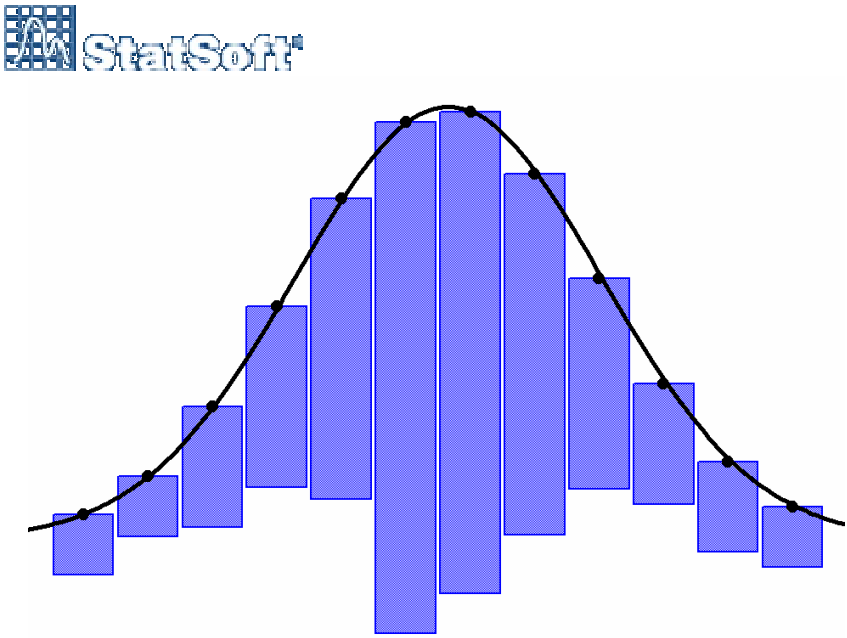
Федеральное агентство по образованию

САНКТ-ПЕТЕРБУРГСКИЙ
ГОСУДАРСТВЕННЫЙ ПОЛИТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ

Н.В.Куприенко, О.А.Пономарева, Д.В. Тихонов

СТАТИСТИКА
Методы анализа распределений
Выборочное наблюдение

Учебное пособие



Санкт-Петербург
Издательство Политехнического университета
2009

УДК 681.322.068:519.23 (075.8)
ББК 65.051я73
К 924

Н. В. Куприенко **Статистика. Методы анализа распределений. Выборочное наблюдение. 3-е изд. :** учеб. пособие. / Н. В. Куприенко, О. А. Пономарева, Д. В. Тихонов. – СПб.: Изд-во Политехн. ун-та, 2009. – 138 с.

В учебном пособии рассматриваются возможности использования пакета прикладных программ (ППП) STATISTICA для реализации статистических методов анализа эмпирических распределений и проведения выборочного статистического наблюдения в объеме, достаточном для решения широкого круга практических задач. Рекомендуется студентам факультета экономики и менеджмента дневного и вечернего отделений, изучающих дисциплину «Статистика». Пособие может быть использовано студентами - дипломниками, аспирантами, научными и практическими работниками, столкнувшимися с необходимостью использования статистических методов обработки исходных данных. Пособие содержит сведения по ППП STATISTICA, не публиковавшиеся на русском языке.

Табл. 8 Рис. 122 Библиогр.: 15 назв.

Печатается по решению редакционно-издательского совета Санкт-Петербургского государственного политехнического университета.

Содержание

Введение

Принятие управленческих решений любого уровня не возможно без полной и точной информации об объекте управления, о среде, его окружающей. Статистические методы анализа помогают извлечь информацию из данных и оценить качество этой информации. Именно методы статистики позволяют выявить и охарактеризовать закономерности распределений, закономерности связей и зависимостей, закономерности развития тех или иных объектов или процессов. В настоящее время, в связи с внедрением компьютерных технологий и распространением статистических ППП, существенно расширяются возможности повседневного практического использования такого мощного аппарата добычи информации, как статистические методы анализа. Статистику, таким образом, можно считать одним из компонентов принятия управленческих решений.

Одним из важнейших направлений анализа исходных данных является статистический анализ распределений, поскольку позволяет получить обширную информацию об объекте исследования (управления). Комплексный анализ рядов распределений включает расчет характеристик центра распределения, его структуры, оценку степени вариации и дифференциации изучаемого признака, характеристику формы распределения, поиск и оценку возможности аппроксимации эмпирического распределения каким-либо теоретическим распределением на основе критериев согласия.

Большинство современных исследований в области экономики, социологии, маркетинга и других сферах проводятся с использованием выборочного наблюдения. Профессионально организованное и проведенное выборочное наблюдение позволяет получить вероятностные оценки параметров интересующего исследователя объекта или процесса в условиях значительной экономии времени, экономических и финансовых ресурсов. Благодаря уменьшению объема изучаемой совокупности, появляется возможность расширить программу наблюдения, т.е. получить более детальную информацию об объекте изучения.

В данном учебном пособии рассматривается реализация анализа распределений и результатов выборочного наблюдения с использованием программы STATISTICA.

STATISTICA - это универсальная интегрированная система, предназначенная для статистического анализа и визуализации данных, управления базами данных и разработки пользовательских приложений, содержащая широкий набор процедур анализа для применения в

научных исследованиях, технике, бизнесе, а также специальные методы добычи данных.

Помимо общих статистических и графических средств, в системе имеются специализированные модули, например, для проведения социологических или биомедицинских исследований, решения технических и, что очень важно, промышленных задач: карты контроля качества, анализ процессов и планирование эксперимента. Работа со всеми модулями происходит в рамках единого программного пакета, для которого можно выбирать один из нескольких предложенных интерфейсов пользователя.

С помощью реализованных в системе STATISTICA мощных языков программирования, снабженных специальными средствами поддержки, легко создаются законченные пользовательские решения, встраиваемые в различные другие приложения или вычислительные среды. Трудно представить, что кому-то могут понадобиться абсолютно все статистические процедуры и методы визуализации, имеющиеся в системе STATISTICA, однако опыт показывает, что возможность доступа к новым, нетрадиционным методам анализа данных (а STATISTICA в значительной мере предоставляет такие возможности) помогает находить новые способы проверки рабочих гипотез и исследования данных.

1. Инсталляция и запуск статистического пакета StatSoft STATISTICA 7.0

Инсталляция системы во многом схожа с установкой любых других Windows-приложений. Система полностью интегрируется с Windows и с некоторыми другими приложениями (например, Microsoft Word, Excel), позволяя вызывать, создавать и добавлять документы STATISTICA в эти приложения и наоборот. Данный ППП взаимодействует с основными графическими пакетами (например, Adobe Photoshop, Illustrator). На жестком диске пакет занимает около 100Мб.

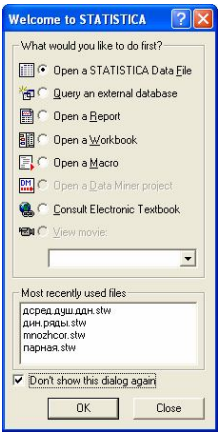


Рис. 1.1. Окно первоначального запуска STATISTICA.

При первоначальном запуске пакета появляется диалоговое окно (рис. 1.1.), с помощью которого система предлагает совершить следующие начальные действия:

1. *Open a STATISTICA Data File*[1] – открыть рабочий лист STATISTICA.
2. *Query an external database* – запросить внешнюю базу данных.
3. *Open a Report* – открыть отчет STATISTICA.
4. *Open a Workbook* – открыть рабочую книгу STATISTICA.

[1] Здесь и далее подчеркнутый курсив это имена кнопок и опций в интерфейсе программы.

5. *Open a Macro* – открыть макрос STATISTICA.

6. *Open a Data Miner project* – открыть проект «Добытчика данных».

7. *Consult Electronic Textbook* – запросит помощь электронного руководства пользователя.

8. *View movie* – просмотреть видео.

9. *Most recently used filed* – в окне представлен список наиболее используемых файлов.

10. *Don't show this dialog again* – метка в этом поле означает, что это диалоговое окно больше не будет выводиться при запуске системы.

Этим меню удобно пользоваться при повторном запуске системы и в условиях, когда ЭВМ пользуется небольшое количество человек. В нашем случае лучше поставить метку в поле *Don't show this dialog again* и нажать *OK*. На первый план выходит открывшийся при запуске рабочий лист, схожий с документом Excel, размером 10 на 10 ячеек (рис. 1.2.) Если имеющийся массив данных не соответствует предлагаемой размерности листа, его можно закрыть и перейти к созданию своего файла (см. раздел 2). Обращаем внимание на то, что при каждом следующем запуске система автоматически загружает последний использовавшийся файл, а если его обнаружить не удается, система опять автоматически создает пустой рабочий лист.

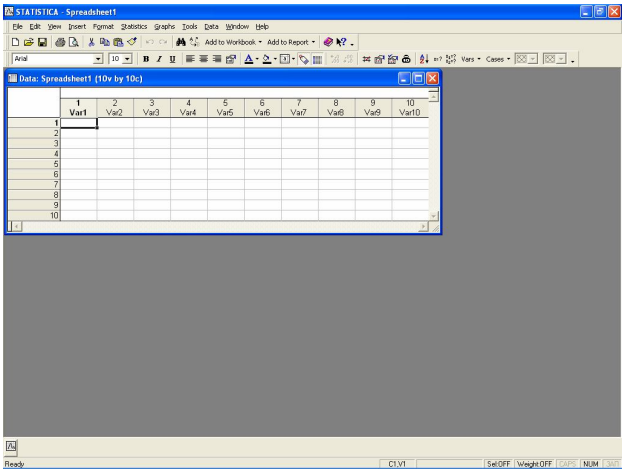


Рис. 1.2. Рабочий лист первоначального запуска STATISTICA.

2. Ввод данных в программу STATISTICA

2.1. Создание рабочей книги

Решение любой задачи статистического анализа с использованием пакетов прикладных программ начинается с процедуры ввода данных. В среде ППП STATISTICA процедура ввода реализуется следующим образом. В первую очередь необходимо создать файл, для чего в главном меню системы выбирается *File* и далее опция *New* (рис. 2.1.).

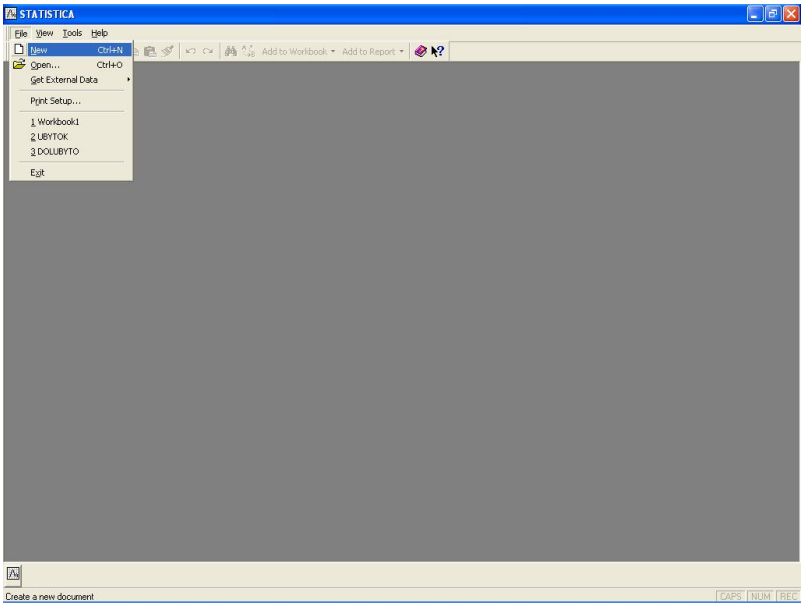


Рис. 2.1. Окно ППП STATISTICA с открытым меню File.

Программа предоставляет возможность создания файлов нескольких типов:

1. *Spreadsheet* – рабочий лист (электронная таблица).
2. *Report* – отчет.
3. *Macro (SVB) Program* – макрос.
4. *Workbook* – рабочая книга.

5. *Browser Window* – ссылка в сети.

Рабочая книга (Workbook) является наиболее общим понятием. В рабочую книгу могут быть включены любые элементы, доступные в STATISTICA, например, рабочие листы, отчеты, документы Microsoft Office.

Для ввода данных потребуется электронная таблица. Для вывода ее на экран необходимо выбрать закладку Spreadsheet (рис. 2.2.). Пользователю предоставляется возможность определить размерность таблицы, т.е. задать число переменных (*number of variables*) и число наблюдений (*number of cases*). Число переменных – это число варьирующих признаков, значения которых зарегистрированы у каждой единицы изучаемой статистической совокупности. Число наблюдений – число единиц совокупности (объем совокупности). Затем выбирается место, где будет храниться создаваемая электронная таблица (*Placement*). Целесообразно выбрать "*In a new Workbook*", что означает создание нового рабочего листа (электронной таблицы) внутри новой рабочей книги. Процедура "*As a stand-alone window*" означает создание отдельного рабочего листа.

Далее можно выбрать различные варианты вывода данных. Поле *Case name length* задает длину имени строчки, поле *MD Code* задает код вывода пропущенных ячеек (по умолчанию они не учитываются при расчете показателей), поле *Default data type* позволяет выбрать формат данных (числовой, текстовый, бинарный; по умолчанию выбран тип *Double* – это означает, что система будет воспринимать как текстовые, так и числовые переменные), поле *Variable length* предназначено для того, чтобы задавать длину данных в случае выбора текстового формата данных. Поле *Display format* задает параметры вывода вводимых или рассчитанных данных (обычный, числовой, процентный и т.д.), поле *Var name prefix* предлагает задать название создаваемых переменных, а поле *Var name start number* задает стартовый порядковый номер первой переменной. Кнопка *Default* позволяет сбросить все настройки по умолчанию.

После того, как все необходимые условия выбраны, нажимаем кнопку *OK*.

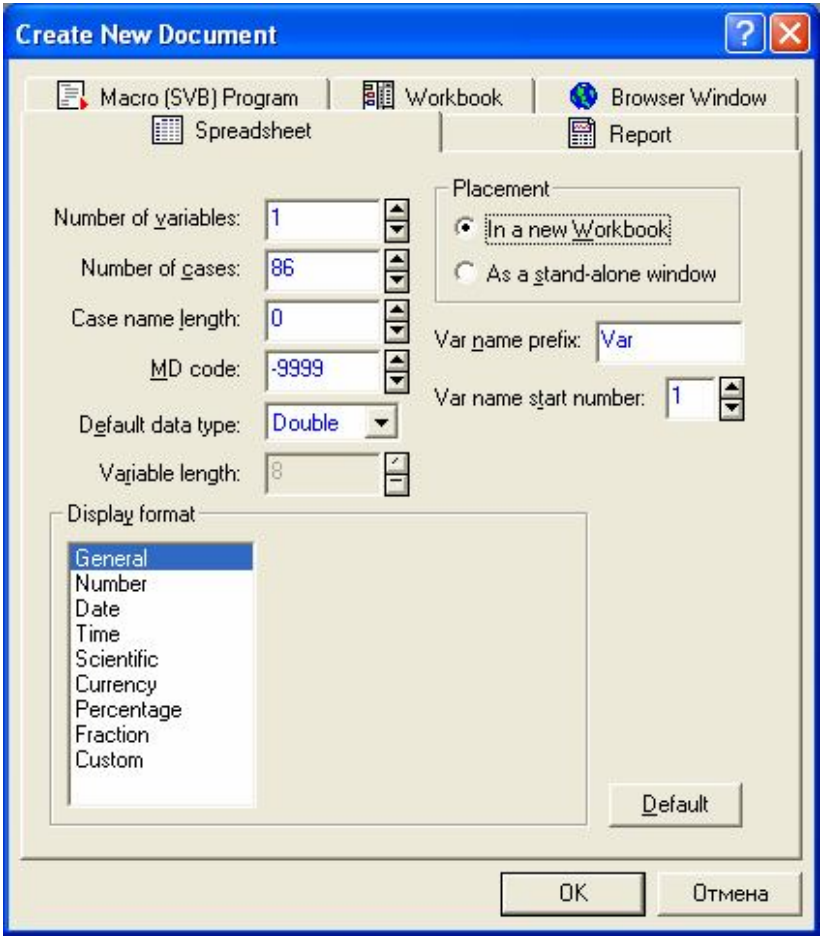


Рис 2.2. Окно выбора параметров создания нового документа STATISTICA.

Появляется рабочий лист, соответствующий заданным условиям, а слева от него дерево рабочей книги (рис. 2.3.). По умолчанию рабочая книга называется Workbook1. Внутри этой рабочей книги будут содержаться все элементы, создаваемые на основе исходных данных. С ними можно производить все элементарные преобразования (сохранение, удаление, переименование). Для этого достаточно щелк-

нуть по ним правой кнопкой мыши в дереве рабочей книги и выбрать соответствующий пункт (рис. 2.4.).

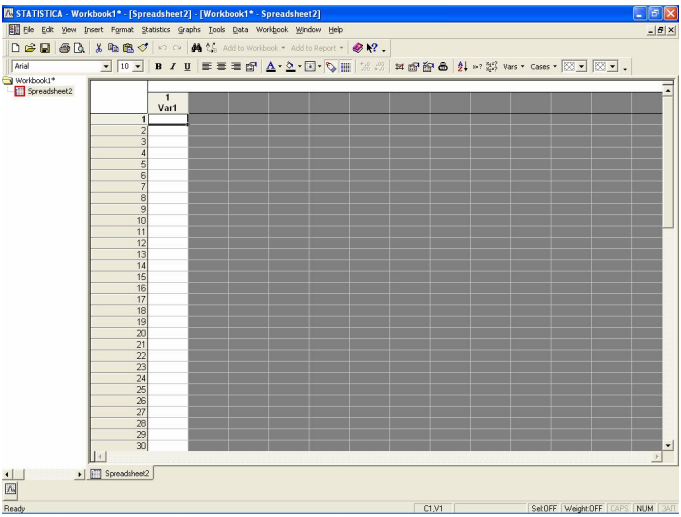


Рис. 2.3. Окно STATISTICA с созданной рабочей книгой и рабочим листом.

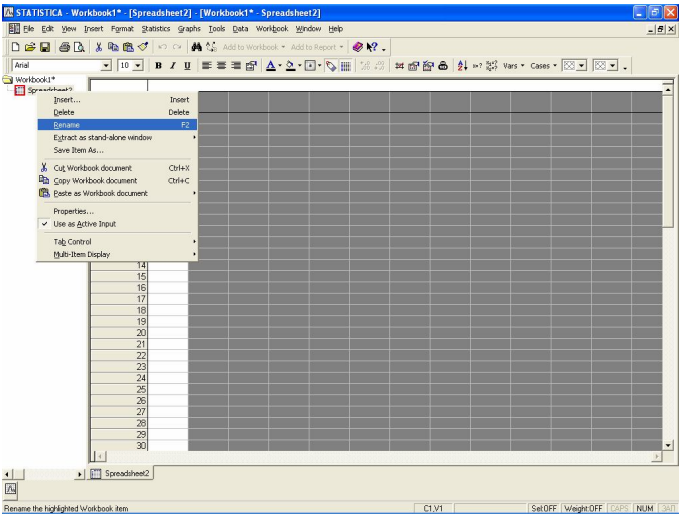


Рис. 2.4. Окно STATISTICA с выбранной функцией переименования документа.

