчитанные частоты). В результате откроется окно с результатами расчета критерия согласия Пирсона χ^2 (рис. 5.4).

Проанализируем полученные результаты (**примечание:** полученные результаты и проводимый далее анализ справедливы для исходных данных, которые использовались при подготовке методических указаний. В каждом конкретном случае результаты и выводы будут различными).

| Upper Boundary | Observed Frequency | Cumulative Observed | Percent Observed | Cumul. % Observed | Expected Frequency | Cumulative Expected | Percent Expected | Cumul. % Expected | Observed-Expected |
|----------------|--------------------|---------------------|------------------|-------------------|--------------------|---------------------|------------------|-------------------|-------------------|
| $-\infty$ | 34 | 34 | 34,00000 | 34,0000 | 20,00000 | 20,0000 | 20,00000 | 20,0000 | 14,0000 |
| 1,00000 | 18 | 52 | 18,00000 | 52,0000 | 20,00000 | 40,0000 | 20,00000 | 40,0000 | -2,0000 |
| 2,00000 | 24 | 76 | 24,00000 | 76,0000 | 20,00000 | 60,0000 | 20,00000 | 60,0000 | 4,0000 |
| 3,00000 | 17 | 93 | 17,00000 | 93,0000 | 20,00000 | 80,0000 | 20,00000 | 80,0000 | -3,0000 |
| 4,00000 | 7 | 100 | 7,00000 | 100,0000 | 20,00000 | 100,0000 | 20,00000 | 100,0000 | -13,0000 |
| $+\infty$ | | | | | | | | | |

Рис. 5.4. Результаты проверки гипотезы H_0

В первой колонке таблицы показаны интервалы группирования значений переменной ($[-\infty, 1)$; $[1, 2)$; ...; $[4, +\infty)$). Далее по колонкам представлены:

- наблюдаемые частоты (Observed Frequency);
- накопленные наблюдаемые частоты (Cumulative Observed);
- частоты (Percent Observed);
- накопленные частоты (Cumulative % Observed);
- расчетные частоты (Expected Frequency);
- накопленные расчетные частоты (Cumulative Expected);
- расчетные частоты (Percent Expected). Расчетные частоты – это аналог вероятности попадания значений исследуемого признака в соответствующий интервал при справедливости проверяемой гипотезы;
- накопленные расчетные частоты (Cumulative % Expected);
- разность накопленных и расчетных частот (Observed- Expected).


В верхней строке представлены следующие результаты:

- переменная, относительно которой проверяется гипотеза H_0 (равномерное распределение);
- вид закона распределения (Distribution: Rectangular (закон равной вероятности) и наименование файла с исходными данными;
- расчетное значение статистики χ^2 ;
- число степеней свободы статистики χ^2 (df);

– значение вероятности p , соответствующее расчетному значению статистики χ^2 .

Собственно три последних значения и позволяют сделать вывод о том, можно или нет принимать гипотезу H_0 . Предположим, что мы хотим проверить гипотезу H_0 на уровне значимости $\alpha = 0,05$.

Обратите внимание на то, что значение вероятности $p = 0,00005$, представленное в верхней строке таблицы результатов, существенно *меньше* установленного нами уровня значимости $\alpha = 0,05$. Последнее означает, что при справедливости гипотезы H_0 такое значение статистики не может быть обусловлено случайными причинами. А это, в свою очередь, говорит о том, что у нас *нет оснований принять* гипотезу H_0 (иными словами, нельзя считать, что выборка подчиняется закону равной вероятности).

Закройте окно результатов и вернитесь в окно анализа. Нажмите на вкладке «Quick» (Быстрое) кнопку  Plot of observed and expected distribution. В результате будет построена гистограмма распределения выборки с наложенной на нее функцией плотности закона равной вероятности (рис. 5.5).

Примечание: соответствующие подписи осей и заголовков графика можно изменить средствами STATISTICA, и в таком виде графики необходимо представлять в отчете по лабораторной работе.

Анализируя график, представленный на рисунке 5.5, и данные заголовка графика, также можно сделать вывод о том, что исходные данные плохо согласуются с законом равной вероятности, несмотря на то, что выборка была получена путем генерации данных, подчиняющихся именно этому закону. Последнее может быть обусловлено тем, что критерий согласия Пирсона χ^2 является чувствительным к объему наблюдений и характеру группирования единиц выборки по интервалам (или группам).

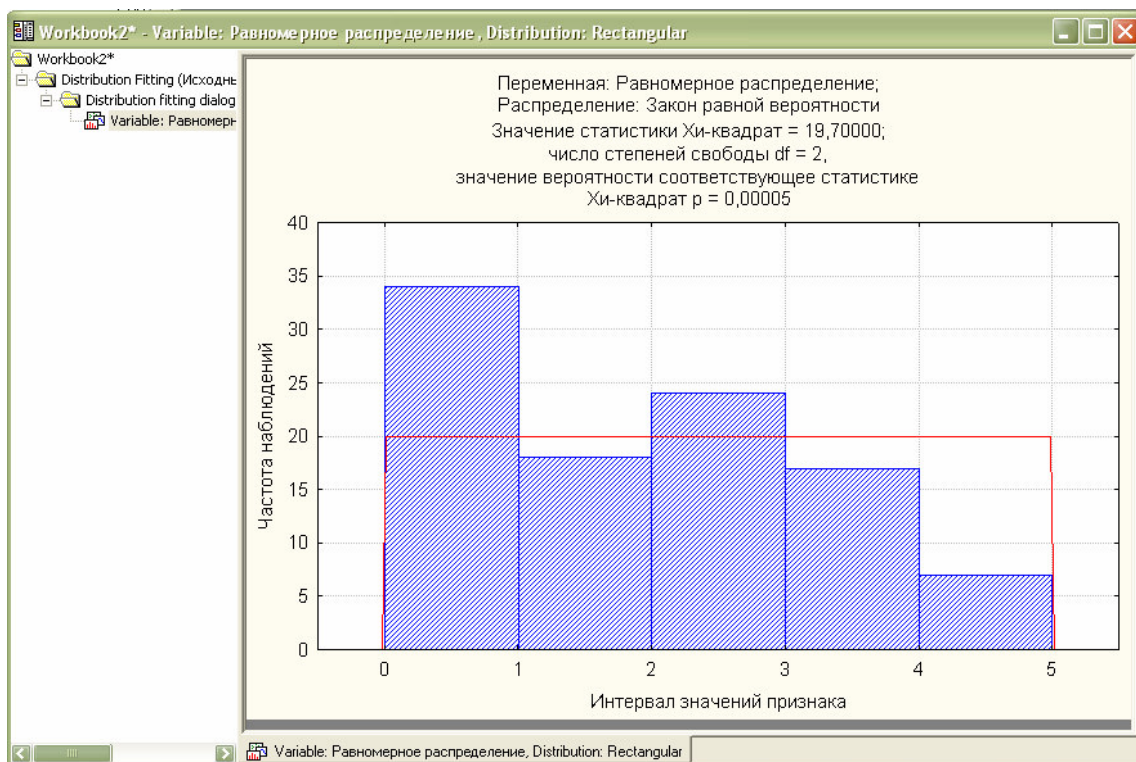


Рис. 5.5. Гистограмма распределения переменной с наложенной функцией плотности закона равной вероятности и результатами расчета статистики χ^2

Для проверки такой чувствительности исследуйте поведение статистики χ^2 при изменении числа интервалов группирования с 5 на какое-либо другое значение. Сделайте соответствующие выводы.

Примечание. Для установления нового числа групп вернитесь в окно анализа на вкладку «Parameters» (Параметры) и в поле Number of categories: установите необходимое вам число групп.

Задание для самостоятельной работы.

Для переменных: Нормальное распределение, Экспоненциальное распределение и распределение Пуассона (не забудьте, что распределение Пуассона является законом распределения дискретных случайных величин) - проверьте гипотезы о соответствии распределений выборочных значений переменных указанным законам распределения случайных величин. Выбор количества интервалов предоставьте сделать системе STATISTICA.

2. Оформление отчета по лабораторной работе.

Отчет по лабораторной работе №5 должен содержать:

- постановку задачи;
- таблицы с результатами проверки гипотез относительно законов распределения для переменных, указанных в задании для самостоятельной работы;
- гистограммы распределений значений признака с наложенной функцией плотности соответствующего закона и результатами расчета статистики χ^2 .

6. Лабораторная работа №6

Тема работы: Дисперсионный анализ (2 часа).

Цель работы: Освоение методов проведения дисперсионного анализа.

Содержание работы:

1. Процедура однофакторного дисперсионного анализа.
2. Процедура многофакторного дисперсионного анализа (на примере двухфакторного).
3. Создание отчета.

Краткая теоретическая справка. Дисперсионный анализ (ДА) используется в статистических исследованиях с целью выявления влияния на изучаемый статистический признак различных факторов, обычно не имеющих количественного выражения. Сущность ДА состоит в разложении общей вариации результирующего показателя на компоненты, отражающие влияние изучаемого (изучаемых) фактора (факторов) и неучтенных факторов. В зависимости от количества исследуемых факторов модели дисперсионного анализа принято подразделять на одно-, двух-, ..., многофакторные.