Создаваемые рабочие книги должны иметь стандартные и легко идентифицируемые с конкретным пользователем имена. Для задания имени рабочей книги выбирается раздел меню *File* и в нём опция *Save as* (рис. 2.5.).

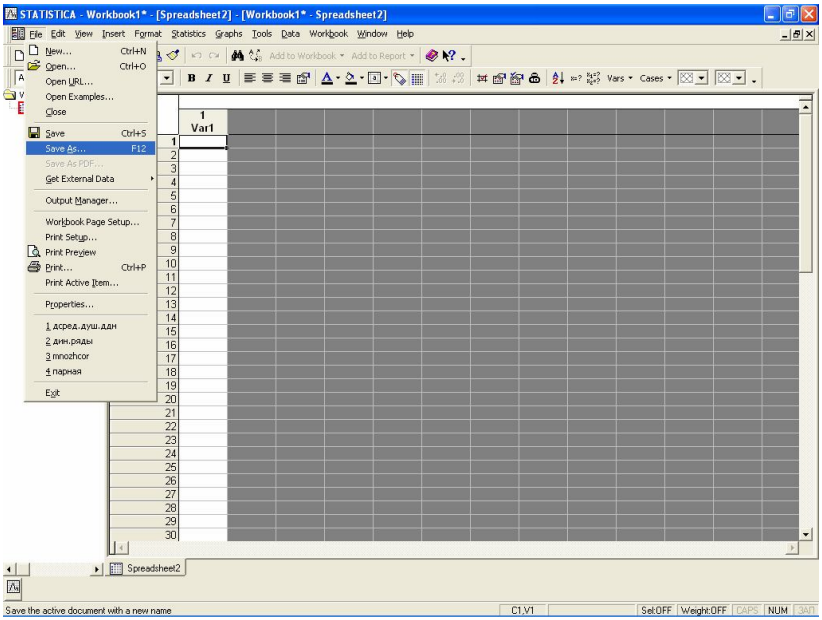


Рис. 2.5. Выбор меню сохранения документа STATISTICA.

Особо отметим, что присвоение имени рабочей книги, при выполнении лабораторных работ студентами, должно осуществляться с соблюдением определенного формата имени: имя начинается с латинских букв ST, далее записываются две цифры, означающие номер компьютера, на котором работает пользователь, в конце записывается номер учебной группы:

<ST> <номер компьютера><номер учебной группы>.

Например, имя файла ST193077/1 означает, что в рабочей книге находится информация студентов группы 3077/1, работающих на 19 компьютере (рис. 2.6.).

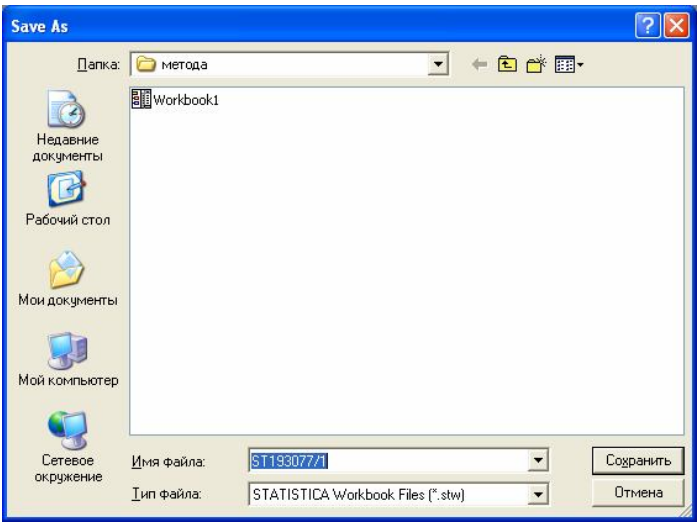


Рис. 2.6. Окно сохранения документа STATISTICA.

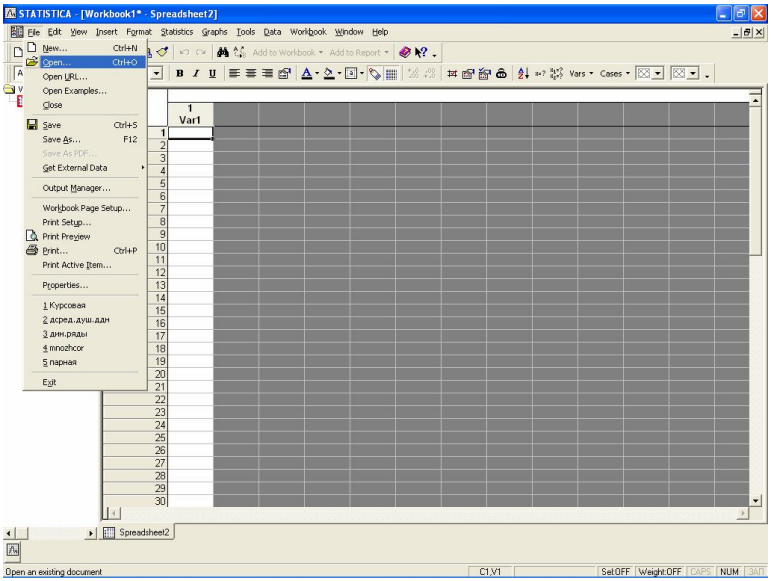


Рис. 2.7. Окно ППП STATISTICA с открытым меню File.

Обращаем внимание, что для рабочей книги характерно расширение *.stw. При запуске STATISTICA открывает файл, который использовался при последнем сеансе. Чтобы открыть нужный файл, необходимо воспользоваться меню *File/Open* и выбрать в проводнике Windows интересующую нас папку и файл (рис 2.7, 2.8.). При этом система предлагает все возможные поддерживаемые форматы (*.stw, *.sta, *.stg).

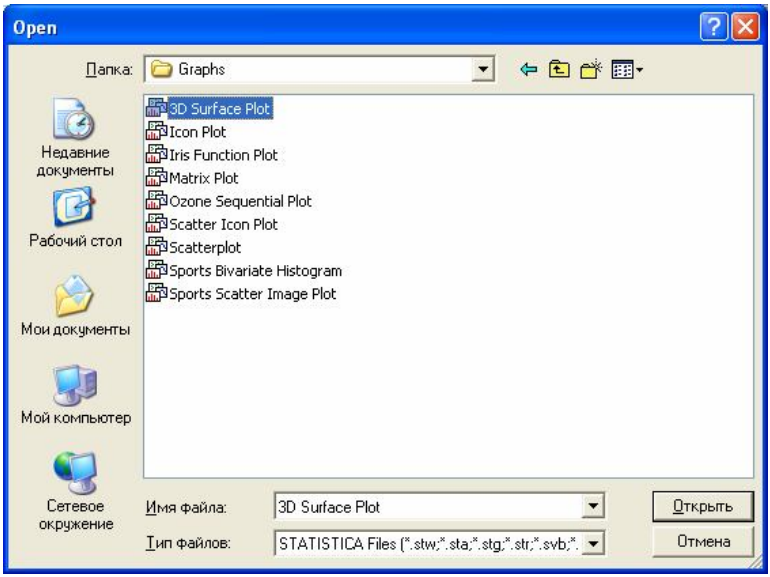


Рис. 2.8. Окно открытия документа STATISTICA.

2.2. Импорт данных в среде STATISTICA

Как уже отмечалось, STATISTICA взаимодействует с другими Windows-приложениями. Это позволяет успешно импортировать и экспортировать данные. Существует несколько способов ввода данных из других приложений в STATISTICA:

1) использование операции копирования данных через буфер обмена Windows;

2) использование механизма динамической связи DDE между данными системы STATISTICA и данными из других приложений;

3) импорт данных из наиболее популярных приложений.

Рассмотрим пример импорта данных из Microsoft Excel. Для этого выберем меню *File* и в нем команду *Open*. Автоматически система предлагает нам выбрать файлы STATISTICA (*.sta, *.stw, и т.д.). Для того чтобы импортировать данные из Excel необходимо в поле Тип файлов выбрать Excel Files (*.xls), где xls – расширение файла (рис. 2.9).

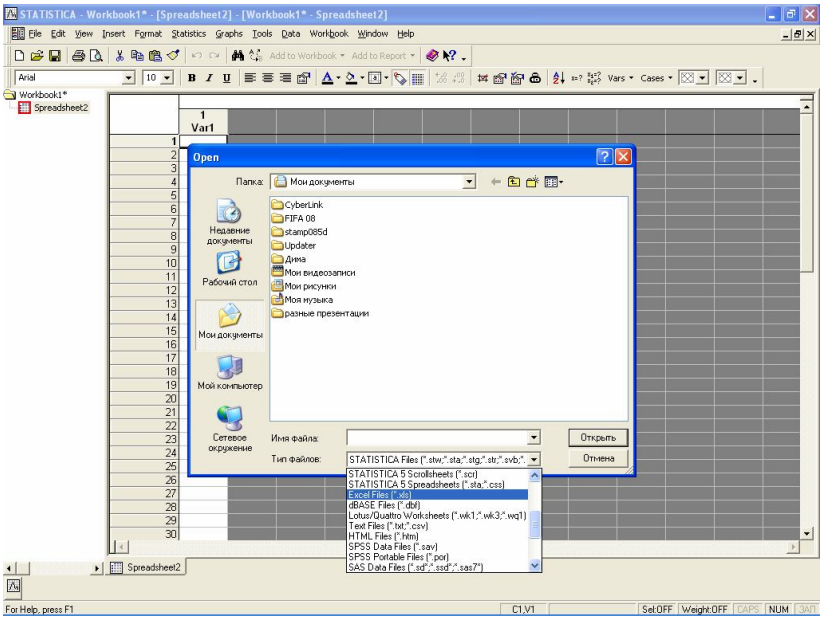


Рис. 2.9. Окно выбора формата открываемого документа.

После того как был выбран нужный файл, появляется диалоговое окно, которое предлагает либо импортировать все имеющиеся данные из файла Excel в рабочую книгу (*Import all sheets to a Workbook*), либо выбрать интересующие исследователя рабочие листы в электронную таблицу STATISTICA (*Import selected sheet to a Spreadsheet*) (рис. 2.10.).

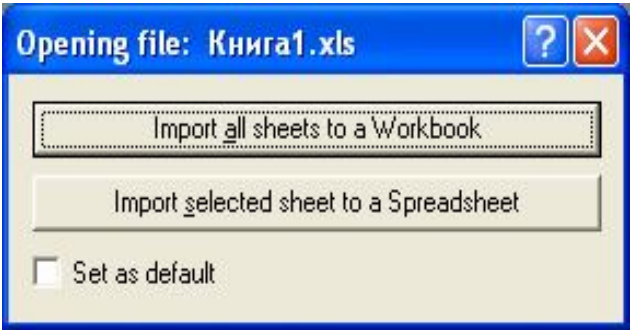


Рис. 2.10. Окно выбора параметров импорта данных.

Выберем второй вариант, после этого появляется диалоговое окно, в котором можно выбрать для импорта данных имеющиеся в файле Excel рабочие листы (рис.2.11.).

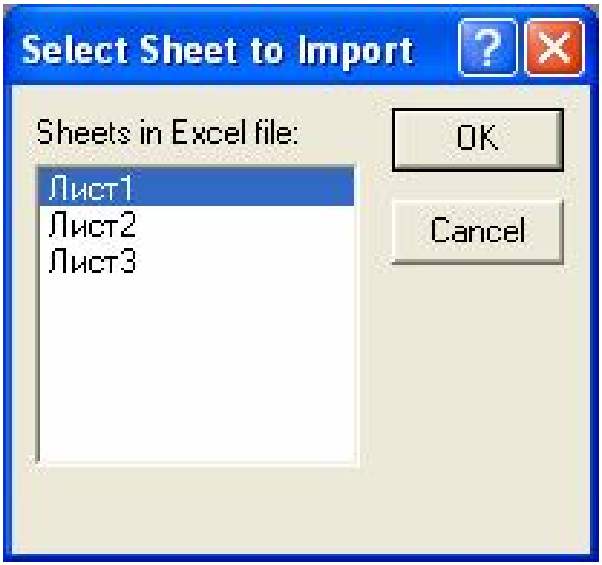


Рис. 2.11. Окно выбора рабочего листа для импорта данных.

После выбора нужного рабочего листа, пользователю предлагается уточнить параметры импорта данных, а именно:

- выбрать для импорта столбцы и строки, указав их номера (*Columns: from….. to; Rows: from….. to*);

- метки в полях *Get variable names from first row* и *Get case names from first column* означают, что переменным и наблюдениям в документе STATISTICA будут присвоены имена из первой строки и первого столбца документа Excel соответственно.

- сохранить заданный в Excel формат ячеек (*Import cell formatting)*. После выбора установок – кнопка *OK* (рис. 2.12).

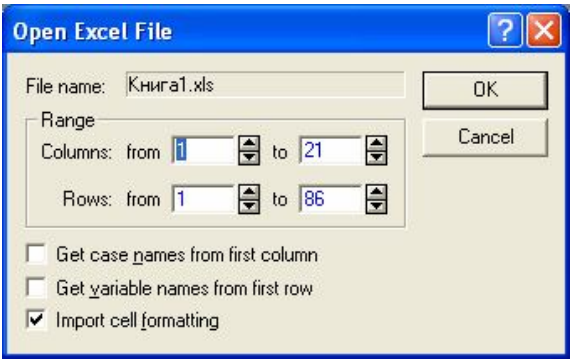


Рис. 2.12. Окно выбора диапазона импортируемых данных.

2.3. Ввод данных

Собственно ввод данных производится стандартным образом и дополнительных комментариев не требует (после записи числа на экране нажимается клавиша «Enter», десятичные дроби вводятся через запятую).

В качестве демонстрационного примера в пособии используются официальные статистические данные о распределении регионов России по показателю «Доля убыточных предприятий» (табл. 2.1) [12]. Поскольку это одномерный массив данных, число переменных (признаков) равно единице, число наблюдений – 86 (г. Москва и С-Петербург исключены как очевидные выбросы). Исходные данные помещены в рабочую книгу с именем Workbook1.stw , под именем Var1 (рис. 2.13.).

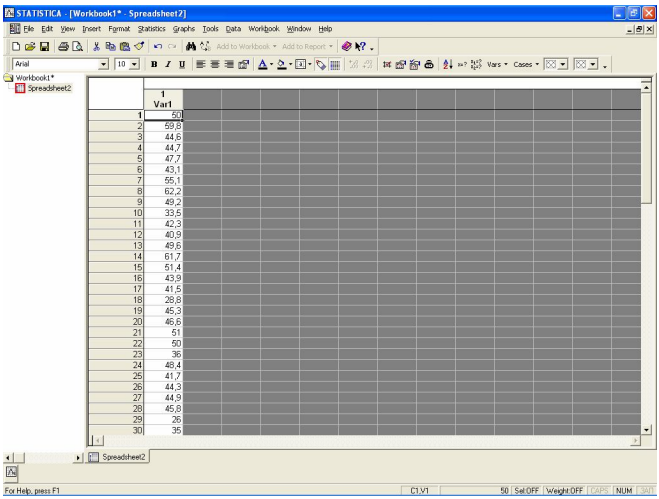


Рис. 2.13. Окно с введенными исходными данными.

Таблица 2.1

Доля убыточных предприятий по регионам России в 2003 год (проценты).

№ региона	Доля убыточных предприятий	№ региона	Доля убыточных предприятий	№ региона	Доля убыточных предприятий	№ региона	Доля убыточных предприятий	№ региона	Доля убыточных предприятий	№ региона	Доля убыточных предприятий
1	50,0	16	43,9	31	45,4	46	49,8	61	42,2	76	48,5
2	59,8	17	41,5	32	55,3	47	41,3	62	47,4	77	57,1
3	44,6	18	28,8	33	47,6	48	52,6	63	38,2	78	42,8
4	44,7	19	45,3	34	46,8	49	45,6	64	56,2	79	36
5	47,7	20	46,6	35	41,6	50	53,5	65	48,2	80	47,4
6	43,1	21	51,0	36	39,4	51	50,9	66	47,3	81	49
7	55,1	22	50,0	37	31,7	52	45,1	67	56	82	74,1
8	62,2	23	36,0	38	37	53	51,1	68	45,8	83	49
9	49,2	24	48,4	39	37,3	54	43,9	69	63	84	49,5
10	33,5	25	41,7	40	46,5	55	43,3	70	66,7	85	56,5
11	42,3	26	44,3	41	30,2	56	57,8	71	45,1	86	51,6
12	40,9	27	44,9	42	38,3	57	65,3	72	54,8		
13	49,6	28	45,8	43	52,5	58	41,9	73	49,9		
14	61,7	29	26,0	44	43,4	59	43,2	74	56,9		
15	51,4	30	35,0	45	41,8	60	36,6	75	50,1		

Для того чтобы представить исходные данные в формате Таблицы 2.1, необходимо скопировать столбец с исходными данными из документа STATISTICA (выделить столбец, щелкнуть по нему правой кнопкой мыши, выбрать функцию *Copy*) и вставить его в Microsoft Excel (рис. 2.14, 2.15.) (в данном приложении таблица редактируется таким образом, чтобы она помещалась на печатном листе и была удобной для чтения).

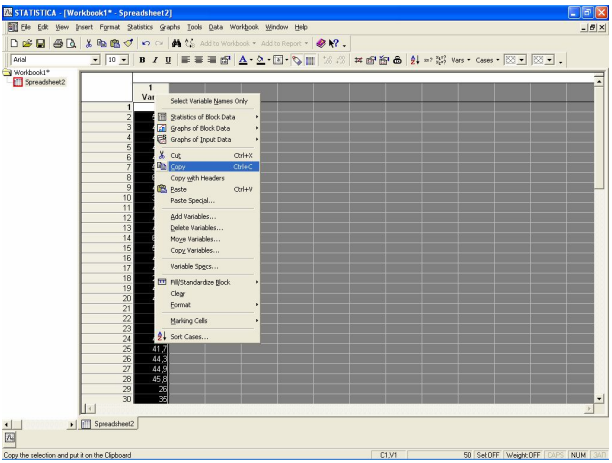


Рис. 2.14. Функция копирования исходных данных.

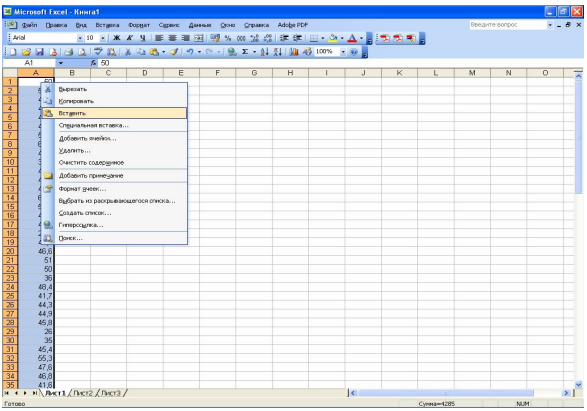


Рис. 2.15. Функция вставки данных в Microsoft Excel.

2.4. Основные операции со строками и столбцами

В этом подразделе будет представлена информация, полезная для использования возможностей программы при решении любых задач. Так же как и элементы в дереве рабочей книги, строки и столбцы на рабочем листе системы можно добавлять, перемещать, удалять, копировать.

Стандартная процедура создания новой переменной состоит в следующем. Необходимо дважды щелкнуть левой кнопкой мыши по серому полю справа от заголовка существующей переменной. Появиться диалоговое окно (рис. 2.16.), предлагающее добавить переменные (*Add Variables*) или наблюдения (*Add Cases*).

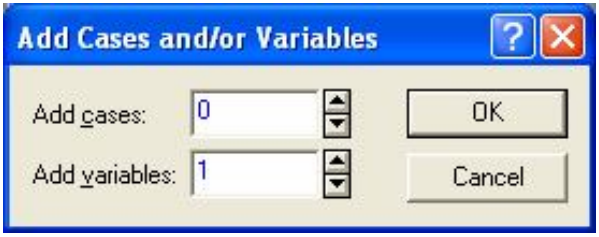


Рис. 2.16. Окно выбора параметров создания переменной.

Добавим, например, одну переменную, поэтому в соответствующем поле ставится цифра 1. Щелкаем *OK*. На экране появляется панель для выбора параметров новой переменной (рис 2.17.).

В появившемся окне можно задать имя новой переменной, например Var2. Кнопки «>>» и «<<», находящиеся под кнопкой *Cancel* в правом верхнем углу окна, позволяют перемещаться по переменным и настраивать их функции. В дальнейшем, при необходимости вызова данного окна, нужно дважды щелкнуть левой кнопкой мыши на заголовке любой из существующих переменных.

В поле *Display Format* можно также задать формат переменной, по умолчанию он обычный (*General*), но нас интересует формат числовой (*Number*). Использование этого формата позволяет устанавливать необходимую точность вычислений, т.е. количество десятичных знаков после запятой (поле *Decimal places*) (см. рис. 2.17.).

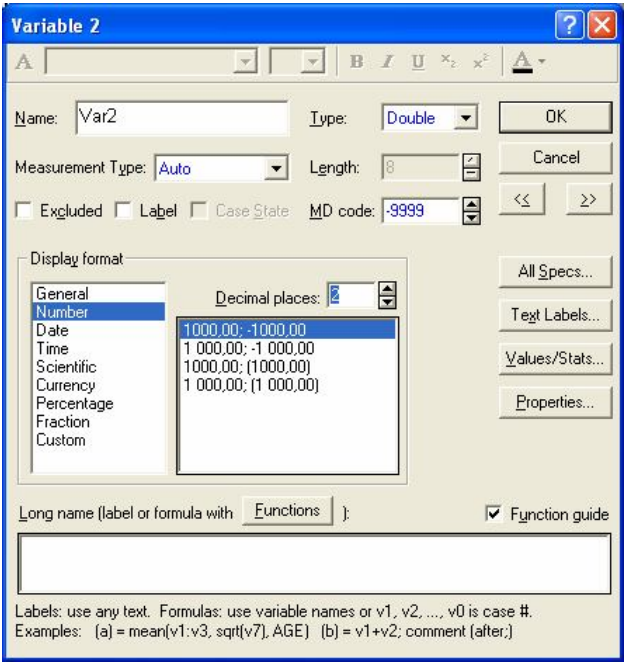


Рис. 2.17. Окно выбора параметров переменной.

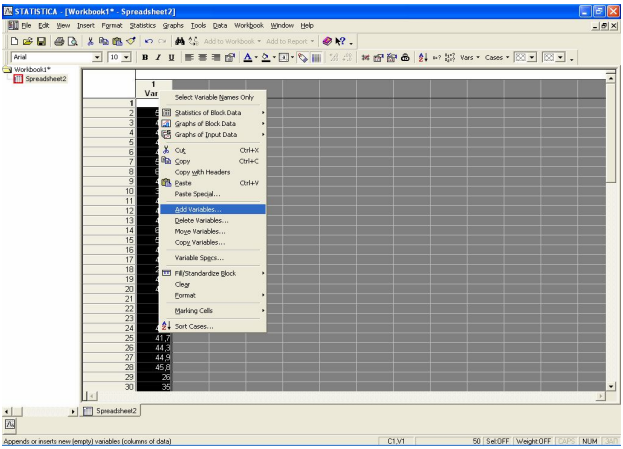


Рис. 2.18. Функция добавления переменных.

Когда на рабочем листе находится несколько переменных, то можно добавлять новые с указанием места их добавления. Для этого необходимо щелкнуть правой кнопкой мыши по заголовку любого столбца и выбрать функцию *Add Variables* (рис. 2.18.) (в этом же меню можно выбрать функции перемещения, удаления и копирования переменных, выполняемые аналогичным образом).

В появившемся окне кроме уже известных нам процедур существует поле *After*, в котором указывается имя переменной, после которой мы хотим поместить вновь создаваемую (рис. 2.19.).

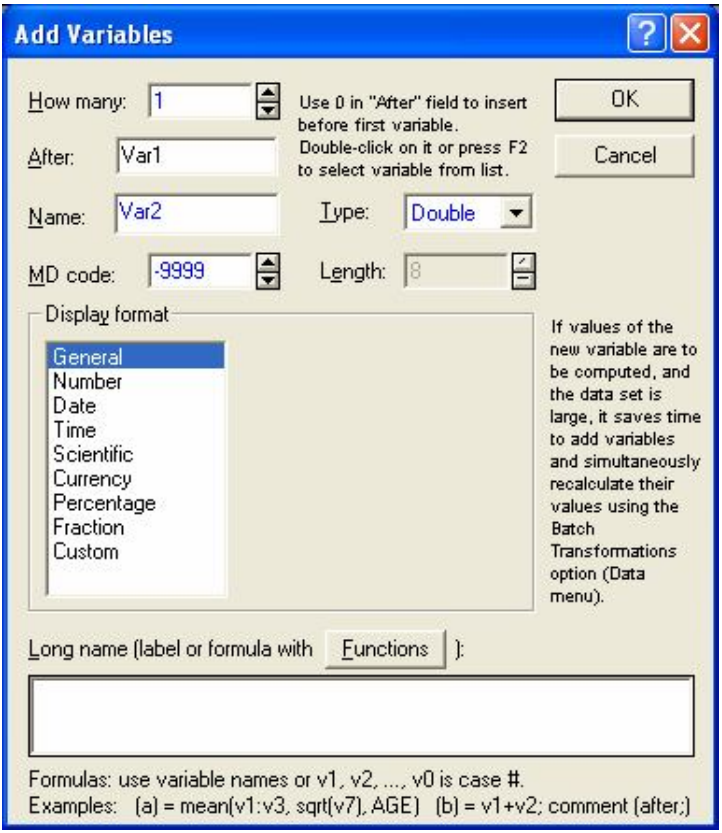


Рис. 2.19. Окно функции добавления новой переменной.

При выборе других функций в этом меню также требуется указание имен переменных, например от какой и до какой переменной следует удалить столбцы.

Данные из столбцов текущего рабочего листа, других рабочих листов, документов Excel можно копировать одинаковым образом. Для этого выделяем интересующую нас область, затем, щелкнув правой кнопкой мыши, в появившемся подменю выбираем функцию *Copy* (рис. 2.20.).

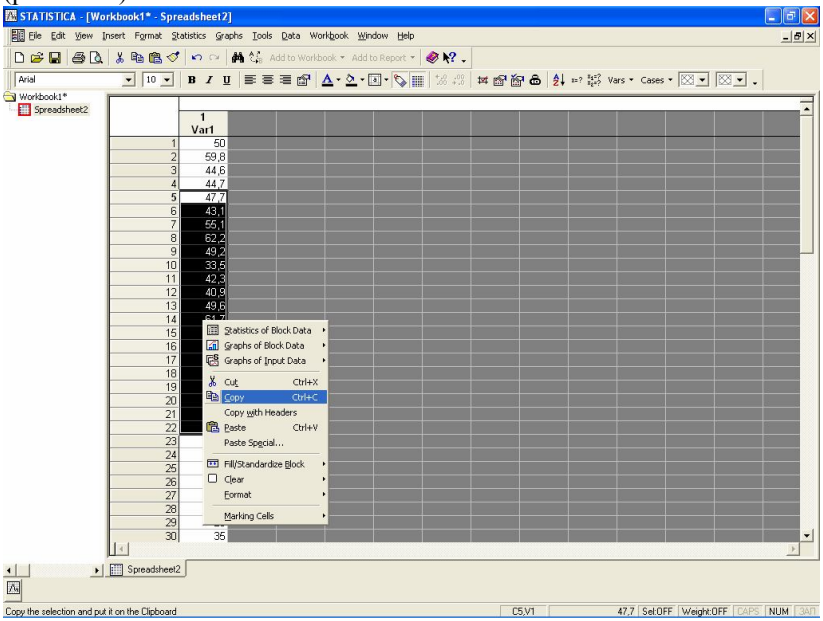


Рис. 2.20. Функция копирования данных в STATISTICA.

Далее щелкаем правой кнопкой мыши на заголовке новой переменной и в появившемся меню выбираем функцию *Paste* (вставка) (рис. 2.21.).

Все эти функции справедливы и для строк. Например, мы хотим удалить строки с 10 по 15. Для этого нужно щелкнуть правой кнопкой мыши по любому номеру строки и выбрать функцию *Case Management/Delete cases* (рис. 2.22.). Далее в появившемся окне в поле *From case* ставим номер строки, начиная с которой нужно удалить данные; в

поле *To case* ставим номер строки, по которую нужно удалить данные включительно (рис. 2.23.).

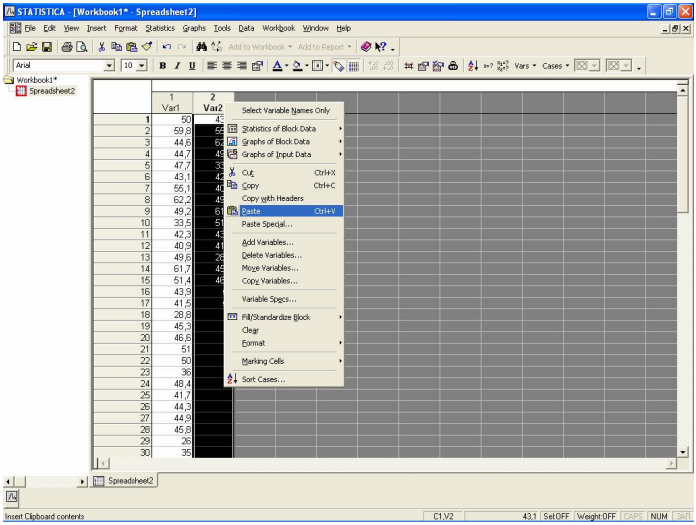


Рис. 2.21. Функция вставки данных в STATISTICA.

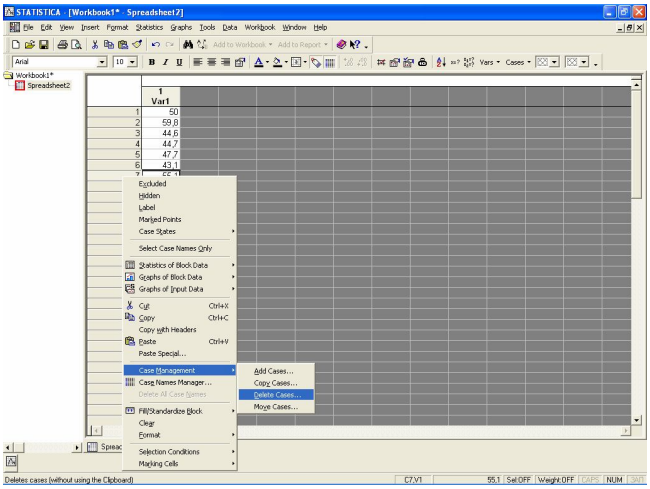


Рис. 2.22. Выбор функции удаления строк.

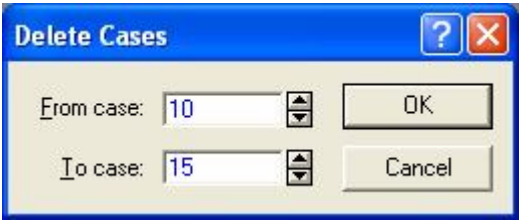


Рис. 2.23. Функция удаления строк в STATISTICA.

Все функции с переменными и наблюдениями можно запустить с помощью кнопок: *Vars* и *Cases* соответственно в панели инструментов (рис. 2.24.).

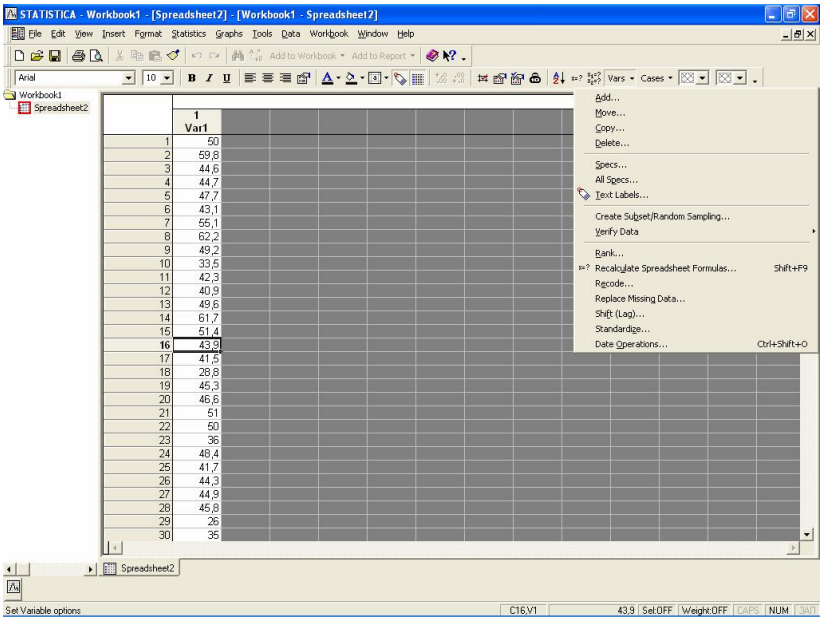


Рис. 2.24. Кнопки управления операциями со столбцами и строками.

3. Анализ эмпирического распределения

Анализ распределений направлен на выявление закономерности изменения частот в зависимости от значений варьирующего признака и анализ различных характеристик изучаемого распределения. Прежде, чем приступить к вычислению специальных статистических показателей, необходимо из исходной совокупности исключить единицы, не подчиняющиеся общей закономерности распределения, так называемые выбросы. Выбросы – это значения признака, резко отличающиеся как в большую, так и в меньшую сторону, от значений признака у основной части единиц совокупности. [4]

ВНИМАНИЕ!!! Вопросы, связанные с выявлением и удалением выбросов обсуждаются с преподавателем на лабораторных занятиях.

Для удобства локализации и устранения выбросов необходимо, прежде всего, ранжировать исходные данные.

Для проведения сортировки выбираем в меню *Data* опцию *Sort* (рис. 3.1.). Либо нажимаем соответствующую кнопку в панели инструментов, либо щелкаем правой кнопкой по заголовку столбца с нашей переменной и выбираем функцию *Sort cases*.

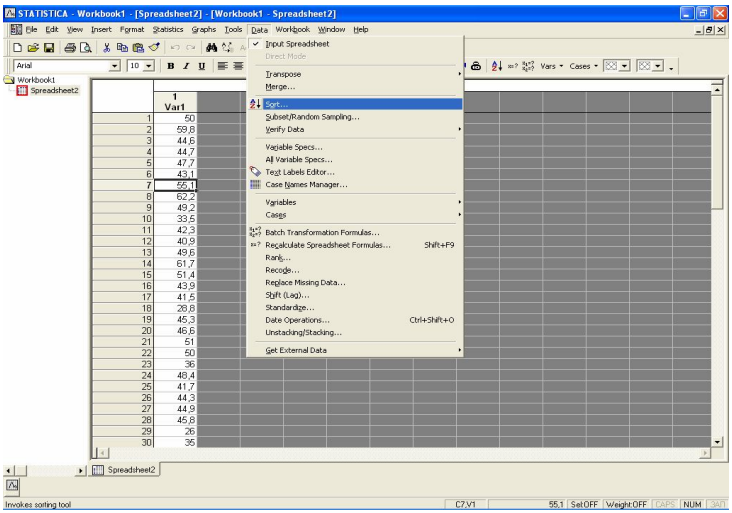


Рис. 3.1. Окно выбора меню сортировки переменных.

В появившемся окне с помощью кнопки *Add Var(s)* добавляем нашу переменную Var1 в список для сортировки (рис. 3.2.).

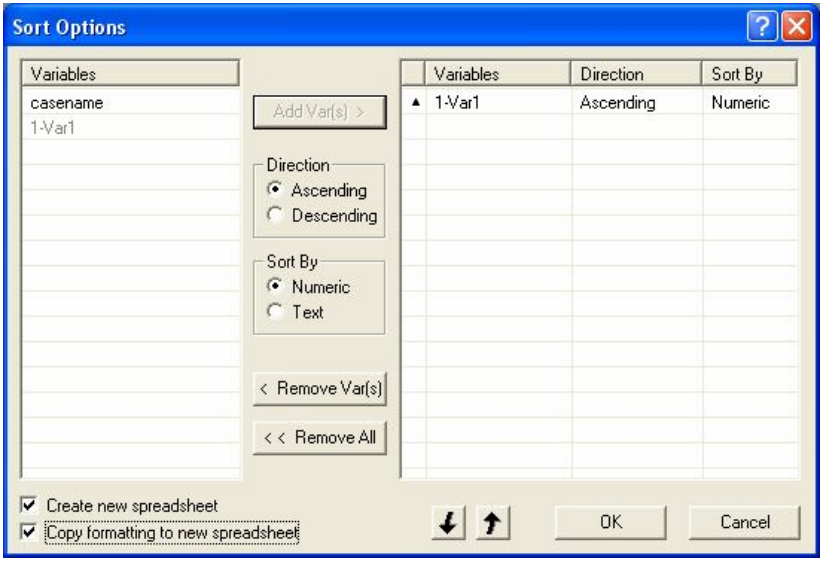


Рис. 3.2. Окно выбора параметров сортировки.