Однофакторный ДА. В рассматриваемом случае исследуется влияние одного фактора. Данные для проведения анализа получают в процессе проведения специальным образом организованных исследований. Исходные данные представляют собой значения результирующего признака, зафиксированные при m уровнях фактора, сведенные в таблицу 6.1. Количество наблюдений при каждом уровне фактора принято равным n .

Таблица 6.1

Исходные данные для проведения однофакторного ДА

| № уровня фактора | Значение результирующего признака | | | | |
|------------------|-----------------------------------|----------|----------|----------|----------|
| | 1 | 2 | 3 | | n |
| 1 | x_{11} | x_{12} | x_{13} | | x_{1n} |
| 2 | x_{21} | x_{22} | x_{23} | | x_{2n} |
| \vdots | \vdots | \vdots | \vdots | \vdots | \vdots |
| m | x_{m1} | x_{m2} | x_{m3} | | x_{mn} |

При проведении однофакторного ДА проверяется статистическая гипотеза:

$$H_0 : a_1 = a_2 = \dots = a_m \text{ против } H_1 : a_1 \neq a_2 \neq \dots \neq a_m, \quad (6.1)$$

где a_1, a_2, \dots, a_m - генеральные средние значения результирующего признака, зафиксированные при первом, втором, ..., m -м уровне фактора соответственно. Проверку гипотезы (6.1) проводят, рассчитывая величины:

$$Q^2 = \sum_{i=1}^m \sum_{j=1}^n (x_{ij} - \tilde{x})^2 ; Q_1^2 = n \sum_{i=1}^m (\tilde{x}_i - \tilde{x})^2 ; Q_2^2 = \sum_{i=1}^m \sum_{j=1}^n (x_{ij} - \tilde{x}_i)^2, \quad (6.2)$$

где \tilde{x}_i - средние арифметические, зафиксированные при i -м уровне фактора соответственно (групповые средние); \tilde{x} - среднее арифметическое всех mn наблюдений (общая средняя); Q^2 - сумма квадратов отклонений наблюдаемых значений от общей средней; Q_1^2 - сумма квадратов отклонений групповых средних от общей средней; Q_2^2 - сумма квадратов отклонений внутригрупповых значений от групповых средних. Между величинами Q^2 , Q_1^2 , Q_2^2 выполняется соотношение:

$$Q^2 = Q_1^2 + Q_2^2. \quad (6.3)$$

Составляющая общей вариации Q_1^2 отражает часть общей вариации, формируемую под влиянием исследуемого фактора, а составляющая Q_2^2 отражает часть общей вариации, формируемую под влиянием всех остальных, неучитываемых факторов.

Для проверки гипотезы (6.1) рассчитывается статистика:

$$F = \frac{Q_1^2 / (m - 1)}{Q_2^2 / m(n - 1)}, \quad (6.4)$$

которая при нормальном распределении данных наблюдений при первом, втором, ..., m - м уровне фактора будет подчиняться распределению Фишера с $(m - 1; m(n - 1))$ степенями свободы. Если расчетное значение величины F

больше правосторонней критической границы распределения Фишера при уровне значимости α , т. е.:

$$F > \bar{F}_{\alpha(m-1; m(n-1))}, \quad (6.5)$$

то нулевая гипотеза (6.1) отвергается и считается, что исследуемый фактор значимо влияет на изменение значений результирующего показателя.

Многофакторный ДА. Влияние на результирующий показатель двух или более факторов устанавливают, проводя процедуру многофакторного ДА. При этом вариация результирующего признака раскладывается на компоненты, отражающие влияние на нее исследуемых и неучтенных факторов.

Так, например, в случае двухфакторного ДА исходные данные можно фиксировать на m уровнях фактора А и на n уровнях фактора В. Общее количество групп наблюдений равно mn . Для упрощения в каждой группе зафиксировано по одному наблюдению и, следовательно, общее число наблюдений $N = mn$. Исходные данные могут быть сведены в таблицу 6.2.

Таблица 6.2

Исходные данные для проведения двухфакторного ДА

| | | Группа наблюдений по фактору В, $j = 1, \dots, n$ | | | | |
|---|----------|---|----------|----------|----------|----------|
| | | 1 | 2 | 3 | ... | n |
| Группа наблюдений по фактору А, $i = 1, \dots, m$ | 1 | x_{11} | x_{12} | x_{13} | ... | x_{1n} |
| | | x_{21} | x_{22} | x_{23} | ... | x_{2n} |
| | \vdots | \vdots | \vdots | \vdots | \vdots | \vdots |
| | m | x_{m1} | x_{m2} | x_{m3} | | x_{mn} |

При проведении двухфакторного ДА проверяют две статистические гипотезы:

$$H_0 : a_1 = a_2 = \dots = a_m; \text{ против } H_1 : a_1 \neq a_2 \neq \dots \neq a_m; \quad (6.6)$$

$$H_0 : a_1 = a_2 = \dots = a_n; \text{ против } H_1 : a_1 \neq a_2 \neq \dots \neq a_n, \quad (6.7)$$

где a_1, a_2, \dots, a_m - генеральные средние значения результирующего признака, зафиксированные при первом, втором, ..., m - м уровне фактора А, а a_1, a_2, \dots, a_n - генеральные средние значения результирующего признака, зафиксированные при первом, втором, ..., n - м уровне фактора В.

Для проверки гипотез (6.6-6.7) рассчитывают величины:

$$Q^2 = \sum_{i=1}^m \sum_{j=1}^n (x_{ij} - \tilde{x})^2; \quad Q_1^2 = n \sum_{i=1}^m (\tilde{x}_{i\bullet} - \tilde{x})^2; \quad Q_2^2 = m \sum_{j=1}^n (\tilde{x}_{\bullet j} - \tilde{x})^2; \\ Q_3^2 = \sum_{i=1}^m \sum_{j=1}^n (x_{ij} - \tilde{x}_{i\bullet} - \tilde{x}_{\bullet j} + \tilde{x})^2, \quad (6.8)$$

где $\tilde{x}_{i\bullet}, i = 1, \dots, m$, $\tilde{x}_{\bullet j}, j = 1, \dots, n$ - средние арифметические, зафиксированные при i - м уровне фактора А, и средние арифметические, зафиксированные при j - м уровне фактора В соответственно (групповые средние по факторам А и В); \tilde{x} - среднее арифметическое по всем mn наблюдениям (общая средняя); Q^2 - сумма квадратов отклонений наблюдаемых значений от общей средней; Q_1^2, Q_2^2 - сумма квадратов отклонений групповых средних по факторам А и В от общей средней; Q_3^2 - остаточная сумма квадратов отклонений. Причем между величинами Q^2, Q_1^2, Q_2^2, Q_3^2 выполняется соотношение:

$$Q^2 = Q_1^2 + Q_2^2 + Q_3^2. \quad (6.9)$$

Для проверки гипотез (6.6-6.7) рассчитывают величины:

$$F_A = \frac{Q_1^2 / m - 1}{Q_3^2 / ((m-1)(n-1))}; \quad F_B = \frac{Q_2^2 / n - 1}{Q_3^2 / ((m-1)(n-1))}, \quad (6.10)$$

подчиняющиеся распределениям Фишера с $(m-1; (m-1)(n-1))$ степенями свободы, и $(n-1; (m-1)(n-1))$ степенями свободы соответственно и, следовательно, если F_A или F_B больше правосторонних критических границ соответствующих распределений при уровне значимости α :

$$F_A > \bar{F}_{\alpha(m-1; (m-1)(n-1))}; \quad F_B > \bar{F}_{\alpha(n-1; (m-1)(n-1))}, \quad (6.11)$$

то гипотезу (6.6) или (6.7) необходимо отвергнуть и считать, что соответствующий фактор значимо влияет на исследуемый признак.

Выполнение работы.

1. Процедура однофакторного ДА.

Рассмотрим процедуру проведения ДА на примере данных, представленных в таблице 6.3. Данные представляют собой значения выручки, полученной в четырех торговых точках, расположенных в четырех районах города за четыре дня недели. Условно районы города назовем «Черемушки», «Центр», «Сельмаш», «Домики». Проверяется влияние расположения торговой точки на объем выручки.

Таблица 6.3

Исходные данные для проведения ДА

| | | Объем реализации, т. руб. (наблюдения) | | | |
|--|-------------|--|------|------|------|
| | | Пн. | Вт. | Ср. | Чт. |
| Район расположения торговой точки (уровень фактора) | «Черемушки» | 14 | 15 | 14,8 | 15 |
| | «Центр» | 14,4 | 14,9 | 14,9 | 15,5 |
| | «Сельмаш» | 14,2 | 15,2 | 14,6 | 15,4 |
| | «Домики» | 14,5 | 15 | 14,7 | 15,2 |

Запустите систему STATISTICA. Создайте файл размерности 3×16 . Первая переменная файла будет предназначена для фиксации района расположения торговой точки, вторая – для фиксации дня недели (понадобится в дальнейшем для проведения двухфакторного ДА), третья переменная – для фиксации выручки. Заполните электронную таблицу, как показано на рис. 6.1.