Далее в поле *Direction* (направление) выбираем тип сортировки:
- по возрастанию значений признака (*Ascending*);
- по убыванию (*Descending*).

Затем в поле *Sort by* выбираем основание для сортировки:
- сортировка по численным значениям (*Numeric*);
- сортировка по текстовым значениям (*Text*).

Далее ставим метку в поле *Create new spreadsheet* (создать новый рабочий лист) – это означает, что отсортированная переменная Var1 будет помещена в новый рабочий лист, а не отсортированная останется в старом. Это делается по той причине, что для второй лабораторной работы нам нужны исходные данные, не подвергавшиеся никакой обработке.

Также ставим метку в поле *Copy formatting to new spreadsheet* – это означает копирование формата переменных, а не только самих значений в ячейках.

Кнопки *Remove Var(s)* и *Remove All* нужны для того, что убрать либо одну, либо все переменные из списка сортировки. Кнопки со стрелками «вверх» и «вниз» используются для перемещения переменных по списку сортировки.

Далее нажимаем кнопку *OK*.

Система создает новый рабочий лист с отсортированной переменной Var1. Этот рабочий лист необходимо добавить в нашу рабочую книгу Workbook1. Для этого нажимаем на кнопку *Add to Workbook* в панели инструментов и выбираем из списка нужную рабочую книгу (рис. 3.3.). Если запущена только одна рабочая книга, то в этом списке она будет первой (над чертой).

Теперь в рабочей книге содержится два рабочих листа, при этом второй (который мы добавили только что) становится активным (подсвечивается красным в дереве рабочей книги (рис. 3.4.). Следует обратить внимание на общее правило работы в ППП STATISTICA: все запускаемые процедуры выполняются с данными активного в текущий момент рабочего листа (выделенного красным цветом).

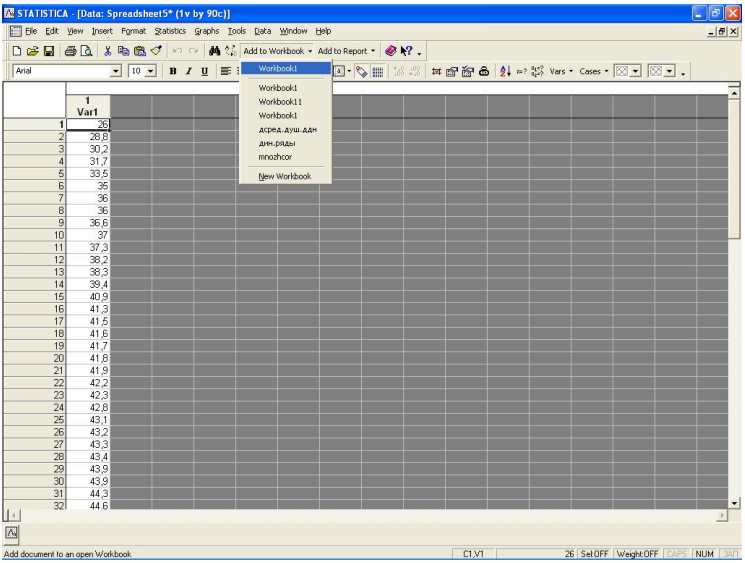


Рис. 3.3. Функция добавления рабочего листа в рабочую книгу.

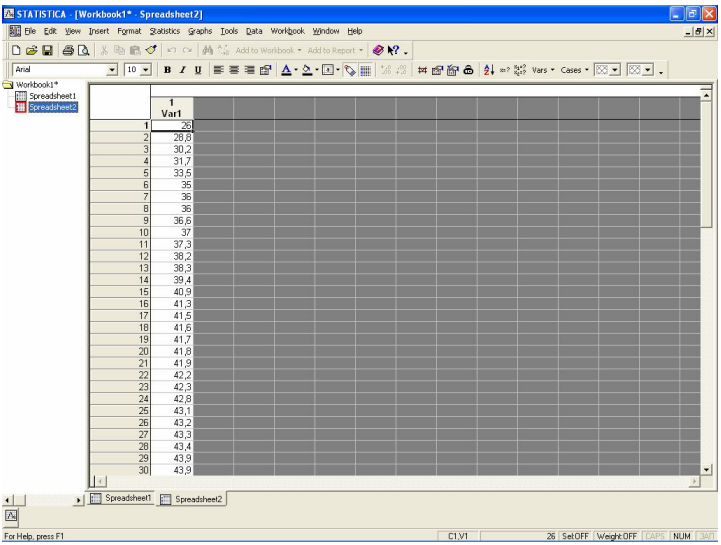


Рис. 3.4. Вид рабочего окна с двумя рабочими листами.

Завершающий этап корректировки новой переменной состоит в элементарном удалении значений, идентифицированных как выбросы. Каждое из них выделяется и стирается простым нажатием кнопки *Delete* на клавиатуре. Также удаление значений (выбросов) из ячеек, можно провести как удаление наблюдений, то есть строк (см. раздел 2.4.).

Заметим, что в отличие от некоторых других статистических пакетов, STATISTICA учитывает только те ячейки, которые заполнены, то есть наличие пустых ячеек допустимо. В нашем примере как выброс удаляется последнее значение ранжированной совокупности, таким образом, окончательный объем совокупности составляет 85 единиц (табл. 3.1).

Если нет необходимости сохранения исходных данных в том виде, какими они были до удаления выбросов, процедура сортировки может быть изменена, т.е. проведена без резервного сохранения исходной переменной. Для этого не надо ставить метку в поле *Create new spreadsheet* (см. рис. 3.2.).

	1 Var1
1	26
2	28,8
3	30,2
4	31,7
5	33,5
6	35
7	36
8	36
9	36,6
10	37
11	37,3
12	38,2
13	38,3
14	39,4
15	40,9
16	41,3
17	41,5
18	41,6
19	41,7
20	41,8

66	51,4
67	51,6
68	52,5
69	52,6
70	53,5
71	54,8
72	55,1
73	55,3
74	56
75	56,2
76	56,5
77	56,9
78	57,1
79	57,8
80	59,8
81	61,7
82	62,2
83	63
84	65,3
85	66,7
86	74,1

Рис. 3.5. Вид начала и конца рабочего листа с ранжированной переменной.

ВНИМАНИЕ !!! Для того чтобы вставить фрагмент таблицы или любой другой элемент из STATISTICA в какой-либо документ как рисунок, необходимо одновременно нажать на клавиатуре клавиши Alt и F3. На экране вместо курсора появится прицел. Нажав и удерживая левую кнопку мыши, выделяем необходимый диапазон. После того как кнопка мыши отпущена, фрагмент автоматически копируется в буфер обмена, и остается только вставить его в нужный документ. Именно таким образом рекомендуется составлять отчеты по лабораторным работам, так как некоторые таблицы невозможно скопировать с помощью стандартной процедуры.

Таблица 3.1

Исходные данные, ранжированные по возрастанию значений признака

№ региона	Доля убыточных предприятий	№ региона	Доля убыточных предприятий	№ региона	Доля убыточных предприятий	№ региона	Доля убыточных предприятий	№ региона	Доля убыточных предприятий	№ региона	Доля убыточных предприятий
1	26	16	41,3	31	44,3	46	47,4	61	50	76	56,5
2	28,8	17	41,5	32	44,6	47	47,4	62	50,1	77	56,9
3	30,2	18	41,6	33	44,7	48	47,6	63	50,9	78	57,1
4	31,7	19	41,7	34	44,9	49	47,7	64	51	79	57,8
5	33,5	20	41,8	35	45,1	50	48,2	65	51,1	80	59,8
6	35	21	41,9	36	45,1	51	48,4	66	51,4	81	61,7
7	36	22	42,2	37	45,3	52	48,5	67	51,6	82	62,2
8	36	23	42,3	38	45,4	53	49	68	52,5	83	63
9	36,6	24	42,8	39	45,6	54	49	69	52,6	84	65,3
10	37	25	43,1	40	45,8	55	49,2	70	53,5	85	66,7
11	37,3	26	43,2	41	45,8	56	49,5	71	54,8	86	74,1
12	38,2	27	43,3	42	46,5	57	49,6	72	55,1		
13	38,3	28	43,4	43	46,6	58	49,8	73	55,3		
14	39,4	29	43,9	44	46,8	59	49,9	74	56		
15	40,9	30	43,9	45	47,3	60	50	75	56,2		

3.1. Графическое и табличное представление вариационного ряда распределения

Вариационным называется ряд распределения, построенный по количественному признаку. Он может быть представлен в виде таблицы и графически. Табличное представление позволяет не только выявить ту или иную закономерность распределения, но и подробно охарактеризовать структуру изучаемой совокупности (см., например, [6,7,8]).

Таблицы вариационных рядов строятся по принципам группировки. Известные проблемы возникают при определении числа групп, поскольку формула Стерджеса (Sturges H.A., The choice of class interval. J. Amer. Assoc. 21 (1926))[2] (3.1), рекомендуемая для этих целей, дает приемлемые результаты только в условиях больших статистических совокупностей. Процесс определения числа выделяемых групп, в значительной степени, носит творческий характер и требует от исследователя применения не только теоретических знаний, но и практического опыта и интуиции.

Формула Стерджеса:

$$k = 1 + 3{,}322 \times \lg N = 1 + 1{,}44 \times \ln N\,, \qquad (3.1.)$$

где k - число групп; N – объем совокупности.

Использование ППП значительно упрощает задачу табличного представления вариационного ряда, поскольку позволяет с малыми временными затратами просмотреть несколько таблиц с разным числом групп и размером группировочного интервала. Конечный вариант таблицы должен отвечать следующим требованиям: в таблице не должно быть малонаполненных и нулевых групп; нужно стремиться к получению мономодального распределения (т.е. по обе стороны от максимальной частоты должно наблюдаться закономерное убывание частот). Если не удается избавиться от многовершинности в распределении, это, как правило, означает, что изучаемая статистическая совокупность неоднородна и требует более детального изучения. В этих условиях следует либо работать с выбросами, либо, если единицы совокупности не подчиняются единой закономерности распределения, разбить совокупность на объективно существующие группы, и анализировать их раздельно.

Для построения таблиц вариационного ряда выбираем следующий путь: меню *Statistics/Basic Statistics Tables* (рис. 3.6.)

---

[2] Обращаем внимание читателя на то, что во всей русскоязычной профессиональной и учебной литературе (исключения нам не известны) принято написание фамилии автора – Стерджесс. Возможно это шлейф первого не вполне профессионального перевода с английского выражения “ Sturges’s formula” и последующих не критических взаимозаимствований.

В появившемся диалоговом окне выбираем пункт *Frequency tables* (таблицы частот) и нажимаем кнопку *OK* (рис.3.7.)

Появляется диалоговое окно процедуры построения таблиц. Любой анализ в системе начинается с выбора анализируемой переменной, для этого используется кнопка *Variables* (рис. 3.8.)

В нашем примере выбираем переменную, содержащую ранжированный ряд с учетом выбросов, и нажимаем кнопку *OK* (рис. 3.9.)

Различные кнопки на закладке *Quick* применительно к данной задаче нас не интересуют, поэтому переходим к закладке *Advanced* (дополнительно) (рис. 3.10.).

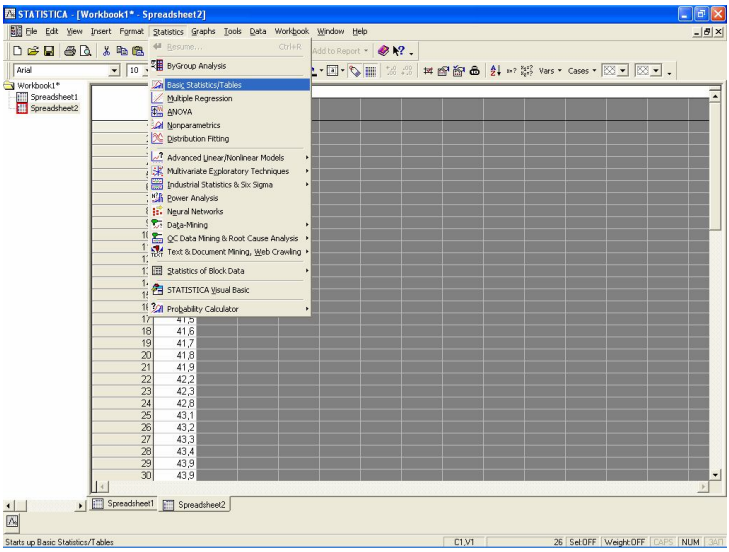


Рис. 3.6. Выбор меню Basic Statistics/Tables.

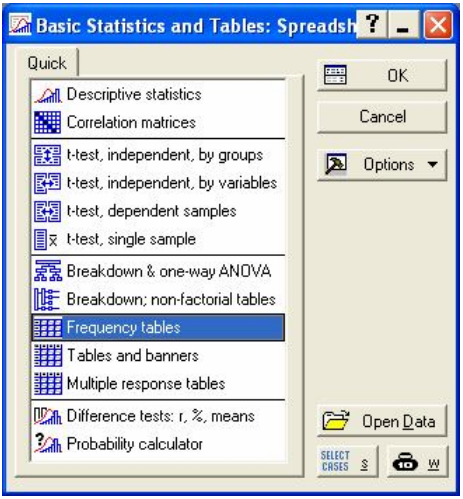


Рис. 3.7. Окно выбора процедуры Frequency Tables.

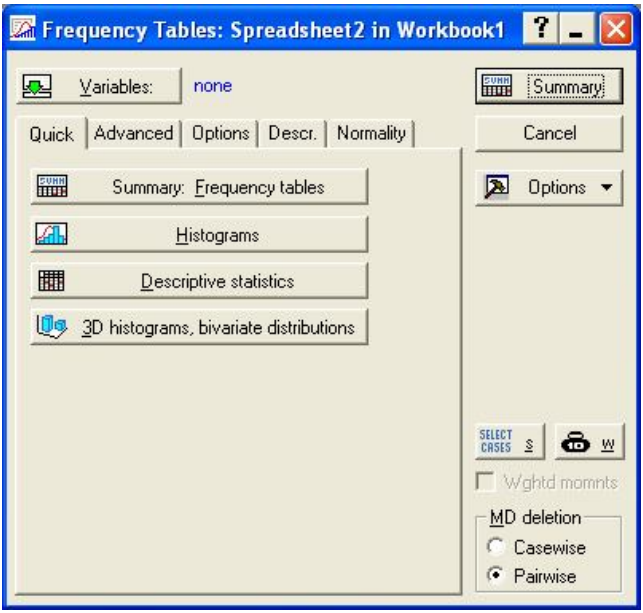


Рис. 3.8. Окно процедуры Frequency tables.

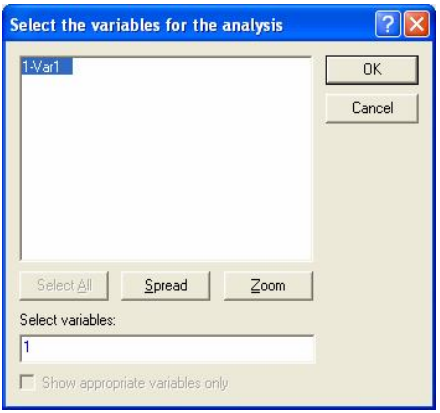


Рис. 3.9. Окно выбора переменных.

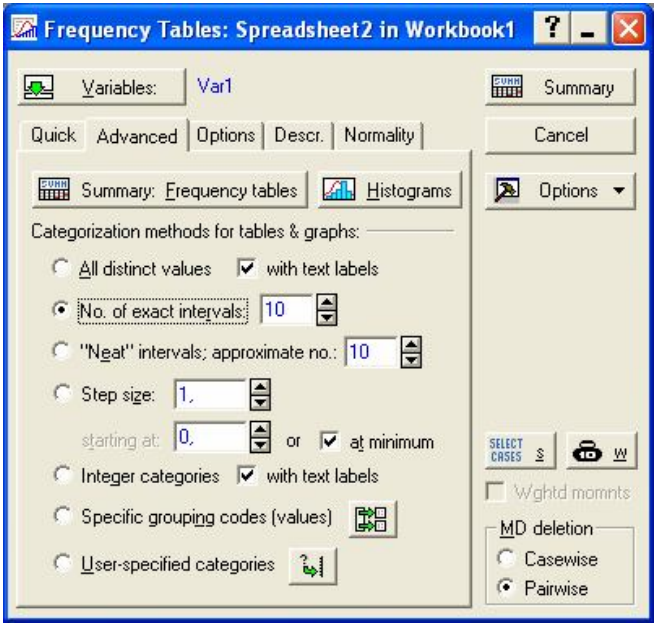


Рис. 3.10. Закладка Advanced процедуры Frequency tables.

В правом верхнем углу находятся три кнопки, которыми можно пользоваться независимо от того, на какой закладке меню находится пользователь:

1. *Summary* – позволяет построить таблицы по заданным на данный момент опциям.
2. *Cancel* – выход из меню.
3. *Options* – позволяет выполнить несколько действий по выбору пользователя:
   3.1 *Close Analysis* - выйти из процедуры текущего анализа;
   3.2 *Creat macro* - создать макрос;
   3.3 *Output* - задать опции для окон анализа и других окон, задать опции для вывода и сохранения выбранных процедур.

При выполнении лабораторной работы, опцией Output необходимо воспользоваться перед началом любой процедуры, так как ранее оговаривалось, что все элементы работы будут сохраняться в одной рабочей книге (рис. 3.11.).

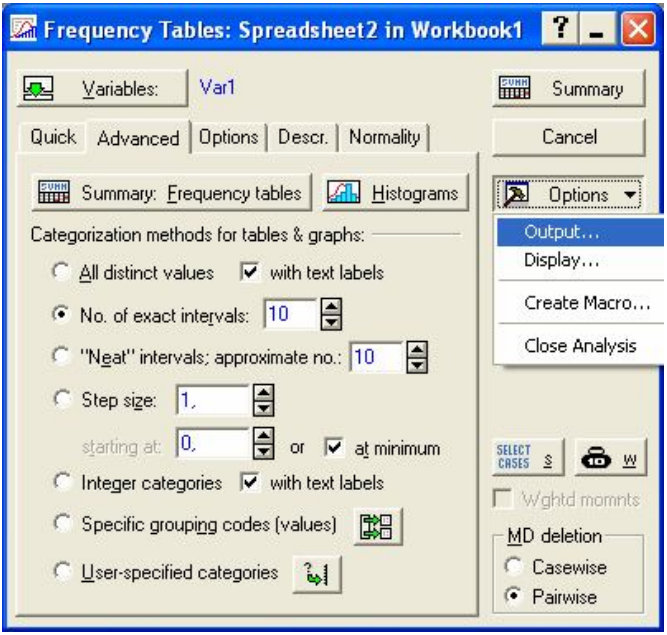


Рис. 3.11. Выбор меню Output.