В пункте меню «Статистика» выберите пункт ANOVA (от *analyze of variance*), который предназначен для проведения процедуры как однофакторного, так и многофакторного ДА. В результате будет открыто стартовое окно модуля (рис. 6.2).

| Data: Исходные_данные_л_р_6 (3v by 16c) | | | |
|---|--|---------------------|--------------|
| | 1 Расположение торговых точек | 2 День недели | 3 Выручка |
| 1 | Черемушк | Пн | 14,0 |
| 2 | Черемушк | Вт | 15,0 |
| 3 | Черемушк | Ср | 14,8 |
| 4 | Черемушк | Чт | 15,0 |
| 5 | Центр | Пн | 14,4 |
| 6 | Центр | Вт | 14,9 |
| 7 | Центр | Ср | 14,9 |
| 8 | Центр | Чт | 15,5 |
| 9 | Сельмаш | Пн | 14,2 |
| 10 | Сельмаш | Вт | 15,2 |
| 11 | Сельмаш | Ср | 14,6 |
| 12 | Сельмаш | Чт | 15,4 |
| 13 | Домики | Пн | 14,5 |
| 14 | Домики | Вт | 15,0 |
| 15 | Домики | Ср | 14,7 |
| 16 | Домики | Чт | 15,2 |

Рис. 6. 1. Исходные данные для проведения ДА

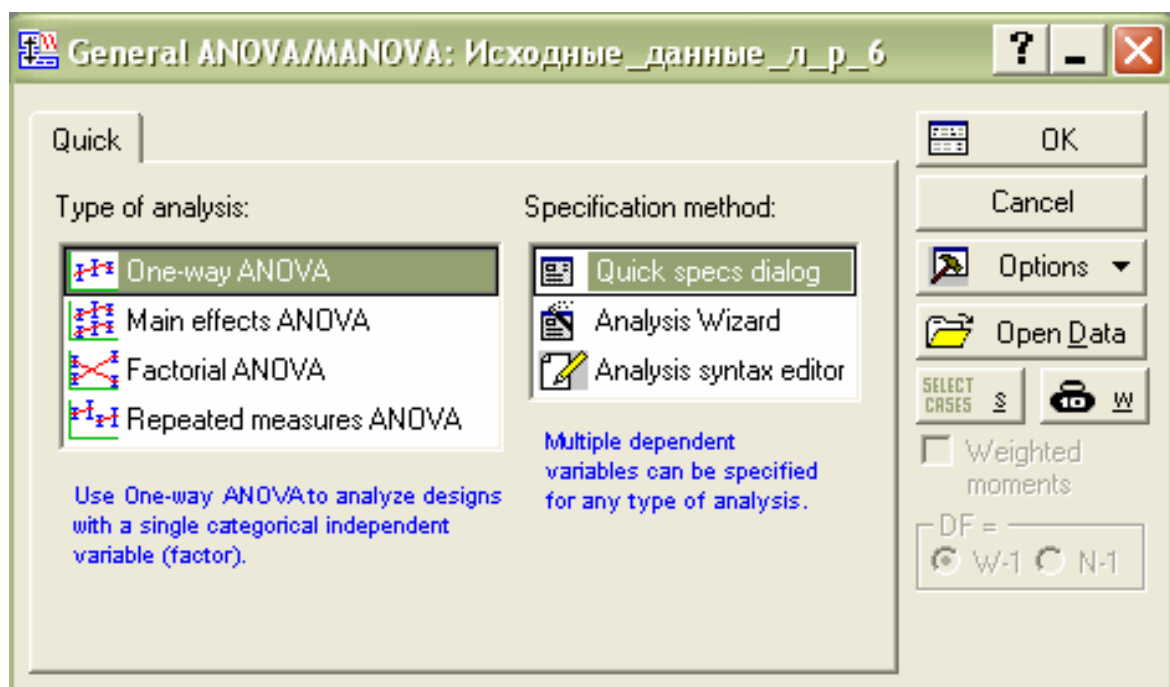




Рис. 6.2. Стартовое окно модуля ANOVA

Для начала проведем процедуру однофакторного ДА. Для этого установите тип анализа  One-way ANOVA и метод  Quick specs dialog. Нажмите «ОК». Откроется окно установки переменных и опций анализа (рис. 6.3). Факторной переменной будем считать переменную «Расположение торговых точек», а результирующей переменной - переменную «Выручка».

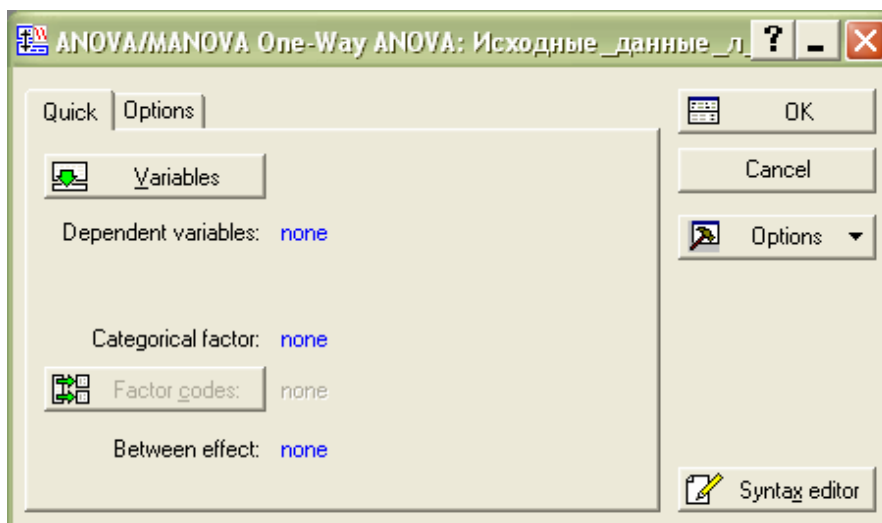



Рис. 6.3. Окно установки переменных и опций анализа

Нажмите кнопку  Variables. В открывшемся окне установите результирующую переменную «Выручка» и факторную переменную «Расположение торговых точек» так, как показано на рисунке 6.4, и нажмите «ОК».

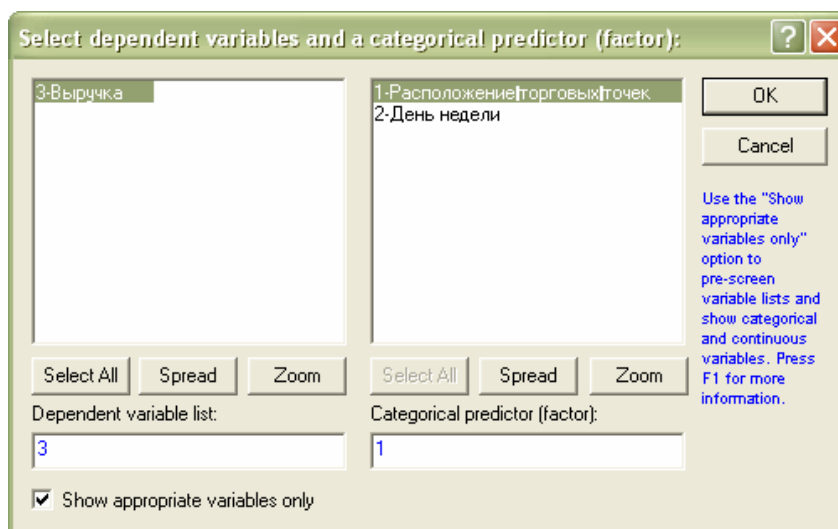




Рис. 6.4. Окно выбора переменных для анализа

Нажмите «OK» и затем нажмите кнопку  Factor codes, в открывшемся окне нажмите кнопку . Окно должно выглядеть так, как показано на рисунке 6.5. Нажмите «OK».

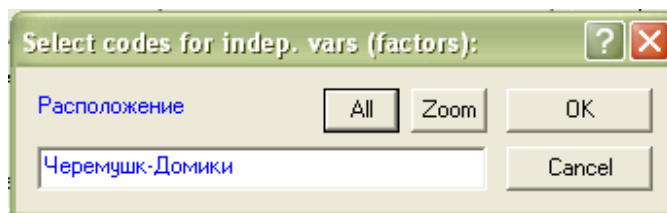
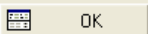


Рис. 6.5. Окно установки уровней факторной переменной

В окне выбора переменных и опций анализа нажмите . В результате будет открыто окно результатов анализа (рис. 6.6).