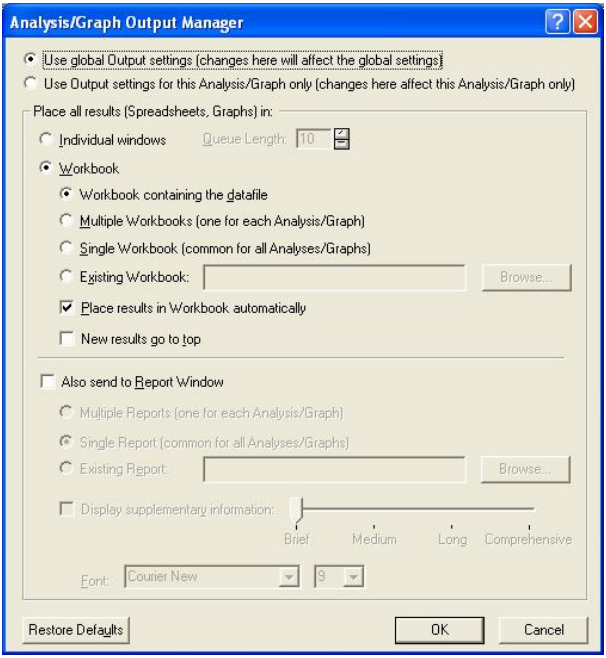


Рис. 3.12. Окно задания параметров вывода результатов анализа.

Ставим метку на первой строке *Use global Output settings* (использовать глобальные установки вывода), что означает использование заданных условий для всех запускаемых процедур в STATISTICA, метка на второй строке означает использование заданных условий только для текущей процедуры (рис. 3.12.).

Далее меню *Place all results* предлагает пользователю несколько вариантов вывода результатов процедуры.

Individual window - вывод результатов в отдельном независимом окне;

Workbook - вывод результатов в рабочую книгу. При этом возможны следующие варианты:

1. *Workbook containing the datafile* – результаты помещаются в рабочую книгу, в которой содержатся исходные данные.

2. *Multiple Workbooks* – результаты каждого анализа помещаются в отдельную рабочую книгу.

3. *Single Workbook* – создание новой рабочей книги, в которую будут помещаться результаты всех анализов.

4. *Existing Workbook* – пользователь самостоятельно выбирает рабочую книгу для сохранения результатов анализа.

При выполнении лабораторной работы удобно воспользоваться первым вариантом (*Workbook containing the datafile)*.

После выбора варианта вывода результатов активизируется (ставится метка) в поле *Place results in Workbook automatically*, что означает автоматическое размещение результатов анализа в определенной папке рабочей книги. Так же данное меню позволяет сразу отправлять результаты анализа в отчет, однако целесообразнее создавать отчет в конце каждой лабораторной работы, во избежание нагромождения информации в рабочей книге. Поэтому поле *Also send to Report Window* оставляем без метки. Так же как и поле *New results go to top* (результаты анализа отображаются на экране выше других окон), так как это задано автоматически.

Заметим, что описанную процедуру достаточно выполнить всего один раз, так как далее она будет выполняться автоматически для любого вновь запускаемого анализа.

Под закладками находятся две кнопки. Первая (*Summary: Frequency Tables*) используется для построения таблиц, вторая (*Histograms*) – для построения гистограммы распределения с наложенным на нее нормальным распределением.

Однако сначала, необходимо задать параметры построения таблиц (см. рис. 3.10.). Для этого рассмотрим поля под вышеназванными кнопками.

1. Метка в поле *All distinct values* означает, что в таблицу будут включены абсолютно все значения, а если справа от этого поля стоит метка в поле *With text labels*, то включены будут не только численные, но и текстовые наблюдения.

2. Метка в поле *No. of exact intervals* означает, что задается точное число интервалов, на которые будет разбита совокупность, и в поле справа от метки ставится точное число интервалов.

3. Метка в поле *"Neat" intervals; approximate no.* означает, что число интервалов, задаваемое пользователем в поле справа, означает примерное число интервалов, и система определит точное число интервалов самостоятельно.

4. С помощью поля *Step size* возможно задать лишь величину интервала (первое окно) и нижнюю границу первого интервала (либо задается

исследователем – метка в поле *starting at*, либо начиная с минимального значения – *at minimum*).

5. Метка в поле *Integer categories* означает, что в таблицу будут включены только целочисленные значения.

6. Поля *Specific grouping codes (values)* и *User-specified categories* позволяют задать определенные критерии отбора элементов для таблицы.

Выбираем вариант с точным заданием числа интервалов (2), поскольку предполагается построение нескольких таблиц с разным числом интервалов (см. рис. 3.10.).

Далее переходим к закладке *Options* (рис.3.13.).

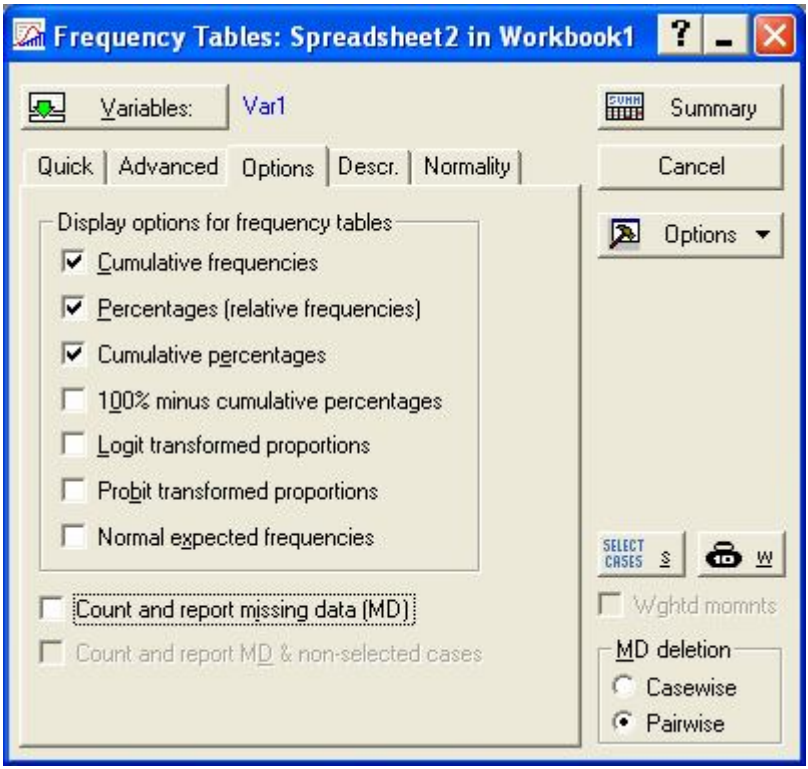


Рис. 3.13. Закладка Options процедуры Frequency Tables.

На закладке необходимо отметить элементы, которые должны быть включены в таблицу. Для более полного представления исследуемой переменной следует установить метки в полях *Cumulative frequencies* (накопленные частоты), *Percentages (relative frequencies)* (относительные частоты) и *Cumulative percentages* (накопленные относительные частоты). Таким образом, для анализа распределения становятся доступны абсолютные, относительные и кумулятивные частоты.

Отсутствие метки в поле *Count and report missing data (MD)* означает, что система не будет учитывать пустые ячейки, если они содержатся в переменной.

Далее нажимаем кнопку *Summary*, и появляется расчетная таблица частот (рис. 3.14.). Для построения таблицы с другим числом интервалов необходимо вернуться к соответствующему диалоговому окну (рис. 3.10.) и изменить значение числа интервалов.

Запущенные направления анализа отображаются в виде кнопок в левом нижнем углу экрана.

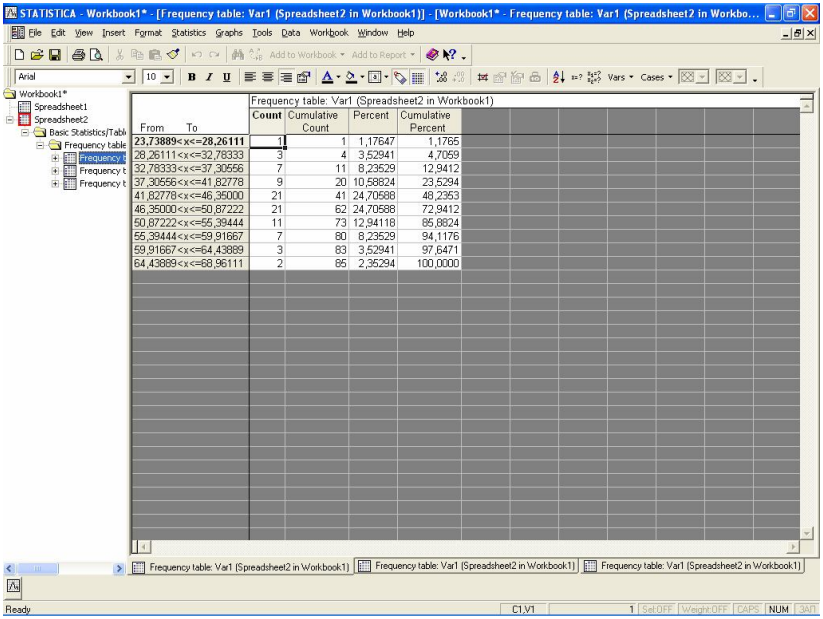


Рис. 3.14. Окно с рассчитанной таблицей частот.

Далее представлены таблицы вариационного ряда, построенные с использованием разного числа интервалов ($k = 15$, $k = 10$ и $k = 8$).

а)

	Frequency table: Var2 (Spreadsheet2 in Workbook1)				
From To	**Count**	Cumulative Count	Percent	Cumulative Percent	
24,54643<x<=27,45357	1	1	1,17647	1,1765	
27,45357<x<=30,36071	2	3	2,35294	3,5294	
30,36071<x<=33,26786	1	4	1,17647	4,7059	
33,26786<x<=36,17500	4	8	4,70588	9,4118	
36,17500<x<=39,08214	5	13	5,88235	15,2941	
39,08214<x<=41,98929	8	21	9,41176	24,7059	
41,98929<x<=44,89643	12	33	14,11765	38,8235	
44,89643<x<=47,80357	16	49	18,82353	57,6471	
47,80357<x<=50,71071	13	62	15,29412	72,9412	
50,71071<x<=53,61786	8	70	9,41176	82,3529	
53,61786<x<=56,52500	6	76	7,05882	89,4118	
56,52500<x<=59,43214	3	79	3,52941	92,9412	
59,43214<x<=62,33929	3	82	3,52941	96,4706	
62,33929<x<=65,24643	1	83	1,17647	97,6471	
65,24643<x<=68,15357	2	85	2,35294	100,0000	

б)

	Frequency table: Var2 (Spreadsheet2 in Workbook1)				
From To	**Count**	Cumulative Count	Percent	Cumulative Percent	
23,73889<x<=28,26111	1	1	1,17647	1,1765	
28,26111<x<=32,78333	3	4	3,52941	4,7059	
32,78333<x<=37,30556	7	11	8,23529	12,9412	
37,30556<x<=41,82778	9	20	10,58824	23,5294	
41,82778<x<=46,35000	21	41	24,70588	48,2353	
46,35000<x<=50,87222	21	62	24,70588	72,9412	
50,87222<x<=55,39444	11	73	12,94118	85,8824	
55,39444<x<=59,91667	7	80	8,23529	94,1176	
59,91667<x<=64,43889	3	83	3,52941	97,6471	
64,43889<x<=68,96111	2	85	2,35294	100,0000	

Рис. 3.15. Распределение регионов России по значению показателя «Доля убыточных предприятий» в 2003г. С числом интервалов: а) $k = 15$, б) $k = 10$.

	Frequency table: Var2 (Spreadsheet2 in Workbook1)				
From To	**Count**	Cumulative Count	Percent	Cumulative Percent	
23,09286<x<=28,90714	2	2	2,35294	2,3529	
28,90714<x<=34,72143	3	5	3,52941	5,8824	
34,72143<x<=40,53571	9	14	10,58824	16,4706	
40,53571<x<=46,35000	27	41	31,76471	48,2353	
46,35000<x<=52,16429	26	67	30,58824	78,8235	
52,16429<x<=57,97857	12	79	14,11765	92,9412	
57,97857<x<=63,79286	4	83	4,70588	97,6471	
63,79286<x<=69,60714	2	85	2,35294	100,0000	

Рис. 3.16. Распределение регионов России по значению показателя «Доля убыточных предприятий» в 2003г. С числом интервалов $k = 8$.

Выбирая окончательный вариант табличного представления вариационного ряда в нашем примере, следует остановиться на группировке с использованием 8 групп. При $k = 15$ получено много малонаполненных групп и наблюдаются три вершины. Группировка с выделением 10 групп дает плосковершинное распределение (равные частоты в двух группах в центре распределения) и также наличие малонаполненных групп.

В таблицах первая непоименованная графа (*From To*) содержит интервалы значений признака («Доля убыточных организаций»)

$$x_i^{\min} < x \le x_i^{\max}$$

в каждой группе;

Count – абсолютные частоты (f_i), т.е. число единиц совокупности, обладающих указанным значением признака;

Cumulative Count – накопленные абсолютные частоты, получаемые последовательным суммированием частот по группам. Сумма накопленных частот по каждой строке означает, какое количество единиц совокупности (регионов) имеет значение признака, не превышающее значения верхней границы данного интервала. Общая сумма накопленных частот соответствует объему изучаемой совокупности (85);

Percent – частости (относительные частоты, w_i ; выражаются в процентах), рассчитываются:

$$w_i = \frac{f_i}{\sum_{i=1}^{k} f_i} 100, \qquad (3.2.)$$

где f_i – число единиц i - й группы;

$\sum_{i=1}^{k} f_i$ – общее число единиц в совокупности;

w_i – доля каждой группы в общем объеме совокупности, т.е. процент регионов, в которых значение признака находится в пределах данного интервала;

Cumulative percent - накопленные частости – это результат последовательного суммирования относительных частот по группам, итоговая сумма, очевидно, равна 100%.

Табличное представление вариационного ряда позволяет получить подробную информацию о составе и структуре изучаемой совокупности, т.е. определить какое количество единиц изучаемой совокупности обладает тем или иным значением признака и какова доля этой группы единиц в общем объеме совокупности, а также выявить закономерность изменения частот.

На основе таблиц строятся графики, наглядно представляющие закономерность распределения анализируемой статистической совокупности. Графическое представление может быть осуществлено как с использованием абсолютных, так и относительных частот.

Традиционно для изображения вариационных рядов распределения в отечественной практике используются графики: гистограмма, полигон, кумулята. Наряду с этим, STATISTICA дает возможность получать графические представления эмпирического распределения, широко используемые в зарубежной статистической литературе, как учебной, так и профессиональной. Речь идет о графиках *Box-and-Whisker Plot*, *Hanging Bars*.

Для построения полигона на основе абсолютных частот необходимо выделить столбец *Count* в таблице частот и щелкнуть на нем правой кнопкой мыши. В появившемся меню выбирается *Graphs of block data/Line Plot: Entire columns* (построение графика с использованием выделенного столбца) (рис. 3.17.).

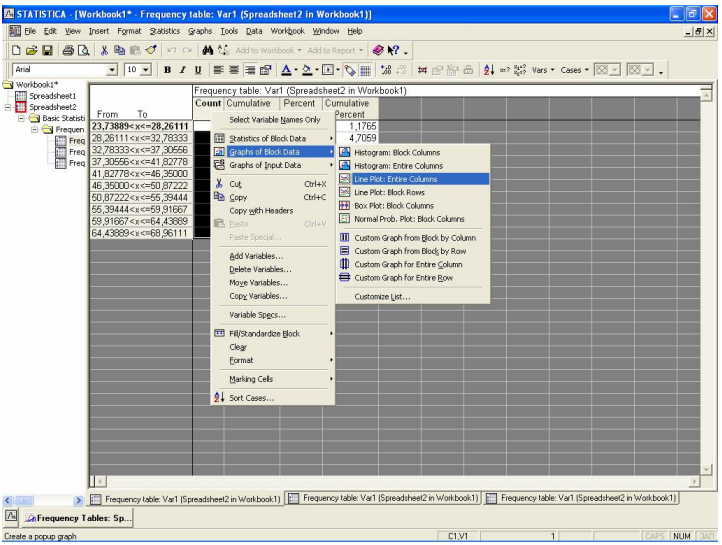


Рис. 3.17. Окно с функциями, необходимыми для построения полигона.

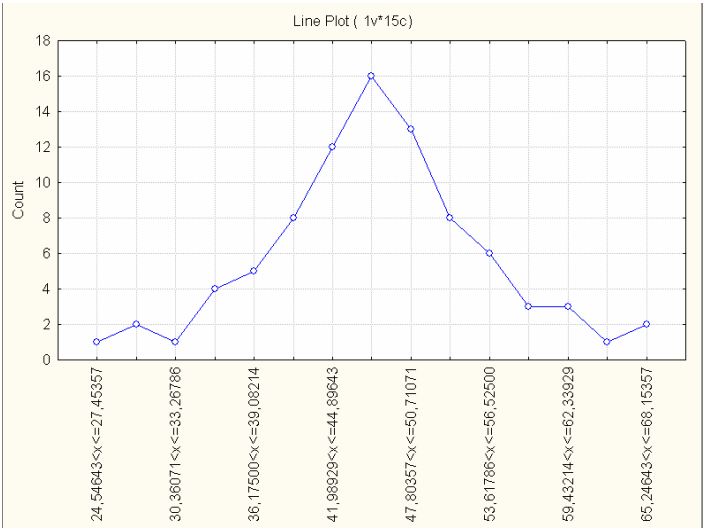


Рис. 3.18. Полигон распределения регионов России по значению показателя «Доля убыточных предприятий» в абсолютных частотах ($k = 15$).