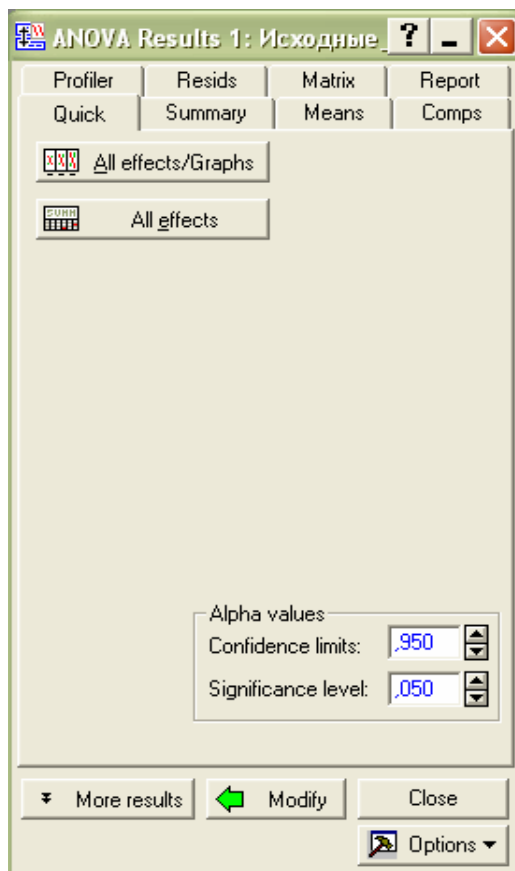
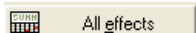


Рис. 6.6. Окно результатов анализа

Оставьте уровень значимости «Significance level» равным $\alpha = 0,05$ и нажмите кнопку . Будет открыто окно с результатами расчета

компонент вариации и проверки гипотезы (6.1) с использованием статистики Фишера (рис. 6.7). Проведем анализ полученных результатов (при анализе игнорируем строку «Intercept»).

| Компоненты вариации, обусловленные | SS Сумма квадратов | Число степеней свободы | MS Средний квадрат | F Статистика | p |
|------------------------------------|-----------------------|------------------------|-----------------------|-----------------|----------|
| Intercept | 3519,456 | 1 | 3519,456 | 16978,28 | 0,000000 |
| Расположением торговых точек | 0,107 | 3 | 0,036 | 0,17 | 0,913350 |
| Неучтенными факторами | 2,488 | 12 | 0,207 | | |

Рис. 6.7. Окно результатов однофакторного ДА

1. Система рассчитала компоненты «Сумма квадратов» (SS) $Q_1^2 = 0,107$ и $Q_2^2 = 2,488$ общей вариации Q^2 (см. формулу 6.3), обусловленные фактором «Расположение торговых точек» и неучтенными факторами.

2. Определила число степеней свободы (Degr. of Freedom) каждой компоненты. В частности, для компоненты вариации, обусловленной факторной переменной, число степеней свободы равно $m - 1 = 3$, а для компоненты, обусловленной неучтенными факторами, число степеней свободы равно $m(n - 1) = 12$.

3. Рассчитала величины «Средний квадрат» (MS) $Q_1^2 / (m - 1) = 0,036$ и $Q_2^2 / m(n - 1) = 0,207$ соответственно.

4. Рассчитала величину статистики Фишера, равную $F = \frac{Q_1^2 / (m - 1)}{Q_2^2 / m(n - 1)} = 0,17$ (см. формулу 6.4), и определила вероятность

$p = 0,913350$ того, что при справедливости гипотезы H_0 (6.1) величина F

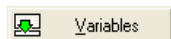
может принять такие же или большие значения. Так как эта вероятность существенно больше уровня значимости $\alpha = 0,05$, то гипотезу о значимости влияния фактора «Расположение торговых точек» на объемы реализации можно отвергнуть.

2. Процедура многофакторного ДА (на примере двухфакторного).

Перейдите в стартовое окно модуля ANOVA. Установите тип анализа



и нажмите «OK». В открывшемся окне нажмите кнопку



и установите результирующую переменную и факторы так, как показано на рисунке 6.8.

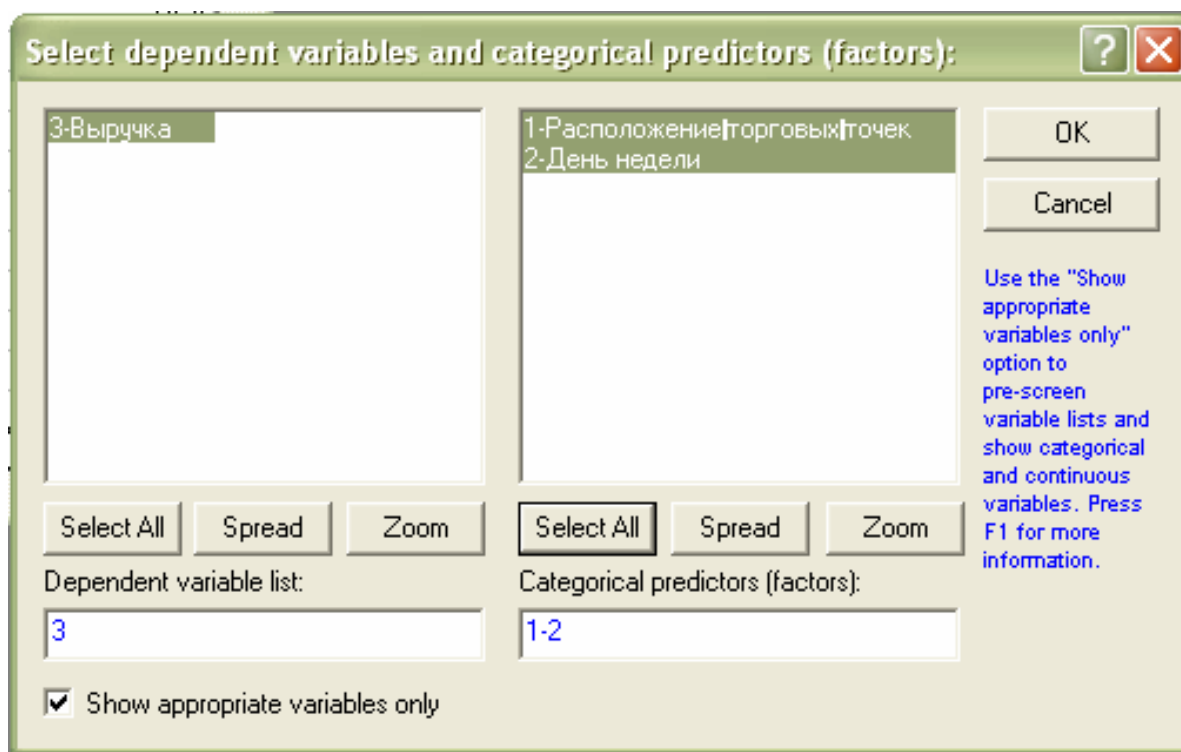


Рис. 6.8. Окно выбора переменных для анализа

Нажмите «OK» и затем кнопку  Factor codes:




нажмите кнопки  напротив обозначения обоих факторов. Окно должно выглядеть так, как показано на рисунке 6.9.



Рис. 6.9. Окно установки уровней факторных переменных

Нажмите «OK». В окне установки переменных и опций анализа также нажмите  OK. В результате будет открыто окно результатов анализа, аналогичное представленному на рисунке 6.6. Оставьте уровень значимости «Significance level» равным $\alpha = 0,05$ и нажмите кнопку  All effects. Будет открыто окно с результатами расчета компонент вариации и проверки гипотез (6.6-6.7) с использованием статистик Фишера (рис. 6.10). Проведем анализ полученных результатов (при анализе также игнорируем строку «Intercept»).

| Компоненты вариации, обусловленные | SS Сумма квадратов | Число степеней свободы | MS Средний квадрат | F Статистика | p |
|------------------------------------|-----------------------|------------------------|-----------------------|-----------------|----------|
| Intercept | 3519,456 | 1 | 3519,456 | 110897,5 | 0,000000 |
| Расположением торговых точек | 0,107 | 3 | 0,036 | 1,1 | 0,390317 |
| Днем недели | 2,202 | 3 | 0,734 | 23,1 | 0,000145 |
| Неучтенными факторами | 0,286 | 9 | 0,032 | | |

Рис. 6.10. Окно результатов двухфакторного ДА

1. Система рассчитала компоненты «Сумма квадратов» $Q_1^2 = 0,107$ и $Q_2^2 = 2,202$ и $Q_3^2 = 0,286$ общей вариации Q^2 (6.9), обусловленные выбранными и неучтенными факторами.

2. Определила число степеней свободы каждой компоненты. В частности, для компоненты вариации, обусловленной переменной «Расположение торговых точек», число степеней свободы равно $m - 1 = 3$, для компоненты, обусловленной переменной «День недели», число степеней свободы равно $n - 1 = 3$, а для компоненты, обусловленной неучтенными факторами, число степеней свободы равно $(m - 1)(n - 1) = 9$.

3. Рассчитала величины «Средний квадрат» $Q_1^2 / (m - 1) = 0,036$, $Q_2^2 / n - 1 = 0,734$ и $Q_3^2 / ((m - 1)(n - 1)) = 0,032$ соответственно.

4. Рассчитала значения статистик Фишера, равные $F_A = \frac{Q_1^2 / (m - 1)}{Q_3^2 / ((m - 1)(n - 1))} = 1,1$ и $F_B = \frac{Q_2^2 / (n - 1)}{Q_3^2 / ((m - 1)(n - 1))} = 23,1$ (6.10), и определила соответствующие значениям F_A и F_B вероятности $p_A = 0,390317$ и $p_B = 0,000145$ того, что при справедливости гипотез H_0 (6.6-6.7) величины F_A и F_B могут принять такие же или большие значения.

Так как вероятность p_A существенно больше, а вероятность p_B существенно меньше уровня значимости $\alpha = 0,05$, то гипотезу о значимости влияния фактора «Расположение торговых точек» на объемы реализации можно отвергнуть, а гипотезу о значимости влияния фактора «День недели» на объемы реализации необходимо принять.

Задание для самостоятельной работы.

Получите вариант исходных данных для выполнения самостоятельной работы у преподавателя. Выполните процедуры однофакторного и двухфакторного ДА по своему варианту исходных данных.

3. Оформление отчета о лабораторной работе.

Отчет о лабораторной работе №6 должен содержать:

- постановку задачи;
- электронную таблицу, заполненную в соответствии с индивидуальным вариантом исходных данных;

– таблицы с результатами проведения процедур однофакторного и двухфакторного ДА с соответствующими выводами относительно наличия или отсутствия влияния факторов на результирующий признак.

7. Лабораторная работа № 7

Тема работы: Корреляционный и регрессионный анализ (2 часа).

Цель работы: Освоение методов анализа парной корреляции и регрессии.

Содержание работы:

1. Создание электронной таблицы с исходными данными.
2. Анализ парной корреляции.
3. Построение и анализ уравнения регрессии.
4. Оформление отчета по лабораторной работе.

Выполнение работы.

1. Создание электронной таблицы с исходными данными.

В системе STATISTICA создайте файл размерности 2×20 и заполните электронную таблицу в соответствии с исходными данными, представленными в таблице 7.1. Сохраните файл и добавьте исходные данные в отчет по лабораторной работе.

Таблица 7.1


Исходные данные для проведения корреляционно-регрессионного анализа

| Год | Личный доход - x_i , млрд \$ | Расходы на мед. услуги - y_i , млрд \$ | Год | Личный доход - x_i , млрд \$ | Расходы на мед. услуги - y_i , млрд \$ |
|------|--------------------------------------|--|------|--------------------------------------|--|
| 1964 | 658,0 | 11,9 | 1974 | 1004,8 | 17,6 |
| 1965 | 700,4 | 12,1 | 1975 | 1010,8 | 17,9 |
| 1966 | 740,6 | 12,1 | 1976 | 1056,2 | 18,0 |
| 1967 | 774,4 | 12,5 | 1977 | 1105,4 | 19,2 |
| 1968 | 816,2 | 12,8 | 1978 | 1162,3 | 18,6 |
| 1969 | 853,5 | 13,6 | 1979 | 1200,7 | 20,1 |
| 1970 | 876,8 | 14,4 | 1980 | 1209,5 | 21,5 |
| 1971 | 900,0 | 14,8 | 1981 | 1248,6 | 22,0 |
| 1972 | 951,4 | 15,7 | 1982 | 1254,4 | 22,4 |
| 1973 | 1007,9 | 16,9 | 1983 | 1284,6 | 23,3 |

Данные, представленные в таблице 7.1, представляют собой значения личного дохода домохозяйств США и их расходов на медицинские услуги в период с 1964 по 1983 гг. Предполагается методами корреляционного

анализа установить наличие корреляционной связи между данными статистическими признаками, а методами регрессионного анализа формализовать (т.е. представить в виде уравнения) данную связь. В системе STATISTICA указанные задачи будут решаться с использованием двух модулей «Basic Statistics/Tables» (Основные статистики/Таблицы) и «Nonlinear estimation» (Нелинейное оценивание).

2. Анализ парной корреляции.

После создания и сохранения электронной таблицы с данными в меню Statistics выберите пункт «Basic Statistics/Tables». В стартовом окне модуля выберите пункт  Correlation matrices - корреляционные матрицы. Будет открыто окно модуля корреляционного анализа (рисунок 7.1).

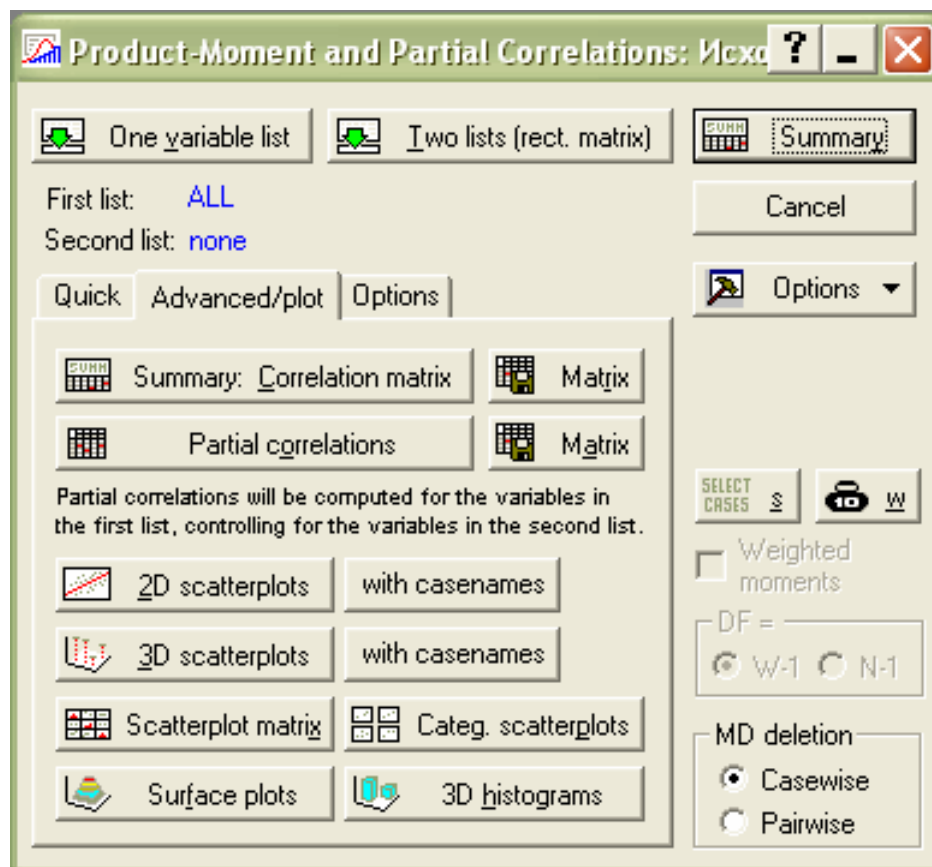



Рис. 7.1. Стартовое окно модуля корреляционного анализа

Нажмите кнопку  One variable list и в открывшемся окне выберите обе переменные для анализа (рисунок 7.2), нажмите «ОК».

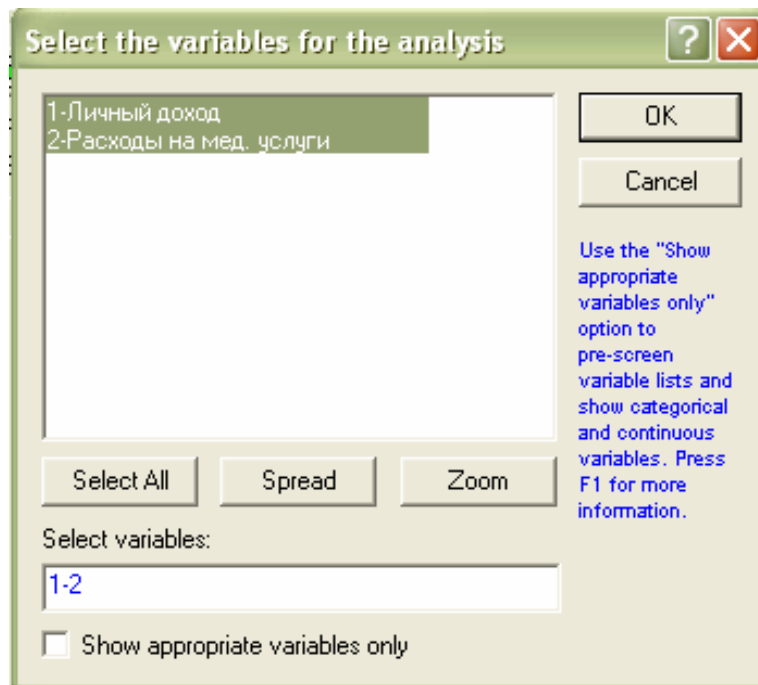


Рис. 7.2. Окно выбора переменных для корреляционного анализа

После выбора переменных для анализа перейдите на вкладку «Options» (Опции) и установите форму отображения корреляционной матрицы «Display simple matrices» (Отображать только матрицу), рисунок 7.3.