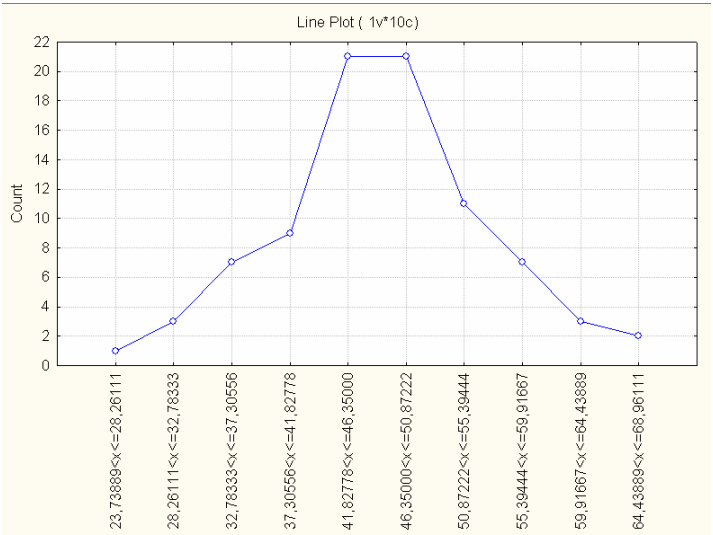


Рис. 3.19. Полигон распределения регионов России по значению показателя «Доля убыточных предприятий» в 2003г. (k =10).

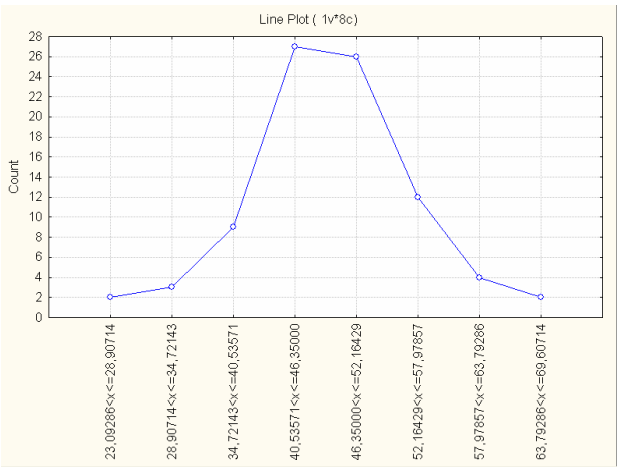


Рис. 3.20. Полигон распределения регионов России по значению показателя «Доля убыточных предприятий» в 2003г. (k =8).

Для построения полигона по относительным частотам, кумуляты по абсолютным и относительным частотам выбираются соответственно столбцы *Percent, Cumulative count , Cumulative percent* в таблице частот.

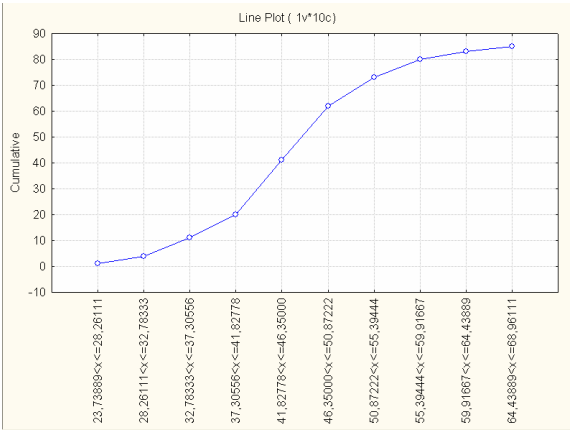


Рис. 3.21. Кумулята распределения регионов России по значению показателя «Доля убыточных предприятий» в 2003г. $k = 10$ (абсолютные частоты).

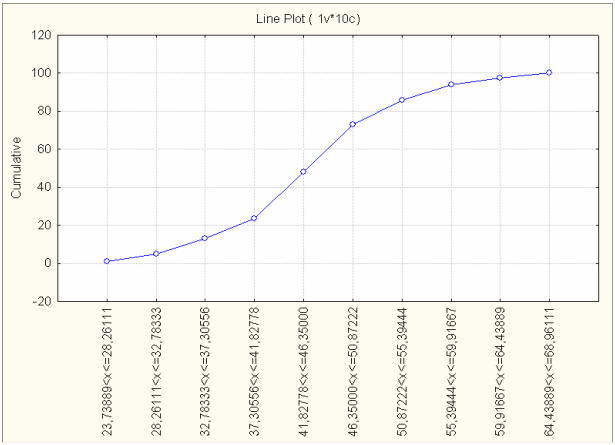


Рис. 3.22. Кумулята распределения регионов России по значению показателя «Доля убыточных предприятий» в 2003г. $k = 10$ (относительные частоты).

Также для графического представления ряда распределения может быть использована гистограмма. Для ее построения удобно воспользоваться кнопкой *Histograms* на закладке *Advanced* меню *Frequency Tables* (см. рис. 3.10., 3.23.), которым мы пользовались для построения таблиц. При этом условия построения гистограмм должны полностью соответствовать условиям построения таблиц (см. поля *No. of exact intervals*, закладку *Options*).

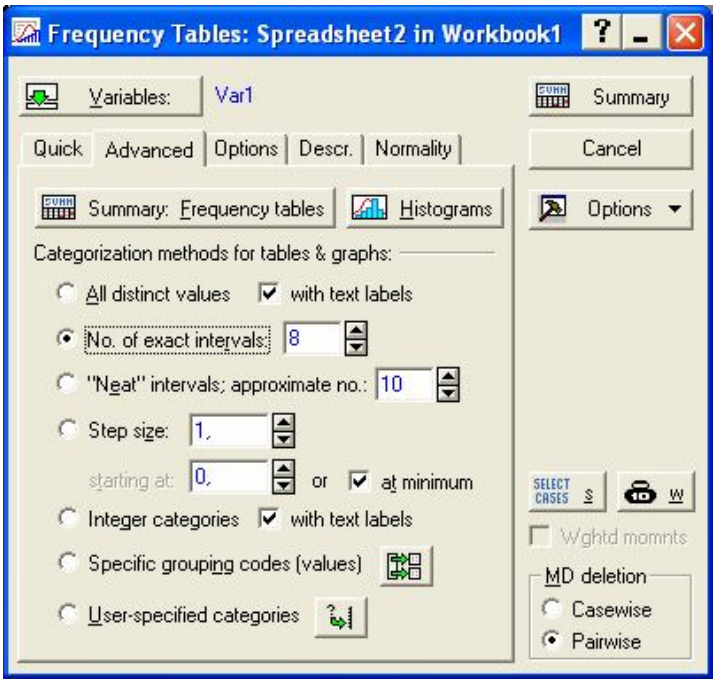
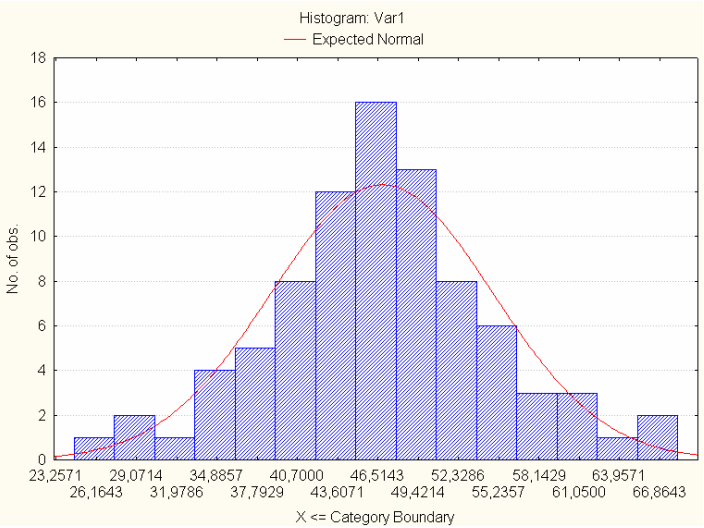


Рис. 3.23. Закладка Advanced меню Frequency Tables.

На построенных графиках (рис.3.24., 3.25., 3.26.) помимо гистограммы нанесена кривая нормального распределения (обозначена красным цветом). В контексте данной задачи она рассматриваться не будет, но это единственный способ получить гистограммы, отражающие данные построенных ранее таблиц.

а)



б)

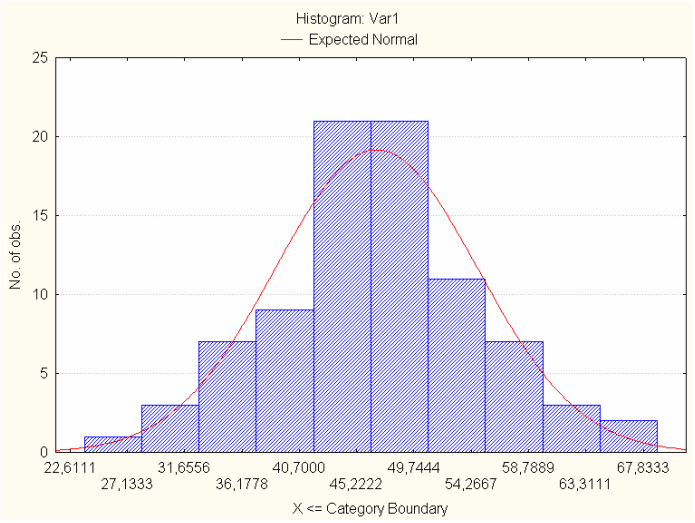


Рис. 3.24. Гистограммы распределения регионов России по значению показателя «Доля убыточных предприятий» в 2003г. с наложенными на них кривыми нормального распределения: с числом интервалов а) $k=15$, б) $k=10$.

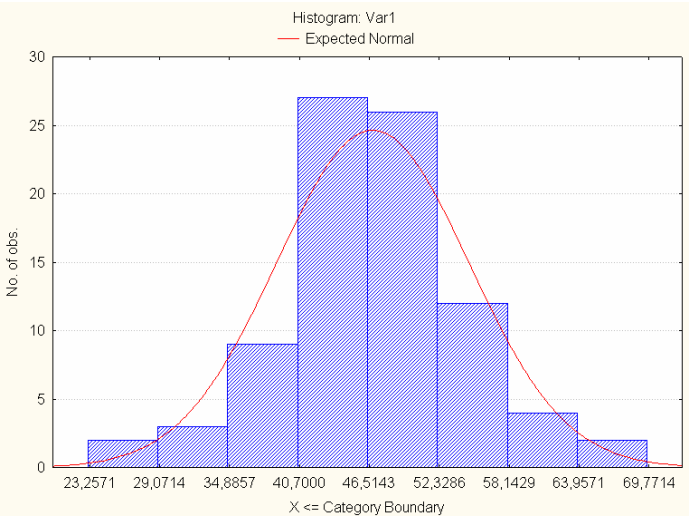


Рис. 3.25. Гистограммы распределения регионов России по значению показателя «Доля убыточных предприятий» в 2003г. с наложенными на них кривыми нормального распределения: с числом интервалов k =8.

Как и в других ППП, в STATISTICA процедура построения гистограмм выделена в отдельное меню. Однако алгоритм группировки данных в этой процедуре отличен от алгоритма при построении таблиц. Поэтому здесь мы лишь рассмотрим этот метод построения гистограмм, но результаты анализа использовать не будем. Этой процедурой удобно пользоваться в условиях, когда не требуется построение таблиц.

Для построения гистограммы таким способом удобнее всего воспользоваться меню *Graphs/Histograms* (рис. 3.26, 3.27.).

Как и в других задачах, предварительно выбирается анализируемая переменная (кнопка *Variables*) Var1. Снимем метку с поля *Normal* (установлена по умолчанию), так как на этом этапе анализа не решается задача сглаживания эмпирического распределения нормальным. Далее задается число интервалов распределения, для этого ставится метка на поле *Categories* (число интервалов) и задается число интервалов. Для завершения настройки процедуры построения гистограммы необходимо перейти к закладке *Advanced* (рис. 3.28.), где можно выбрать тип гистограммы (*Graph type*):

- *Regular* – обычная гистограмма;
- *Multiple* – наложение нескольких гистограмм на один график;
- *Double-Y* – с двумя разномасштабными осями ординат (для одновременного представления двух переменных).

В поле вывода типа гистограммы (*Showing type*) можно выбрать:

- *Standard* – вывод гистограммы в обычном виде (с использованием абсолютных или относительных частот).
- *Cumulative* – гистограмма накопленных частот.
- *Hanging Bars* – описание данной функции см. на стр. 60.

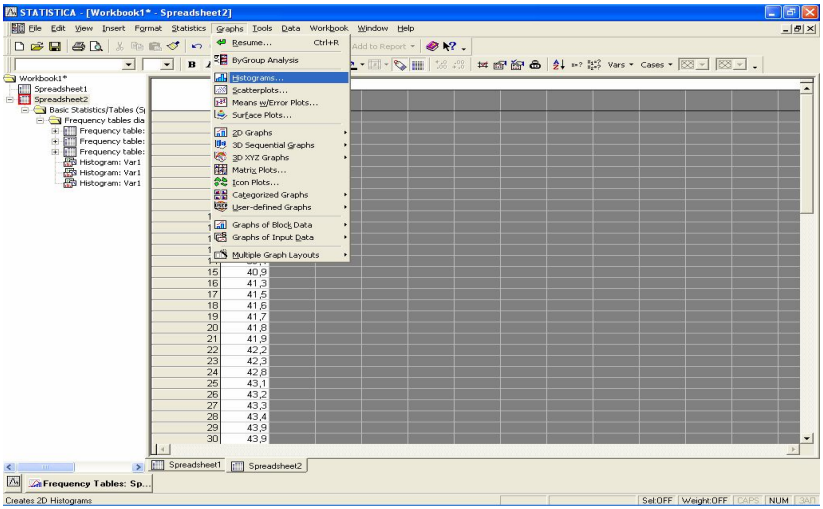


Рис. 3.26. Окно выбора меню построения гистограмм.

В поле *Break between columns* задаются разрывы между столбцами гистограммы, что бывает полезно в демонстрационных целях.

Show percentages – выводит на каждом столбце значение относительных частот в процентах.

Меню *Y axis* позволяет выбрать вариант маркировки оси ординат (работает только для *Regular* и *Multiple*):

N – абсолютные частоты;

% – относительные частоты;

N&%, %&N – их комбинация.

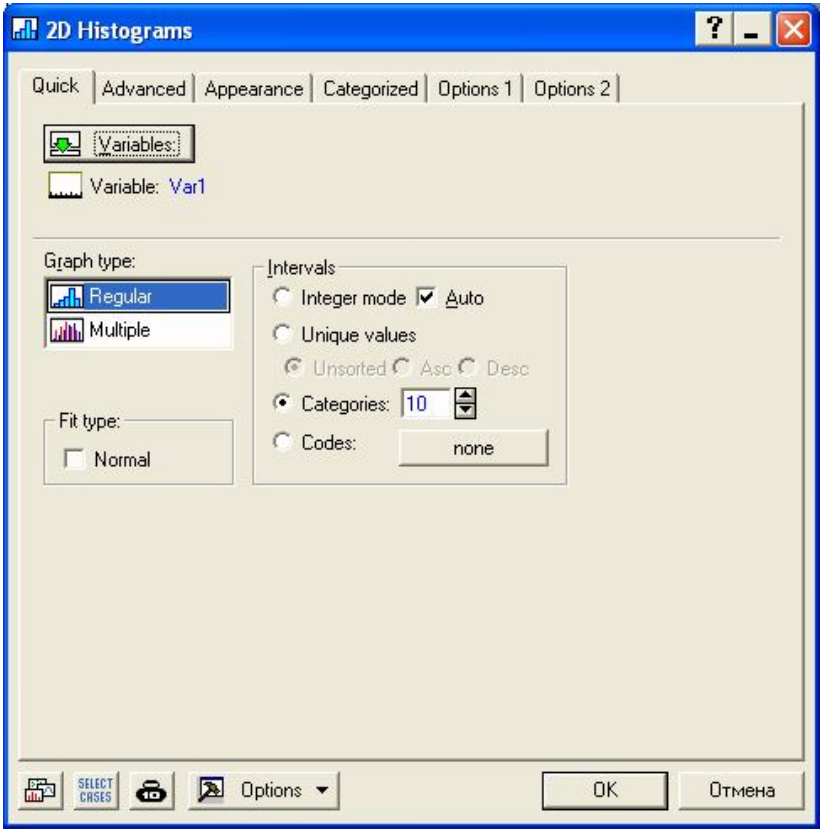


Рис. 3.27. Окно меню для построения гистограмм.

С помощью меню *Fit type* можно наложить на гистограмму некоторые виды теоретических распределений. Поле *Descriptive statistics* выводит под гистограммой основные статистики, поле *Total count* предоставляет возможность указать под гистограммой объем совокупности. В меню *Intervals* ставится нужная метка в поле *Categories* и задается необходимое число интервалов.

Рис. 3.28. Окно выбора параметров построения гистограмм.

Далее переходим к закладке *Options 1* и в поле *Common title* вводим название гистограммы. Затем нажимаем *OK*. (рис. 3.29.)

Автоматически устанавливается рекомендуемое системой число интервалов. В нашем случае – 10. Чтобы построить гистограмму с другим числом интервалов, необходимо ввести его в поле *Categories*.

В методических целях, для получения наглядного представления об изменении формы графика распределения при изменении числа интервалов группировки, строится несколько графиков с различным числом интервалов. По оси ординат расположены абсолютные частоты.

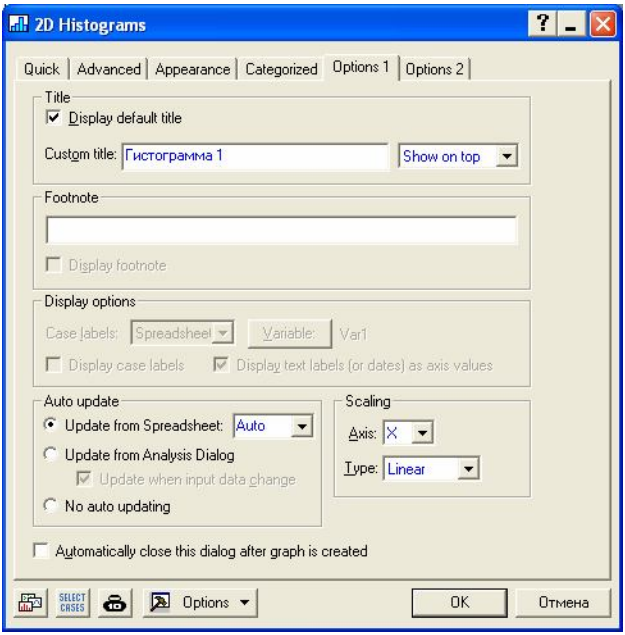


Рис. 3.29. Окно настройки подписей гистограммы.

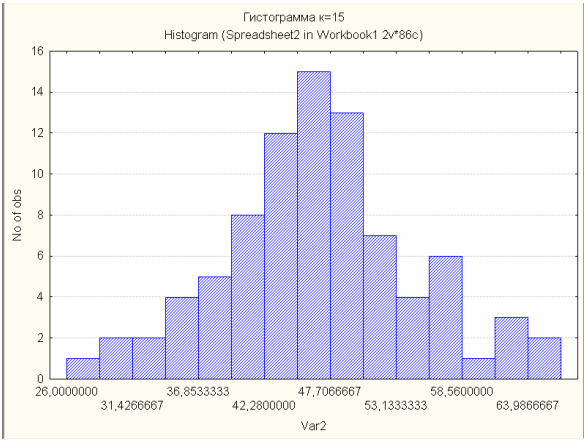


Рис. 3.30. Гистограмма распределения регионов России по значению

показателя «Доля убыточных предприятий» в 2003г. (k = 15).

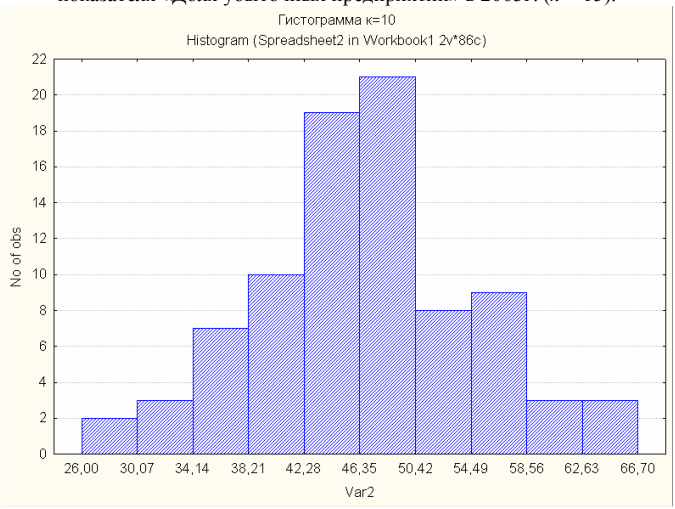


Рис. 3.31. Гистограмма распределения регионов России по значению показателя «Доля убыточных предприятий» в 2003г. (k = 10).

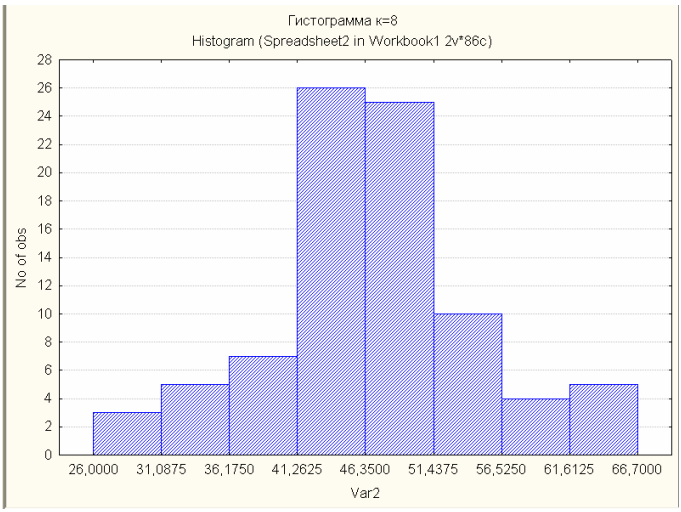


Рис. 3.32. Гистограмма распределения регионов России по значению показателя «Доля убыточных предприятий» в 2003г. (k =8).

ППП STATISTICA предоставляет пользователю широкие возможности по настройке и редактированию различных типов графиков. Если щелкнуть правой кнопкой мыши в свободном поле графика и выбрать в появившемся подменю свойства графика (*Graph Properties (All Options)*) (рис. 3.33.), то на экране появится диалоговое окно, состоящее из 17 закладок для настройки определенных элементов графика (рис. 3.34.). Настройке поддаются практически все элементы графика. Например, с помощью закладки *Graph Titles/Text* можно редактировать оформление текстовых составляющих графика: их размер, цвет, стиль, выравнивание и так далее. А закладка *Plot: General* позволяет выбирать тип, размер и цвет линий графика, маркеры, заливки и другие художественные элементы графика. Настройка всех этих элементов соответствует интерфейсу основных Windows-приложений (Word, Excel).

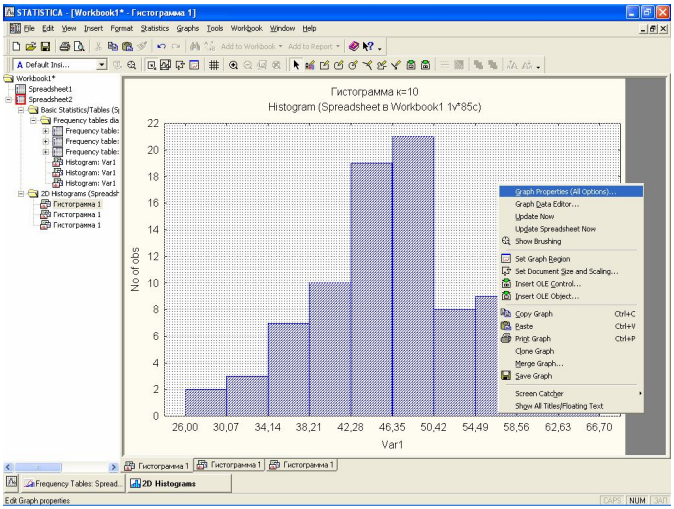


Рис. 3.33. Выбор меню настроек графических изображений.

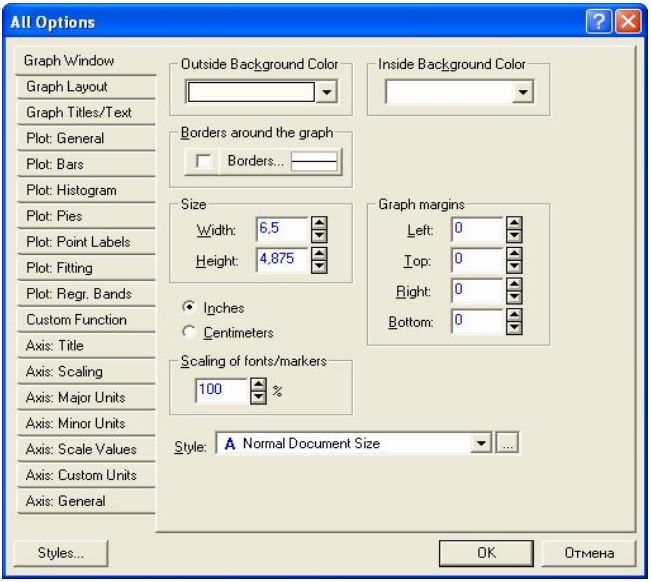


Рис. 3.34. Окно настройки графических изображений.

Для более полного представления графических возможностей системы в контексте анализа распределений приведем примеры графиков, не ставших рутинными инструментами в отечественной практической статистике.

Одним из приемов компактного изображения статистической совокупности, находящимся вне отечественной традиции, является "Box-and-Whisker Plot" — "ящик с усами". Рассматриваемая процедура обеспечивает как диагностическую, так и описательную информацию об исследуемой совокупности.

Для ее реализации запускаем процедуру *Graphs/2D Graphs/Box Plots* (рис. 3.35.). В появившемся окне (рис. 3.36.) удобно сразу же выбрать интересующий нас тип графика (в поле *Graph Type* выбираем *Box-Whiskers*). Остальные свойства графика удобнее всего настроить, перейдя на закладку *Advanced* (рис. 3.37.). Изначально выбираем переменную с помощью кнопки *Variables* (рис. 3.38.). В нашем случае необходимо выбрать переменную Var1 в поле *Dependent variable*.

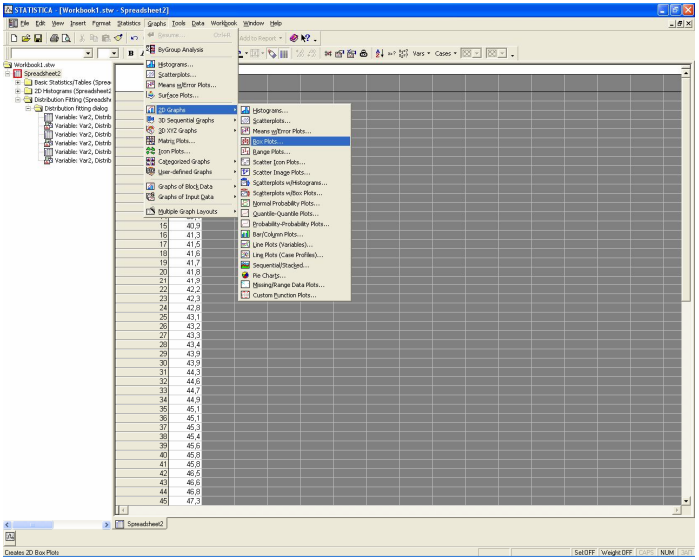


Рис. 3.35. Запуск процедуры построения Box-and-Whisker Plot.

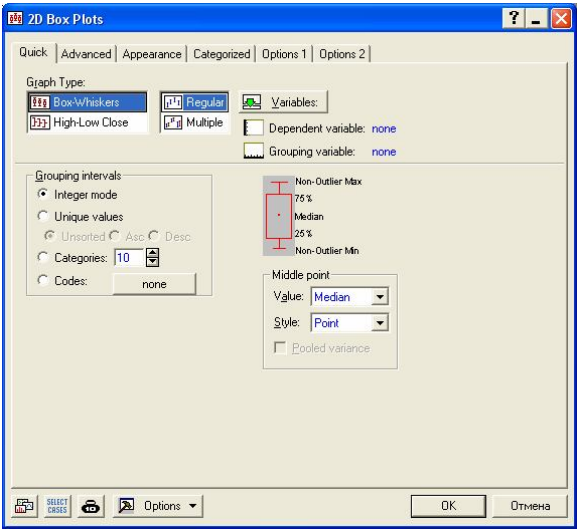


Рис. 3.36. Вид закладки *Quick* процедуры *2D Box Plots*.

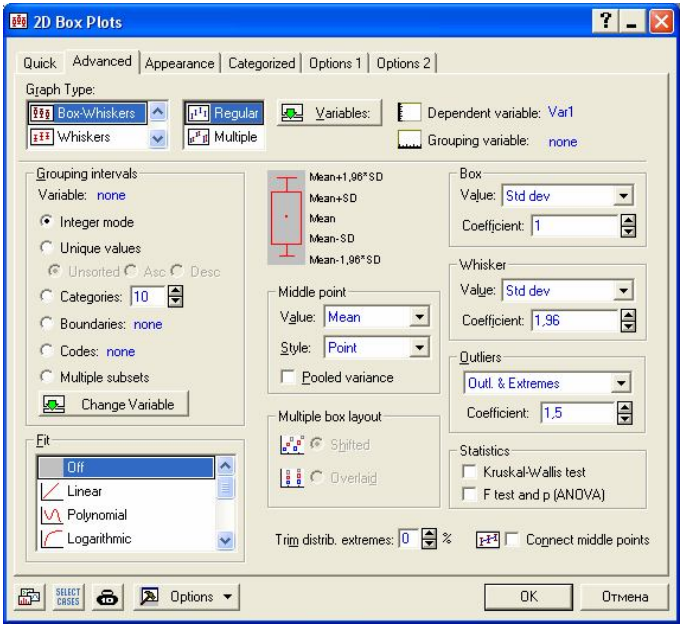


Рис. 3.37. Вид закладки *Advanced* процедуры *2D Box Plots*.

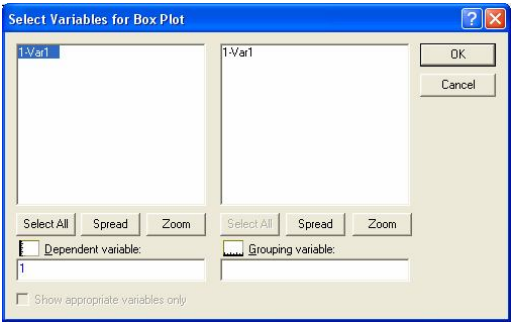


Рис. 3.38. Окно выбора переменных.

Поскольку у нас отсутствует группировочная переменная, поле *Grouping intervals* не рассматривается, то же самое можно сказать о полях *Fit* и *Multiple box layout*, поскольку мы оперируем только одной

переменной (см. рис. 3.37.). Обратим внимание на возможные варианты вывода интересующего нас графика.

Раздел *Middle point* содержит поля *Value* и *Style*. Первое позволяет выбрать центральное значение (точку отсчета) «ящика» и имеет два варианта:

- *Mean* – средняя арифметическая;
- *Median* – медиана.

Второе поле позволяет выбрать способ отображения этих значений:

- *Point* – в виде точки;
- *Line* – в виде линии.

Далее рассмотрим настройку границ «ящика» с помощью раздела *Box*. Первое поле *Value* показывает, какая величина будет прибавлена к значению средней арифметической или медианы при определении границ «ящика». Поле *Coefficient* задает коэффициент, на который умножается эта величина.

Если в качестве центральной точки выбрана средняя арифметическая, то возможны следующие варианты границ «ящика»:

- *Std dev* – среднеквадратическое отклонение;
- *Std error* – средняя ошибка;
- *Conf. interval* – доверительный интервал;
- *Min-Max* – минимальное и максимальное значение;
- *Constant* – постоянная величина.

Если в качестве центральной точки выбрана медиана, то доступны следующие варианты:

- *Min-Max*;
- *Constant*;
- *Percentiles* – процентили.

Для определения длины «усов» используется раздел *Whiskers*. Его содержание также зависит от выбранной точки отсчета. Если выбрана средняя арифметическая, то содержание поля *Value* аналогично одноименному поля в разделе *Box*, за исключением добавившегося пункта *Non-outlier range*, который означает прибавление длины «ящика» к центральной точки для определения границ «усов». Такая же ситуация наблюдается при выборе медианы в качестве точки отсчета. Поле *Coefficient* имеет те же функции, что и одноименное в разделе *Box*.

Теперь остается настроить раздел *Outliers*, с помощью которого на график можно наносить (или не наносить) определенные точки двух видов:

1. Outliers – точки, которые находятся далеко от центра распределения и не являются характерными для него (возможно, являются результатами ошибок наблюдения или выбросами). Определяются следующим образом:

- значения, больше чем (верхняя граница «ящика» + коэффициент*(верхняя граница «ящика» – нижняя граница «ящика»));
- значения, меньше чем (нижняя граница «ящика» – коэффициент*(верхняя граница «ящика» – нижняя граница «ящика»)).

Коэффициент задается пользователем в поле *Coefficient* раздела *Outliers*.

2. *Extremes* – определение точек тоже, но в формуле их определения коэффициент умножается на 2.

Напомним, что на закладке *Options 1* можно задать название графика, а его построение запускается кнопкой *OK*.

График появляется именно в таком виде (рис. 3.39.) однако, на практике принято рассматривать его горизонтально. Для того чтобы повернуть график на 90 градусов, нужно щелкнуть в поле графика правой кнопкой мыши и выбрать меню *Graph Properties (All options)* (рис. 3.40.).

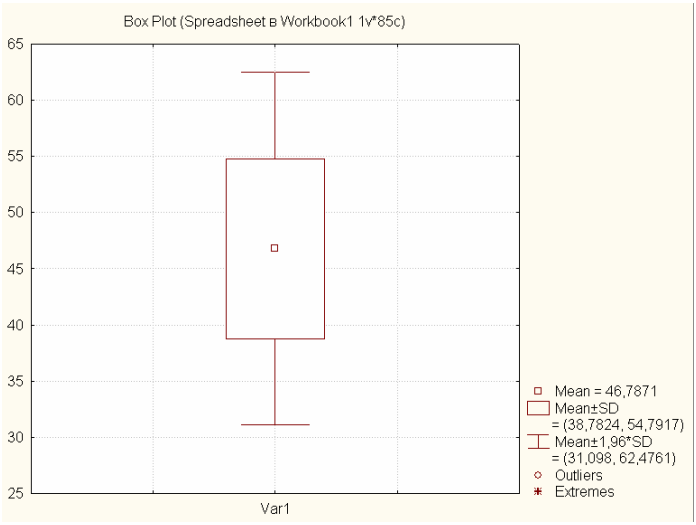


Рис. 3.39. Результат вывода диаграммы “Box & Whisker Plot” для переменной Var1.

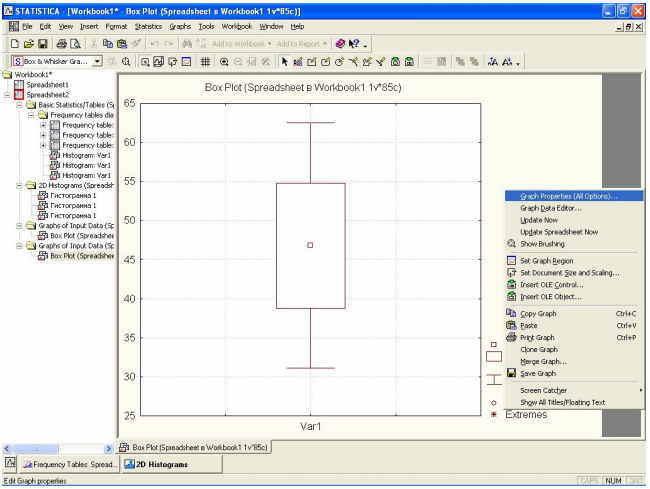


Рис. 3.40. Выбор меню задания свойств графика.

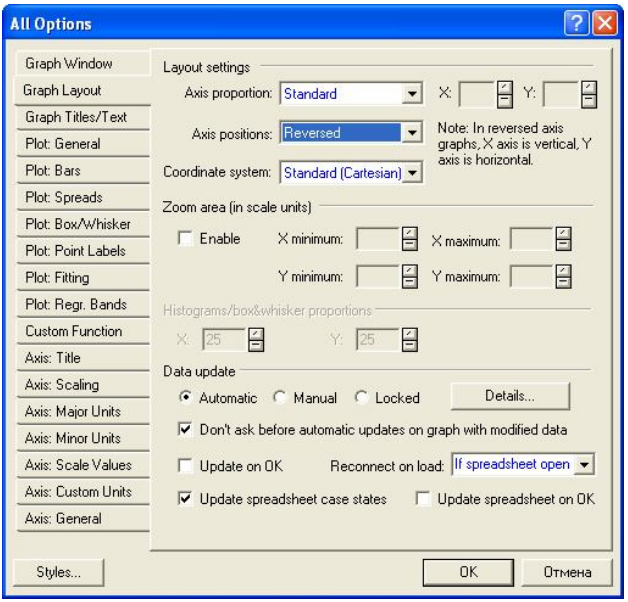


Рис. 3.41. Функция поворота графического изображения на 90 градусов.

Далее переходим к закладке *Graph layout*: в поле *Axis position* вместо функции *Standard* выбираем *Reserved*, то есть обратное положение оси абсцисс. Нажимаем *OK* (рис. 3.41.).