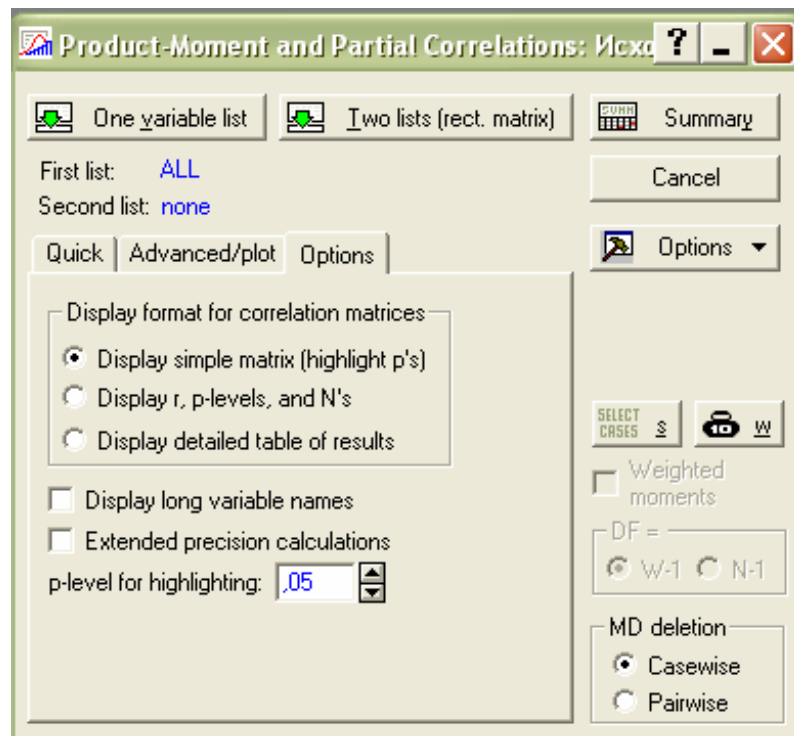
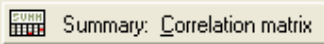
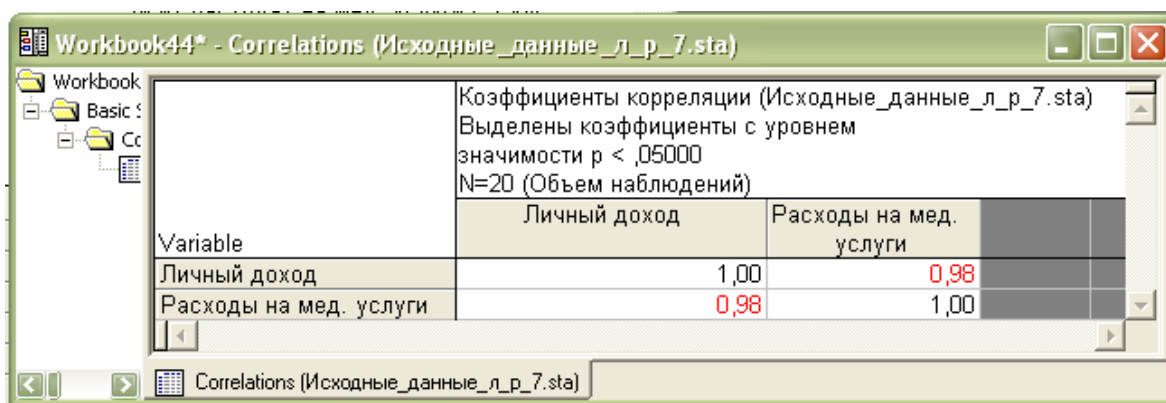


Рис. 7.3. Установка формы отображения корреляционной матрицы

Вернитесь на вкладку «Advanced/plot». Нажмите  (Итог. Корреляционная матрица). Система произведет требуемые расчеты и представит итог в виде таблицы (рисунок 7.4).



Workbook44* - Correlations (Исходные_данные_л_р_7.sta)

Кoeffициенты корреляции (Исходные_данные_л_р_7.sta)
Выделены коэффиценты с уровнем значимости $p < .05000$
N=20 (Объем наблюдений)

| Variable | Личный доход | Расходы на мед. услуги |
|------------------------|--------------|------------------------|
| Личный доход | 1,00 | 0,98 |
| Расходы на мед. услуги | 0,98 | 1,00 |



Correlations (Исходные_данные_л_р_7.sta)

Рис. 7.4. Результаты расчета корреляционной матрицы

Анализ полученных результатов позволяет установить, что выборочный коэффициент корреляции между признаками «Личный доход» и «Расходы на мед. услуги» равен $\hat{\rho}_{xy} = 0,98$. Его значение является значимым (в таблице результатов значение коэффициента корреляции выделено красным цветом). Добавьте результаты расчета в отчет.

Примечание: STATISTICA наряду с оценкой коэффициента корреляции осуществляет и подробную оценку его значимости (рассчитывая соответствующую статистику Стьюдента). Для этого необходимо при выборе опций отражения корреляционной матрицы выбрать пункт «Display detailed table of results» (Отражать таблицу результатов детально).

3. Построение и анализ уравнения регрессии.

Закройте модуль анализа «Basic Statistics/Tables». В меню Statistics последовательно выберите пункты  (Дополнительные линейные/нелинейные модели) и  (Нелинейное оценивание). В открывшемся окне установите положение курсора в соответствии с рисунком 7.5.

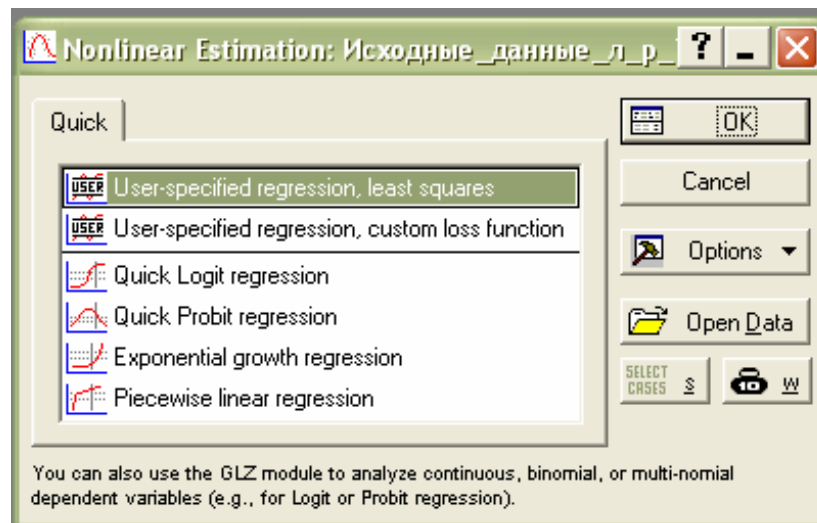



Рис. 7.5. Стартовое окно модуля нелинейное оценивание

Вы указали, что модель связи между факторной и результирующей переменной будет определяться пользователем «User-specified regression», а коэффициенты будут оцениваться по методу наименьших квадратов «least squares». Нажмите «OK». В открывшемся окне нажмите  , после чего введите уравнение линейной регрессии в поле «Estimated function:» так, как показано на рисунке 7.6.

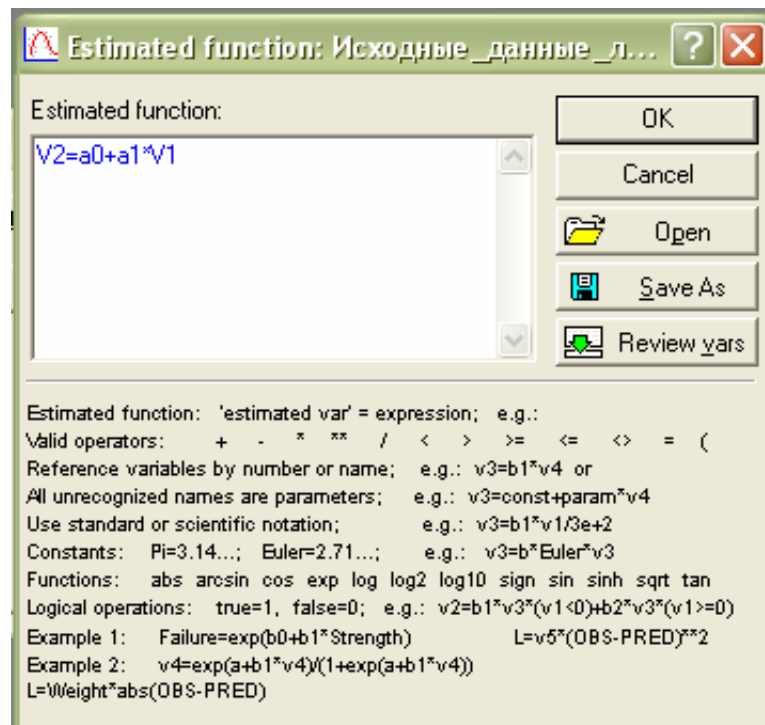


Рис. 7.6. Окно ввода уравнения модели

Дважды нажмите «ОК» и перейдите к окну задания условий расчета (рисунок 7.7), в котором будут описаны заданные вами условия.

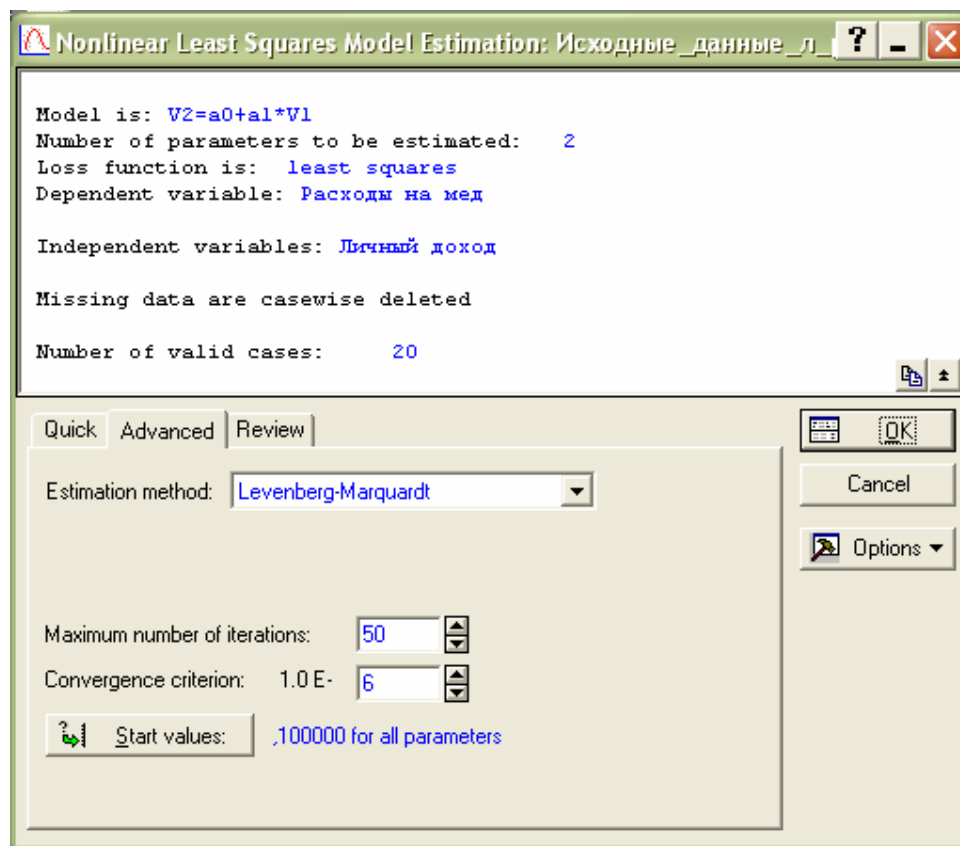


Рис. 7.7. Окно задания условий расчета

В соответствии с рисунком 7.7 заданы следующие условия проведения регрессионного анализа:

- модель для расчета - $V2 = a_0 + a_1 \cdot V1$;
- количество оцениваемых параметров – 2 (параметры a_0 и a_1);
- метод оценки параметров модели – метод наименьших квадратов;
- результирующая (Dependent) переменная – Расходы на мед. услуги;
- факторная (Independent) переменная – Личный доход;
- количество наблюдений (объем выборки) – 20.

Оставьте все параметры без изменений и нажмите «ОК». Система произведет расчет необходимых величин и представит панель результатов расчета (рисунок 7.8).

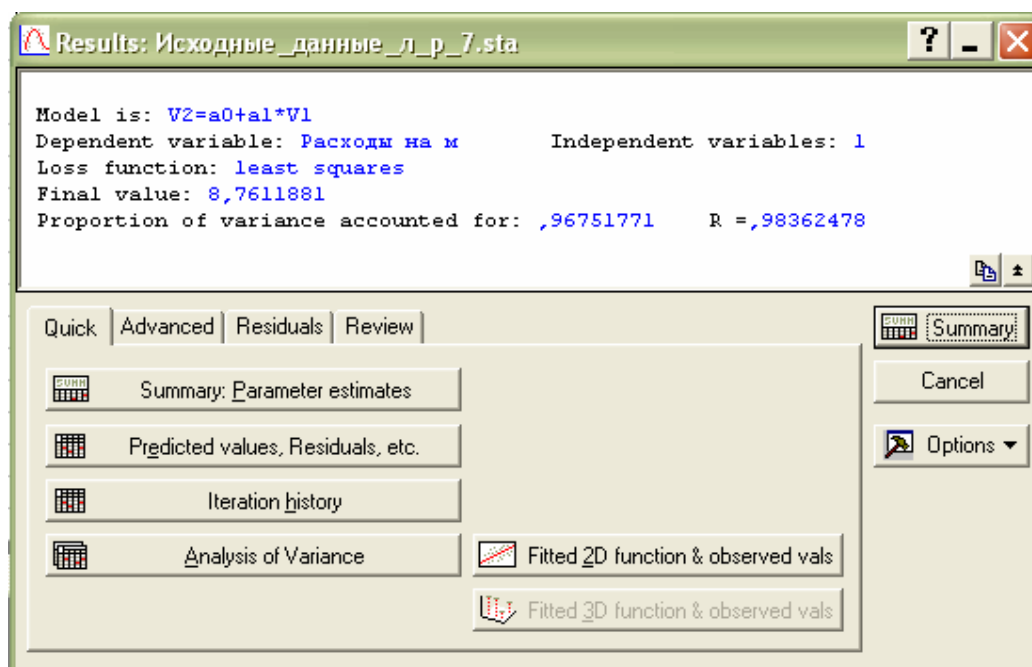
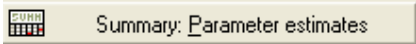


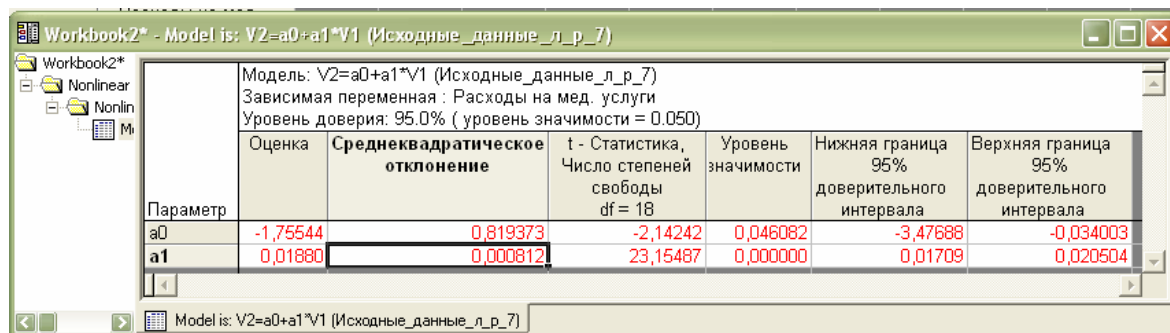
Рис. 7.8. Панель результатов расчета

На заголовке окна еще раз представлены:

- модель связи (уравнение регрессии) переменных;
- название зависимой и номер независимой переменных;
- метод оценки параметров - метод наименьших квадратов;
- значение минимума суммы квадратов отклонений расчетных значений результирующего признака от его наблюдаемых значений – 8,7611881;
- коэффициент детерминации - $R^2 = 0,9675$ и значение квадратного корня из коэффициента детерминации $R = 0,9836$, которое в случае парной линейной регрессии равно значению коэффициента корреляции (сравните с уже полученными результатами).

Нажмите кнопку  (Итог: Оценки параметров). Результаты оценивания параметров уравнения регрессии,


среднеквадратические отклонения параметров, t - статистики для параметров, уровни значимости параметров и границы 95% доверительных интервалов для параметров будут приведены в таблице, аналогичной представленной на рисунке 7.9.




| Параметр | Оценка | Среднеквадратическое отклонение | t - Статистика, Число степеней свободы df = 18 | Уровень значимости | Нижняя граница 95% доверительного интервала | Верхняя граница 95% доверительного интервала |
|----------|----------|---------------------------------|--|--------------------|---|--|
| a0 | -1,75544 | 0,819373 | -2,14242 | 0,046082 | -3,47688 | -0,034003 |
| a1 | 0,01880 | 0,000812 | 23,15487 | 0,000000 | 0,01709 | 0,020504 |

Рис. 7.9. Результаты оценки параметров уравнения регрессии переменной «Расходы на мед. услуги» на переменную «Личный доход»

Анализируя полученные данные, можно сделать вывод о значимости обоих параметров уравнения регрессии, поскольку уровень значимости каждого из параметров меньше установленного критического уровня значимости $\alpha = 0,05$.

Добавьте полученную таблицу в отчет и закройте окно результатов оценки параметров уравнения регрессии. На панели результатов расчета (см. рисунок 7.8.) нажмите кнопку  Fitted 2D function & observed vals (Подобранная функция и наблюдаемые значения). STATISTICA осуществит построение графика линейной регрессии на корреляционном поле значений переменных «Личный доход» и «Расходы на мед. услуги» (рисунок 7.10).

После добавления в отчет закройте график. На панели результатов расчета нажмите кнопку  Predicted values, Residuals, etc. (Расчетные значения, остатки и т.д.) и просмотрите расчетные значения величины «Расходы на мед. услуги» в сравнении с ее наблюдаемыми значениями, а также разность (остатки) между наблюдаемыми и расчетными значениями (рисунок 7.11). Просмотрите результаты, добавьте таблицу в отчет и закройте окно.

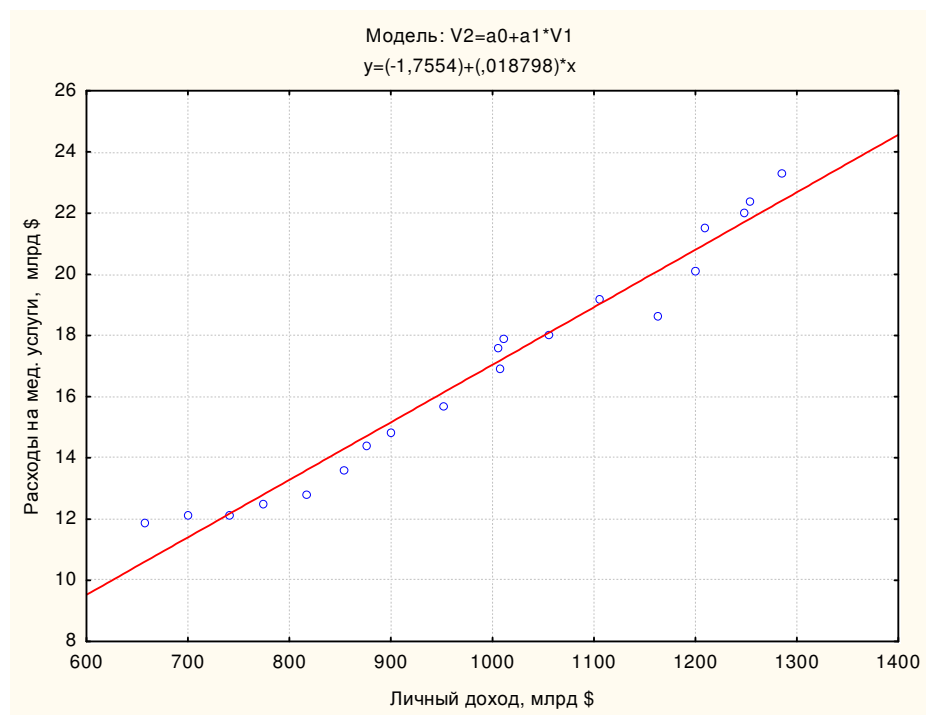


Рис. 7.10. График линейной регрессии переменной «Расходы на мед. услуги» на переменную «Личный доход»

Workbook4* - Model is: $V2=a0+a1*V1$ (Исходные_данные_л_п_7)

Модель: $V2=a0+a1*V1$ (Исходные_данные_л_п_7)
Зависимая переменная: Расходы на мед. услуги

| | Наблюдаемые значения | Расчетные значения | Остатки |
|------|----------------------|--------------------|----------|
| 1964 | 11,90000 | 10,61358 | 1,28642 |
| 1965 | 12,10000 | 11,41062 | 0,68938 |
| 1966 | 12,10000 | 12,16629 | -0,06629 |
| 1967 | 12,50000 | 12,80166 | -0,30166 |
| 1968 | 12,80000 | 13,58741 | -0,78741 |
| 1969 | 13,60000 | 14,28858 | -0,68858 |
| 1970 | 14,40000 | 14,72657 | -0,32657 |
| 1971 | 14,80000 | 15,16268 | -0,36268 |
| 1972 | 15,70000 | 16,12889 | -0,42889 |
| 1973 | 16,90000 | 17,19097 | -0,29097 |
| 1974 | 17,60000 | 17,13270 | 0,46730 |
| 1975 | 17,90000 | 17,24549 | 0,65451 |
| 1976 | 18,00000 | 18,09891 | -0,09891 |
| 1977 | 19,20000 | 19,02377 | 0,17623 |
| 1978 | 18,60000 | 20,09337 | -1,49337 |
| 1979 | 20,10000 | 20,81521 | -0,71521 |
| 1980 | 21,50000 | 20,98063 | 0,51937 |
| 1981 | 22,00000 | 21,71563 | 0,28437 |
| 1982 | 22,40000 | 21,82466 | 0,57534 |
| 1983 | 23,30000 | 22,39236 | 0,90764 |

Model is: $V2=a0+a1*V1$ (Исходные_данные_л_п_7)

Рис. 7.11. Наблюдаемые, расчетные значения и остатки для переменной «Расходы на мед. услуги»

Примечание: STATISTICA позволяет осуществить и более детальный анализ уравнения регрессии. Например, провести дисперсионный анализ и рассчитать значение критерия Фишера, которое позволяет определить уровень значимости всего уравнения регрессии в целом. Провести анализ распределения остатков, анализ коррелированности параметров уравнения регрессии и т.д. Для проведения такого анализа необходимо на панели результатов расчета выбирать соответствующие пункты.

Задание на самостоятельную работу.

Получите вариант исходных данных для выполнения самостоятельной работы у преподавателя. Выполните процедуры расчета корреляционной матрицы и постройте график уравнения линейной регрессии по своему варианту исходных данных. Оформите отчет о лабораторной работе.

4. Оформление отчета по лабораторной работе:

Отчет по лабораторной работе №7 должен содержать:

- постановку задачи;
- электронную таблицу, заполненную в соответствии с вариантом исходных данных;
- результаты расчета корреляционной матрицы;
- результаты регрессионного анализа связи между результирующей и факторной переменной (таблица расчета параметров модели, таблица наблюдаемых, расчетных значений и значений остатков для результирующей переменной);
- график уравнения регрессии результирующей переменной на факторную переменную.

Правила оформления отчета по лабораторной работе

Отчет по лабораторной работе выполняется и представляется на рецензию преподавателю в виде принтерной распечатки, напечатанной шрифтом "Times New Roman" 14 кегль через 1,5 интервала (40 строк на листе) на листах белой бумаги формата А4 (210×297 мм).

Текст отчета печатают на одной стороне листа с полями: слева – 30 мм, справа – 15 мм, сверху и снизу по 20 мм. Отчет по лабораторной работе должен иметь титульный лист, оформленный в соответствии с приложением.

Текст отчета может включать обязательные и необязательные элементы. Обязательные элементы перечислены в конце каждой лабораторной работы. Все остальные элементы, приводимые автором как пояснения или теоретический материал, являются необязательными.

Иллюстрации (рисунки) и таблицы располагают в тексте отчета после первого упоминания о них. Все таблицы и рисунки нумеруются (Рисунок 1, Таблица 5). Каждый рисунок сопровождают содержательной подписью, которую располагают справа от слова "Рисунок". Название таблицы указывается на строке, следующей за строкой с номером таблицы. Рисунки должны быть выполнены средствами системы «STATISTICA». Не допускается выполнение рисунков «от руки». В таблицах и на графиках все обозначения и названия должны быть представлены на русском языке.

Ссылки на использованную литературу даются в квадратных скобках с указанием номера по списку использованной литературы (например, [2], [5] и т. д.). Нумерация страниц в пояснительной записке дается сквозной. Список литературы включается в общую нумерацию страниц.

Работы, выполненные без соблюдения указанных требований, не оцениваются и возвращаются студенту на доработку.

Образец оформления титульного листа

Министерство образования и науки Российской Федерации
Федеральное агентство по образованию
Рубцовский индустриальный институт (филиал) ГОУ ВПО
«Алтайский государственный технический университет им. И.И. Ползунова»

Кафедра «Менеджмент и экономика»

Отчет по лабораторной работе №_ по курсу
"СТАТИСТИКА"

Тема работы «.....»

Выполнил:

| | | |
|---------------|--------|---------|
| Студент _____ | _____ | _____ |
| группа | Ф.И.О. | Подпись |

Проверил:

| | | |
|------------------------|-------|---------------|
| _____ | _____ | _____ |
| должность, уч. степень | Ф.И.О | Подпись, дата |

Рубцовск 200_

Шашок Александр Васильевич

Лабораторный практикум в системе «STATISTICA»

Методические указания к выполнению лабораторных работ по дисциплине
«Статистика» для студентов всех форм обучения специальности 080507
«Менеджмент организации»

Редактор Е.Ф. Изотова

Подготовка оригинала-макета – А.В. Шашок

Подписано к печати Формат 84х108 /16.

Усл. печ. л. 5,06. Тираж 100 экз. Зак. Рег. №

Отпечатано в РИО Рубцовского индустриального института
658207, Рубцовск, ул. Тракторная, 2/6.