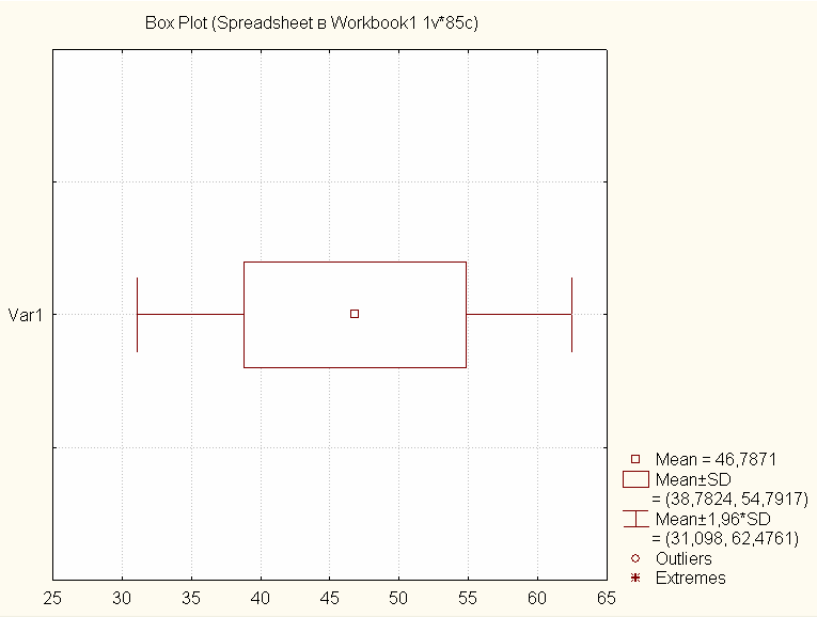


Рис. 3.42. Результат поворота диаграммы “Box & Whisker Plot” на 90 градусов.

Центральный прямоугольник на графике это и есть тот "Box", из-за которого метод получил свое название. Здесь в центре ящика находится значение средней арифметической (*Mean*), правая и левая границы прямоугольника определяют интервал $\bar{x} \pm \sigma$ (*Mean* $\pm$ *SD*).

Длина «усов» определяется как $\bar{x} \pm 1{,}96\sigma$ (*Mean* $\pm$ *1.96 SD*). Элементы совокупности, выпадающие за установленные границы, отмечаются графически (точками, звездочками) и интерпретируются как возможно аномальные значения, требующие внимательного анализа.

Возможен иной вариант построения “Box & Whisker Plot” (рис. 3.43.). Левая граница “ящика” соответствует нижнему квартилю распределения, правая — верхнему квартилю. Внутри ящика прямоугольной отметкой определена медиана. Справа и слева от “ящика” поме-

щены "усы". Длина "усов" определяется как 1,5 межквартильного расстояния: $(Q_3-Q_1)\times 1.5$.

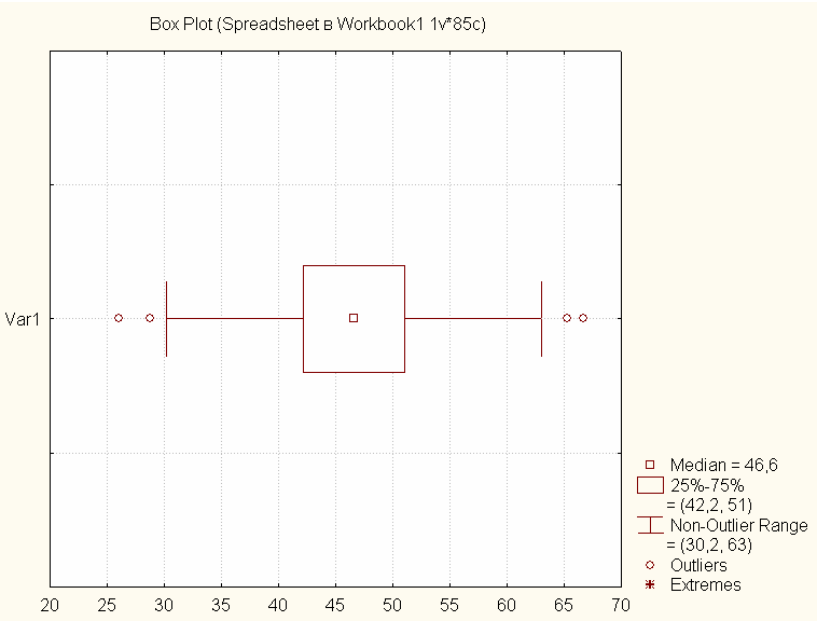


Рис. 3.43. Диаграммы “Box & Whisker Plot”.

Метод "Box-and-Whisker Plot" также дает полезную информацию о концентрации, дисперсии и асимметрии распределения, но наряду с этим исследователь получает наглядное представление о том, что происходит на концах распределения.

В качестве дополнения отметим, что система дает также возможность получить нетрадиционное графическое представление о том, как соотносятся между собой эмпирическое распределение и его нормальная аппроксимация. Речь идет о графике "Hanging Histobars" или “Hanging Bars” (в весьма вольном переводе – «висячие полоски»). Канонизированного термина на русском языке нет, потому будем обозначать рассматриваемую процедуру как "НН-график" (рис 3.44.).

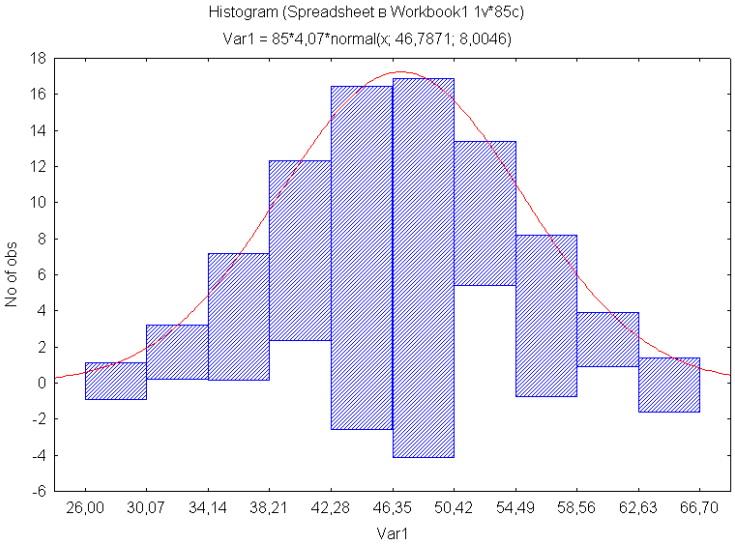


Рис. 3.44. Диаграммы “Hanging Bars” для переменной Var1.

НН-процедура выводит на экран изображение близкое к гистограмме, с тем только различием, что столбцы гистограммы не опираются на горизонтальную ось, а "подвешены" (Hanging) к кривой нормального распределения в точках, соответствующих серединам интервалов группировки. Степень близости эмпирического распределения к нормальному оценивается по отклонениям "подвешенных" элементов графика от горизонтальной оси, проходящей через 0. "Hanging Histobars", как форма сопоставления эмпирического распределения и его теоретической аппроксимации, часто используется в зарубежной статистической литературе и к встрече с ней надо быть готовым. Обращаем внимание на то, что процедура работает только в приложении к нормальному распределению. В отечественной литературе и практике используются другие изображения, о которых будет сказано ниже.

3.2. Расчет основных характеристик вариационного ряда

Статистический анализ вариационных рядов распределения предполагает расчет характеристик центра распределения, его струк-

туры, оценку степени вариации и дифференциации изучаемого признака, изучение формы распределения.

В качестве показателей центральной тенденции распределения используются: среднее арифметическое значение, мода и медиана. Основными показателями вариации являются: размах вариации, дисперсия, среднее квадратическое отклонение, коэффициент вариации. Для характеристики структуры распределения используются следующие показатели: медиана, квартили, децили и прочие перцентили. Изучение формы распределения предполагает оценку асимметрии и эксцесса (куртозиса). Перечисленные показатели имеют самостоятельное аналитическое значение, поскольку отражают разные свойства изучаемой совокупности, а все вместе они позволяют получить комплексную характеристику эмпирического распределения [см. 6,7,8].

В программе STATISTICA, как и в других статистических ППП, есть возможность получить все перечисленные показатели, пользуясь одной процедурой.

В главном меню раздел *Statistics,* активизируем опцию *Basic Statistics/Tables* (рис. 3.45.).

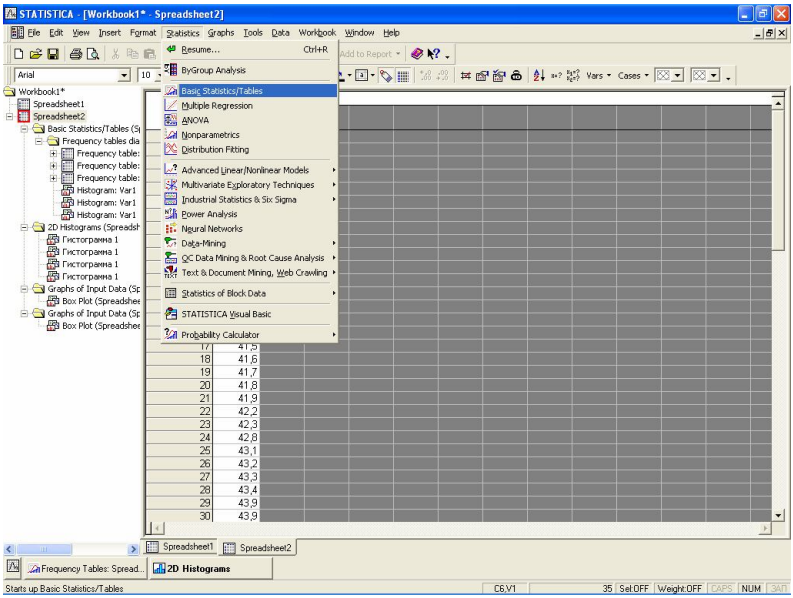


Рис. 3.45. Окно STATISTICA с выбранной функцией *Basic Statistics/Tables*.

Затем, в появившемся контекстном окне, выбирается процедура *Descriptive statistics* – описательные статистики (рис. 3.46.).

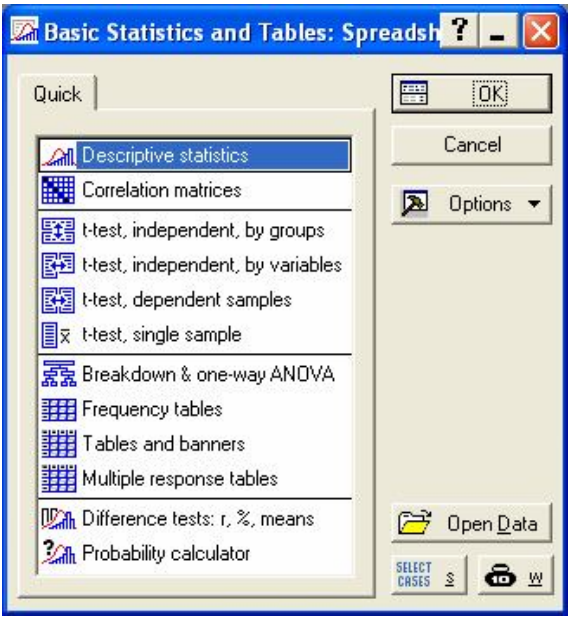


Рис. 3.46. Окно выбора процедуры в меню *Basic Statistics/Tables.*

Процедура *Descriptive statistics* предлагает пользователю широкий набор функций, а также опций к ним. Работа начинается с выбора переменной или переменных, по которым будет проведен анализ. Для этого из базы данных, вызываемой кнопкой *Variables* – переменные, выбирается нужная переменная, в рассматриваемом примере это Var1 (рис 3.47.).

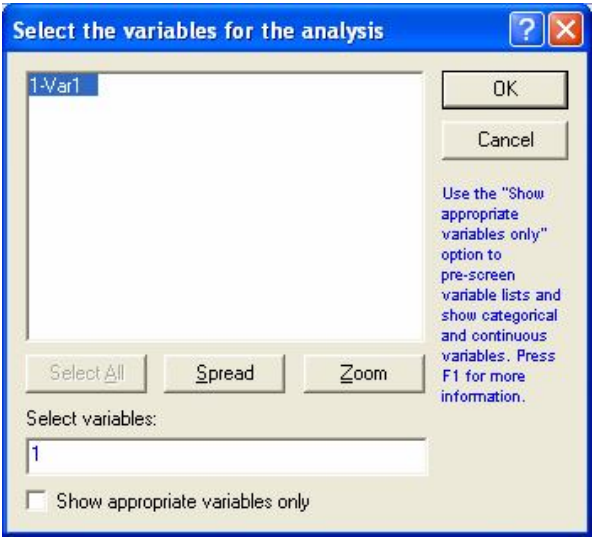


Рис. 3.47. Окно выбора переменных.

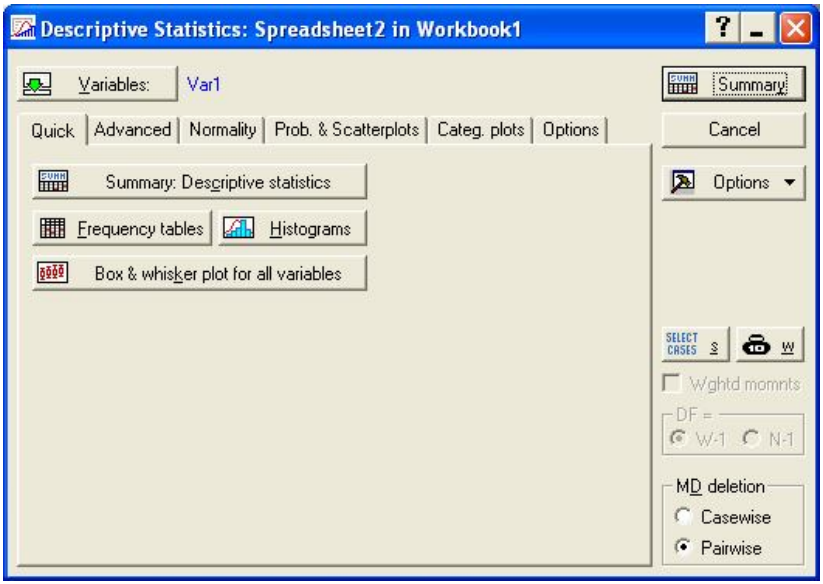


Рис. 3.48. Окно процедуры *Descriptive Statistics*.

Проведем обзор кнопок и закладок данного меню (рис. 3.48.). В правом верхнем углу находятся три кнопки, которыми можно пользоваться независимо от того, на какой закладке меню находится пользователь:

1.Summary – позволяет рассчитать основные статистики по заданным на данный момент опциям.

2.Cancel – выход из меню *Descriptives statistics*.

3.Options – позволяет выполнить несколько действий по выбору пользователя:

3.1 *Close Analysis* - выйти из процедуры текущего анализа;

3.2 *Creat macro* - создать макрос;

3.3 *Output* - задать опции для окон анализа и других окон, задать опции для вывода и сохранения выбранных процедур.

Перейдем к описанию закладок в меню *Descriptive statistics*. *Quick* - позволяет выполнять определенные функции, основываясь на опциях заданных по умолчанию, и содержит четыре кнопки:

1.*Summary: Descriptive statistics* – кнопка аналогична кнопке Summary в правом углу меню.

2. *Frequency tables* – строятся частотные таблицы с заданными по умолчанию показателями.

3. *Histograms* – строится гистограмма распределения по заданным по умолчанию показателям и с наложением на нее кривой нормального распределения.

4. *Box & whisker plot for all variables* – строится «ящик с усами». Закладки *Prob. & Scatterplots, Categ.plots, Options* не используются в анализе одномерного массива данных. Эти закладки служат для построения различных графических изображений для двух переменных, а так же для построения 3-D графиков. Последние две закладки практически полностью посвящены настройке опций для *Box & whisker plot*.

Для расчета основных статистических характеристик ряда распределения необходимо воспользоваться закладкой *Advanced* (рис. 3.49.) (дополнительно), так как нажатие кнопки *Summary* на закладке *Quick* позволяет получить лишь четыре показателя.

Закладка *Advanced* предлагает пользователю сформировать набор вычисляемых статистик, отвечающих целям анализа. При этом есть возможность выбрать все характеристики (кнопка *Select all stats*), что, естественно, и следует сделать в рамках выполняемой лабораторной работы. Отказ от вычисления тех или иных статистик реализуется

кнопкой *Reset*. Для запуска нужной процедуры необходимо воспользоваться кнопкой *Summary*: *Descriptive statistics или Summary*.

Рассчитанные статистики выводятся одной строкой, что создает трудности при оформлении отчета. В связи с этим следует обратить внимание на предоставляемую пользователю сервисную услугу, а именно транспонирование матрицы результатов. В нашем примере строка результатов должна быть транспонирована в столбец, что гораздо удобнее. Для транспонирования матрицы результатов необходимо обратиться к пункту главного меню *Data* (рис. 3.50.). Затем пройти по пути: *Data /Transponse/File*.

ВНИМАНИЕ!!! При транспонировании матрицы модальное значение исчезает и вместо него появляется число, означающее количество модальных значений в распределении. Поэтому модальное значение необходимо выписать перед транспонированием матрицы, а затем вручную занести в столбец с показателями.

Результаты расчета основных статистических характеристик представлены на рис. 3.51.

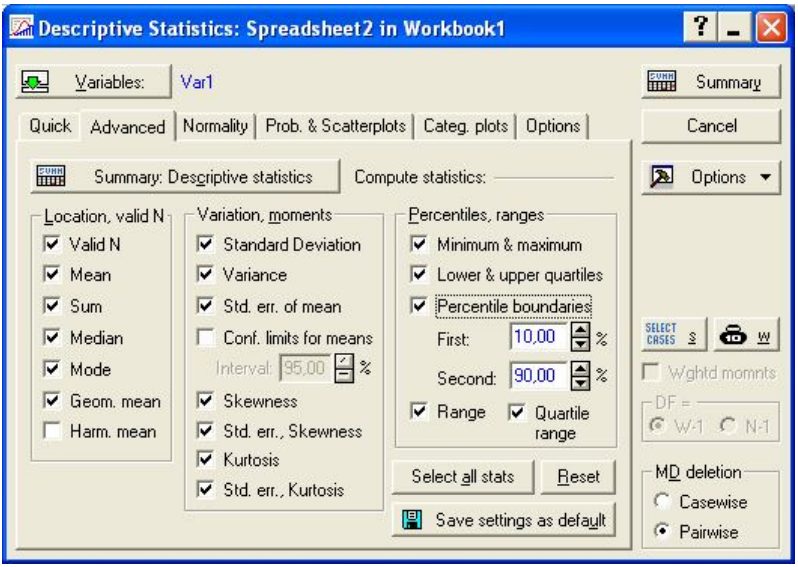


Рис. 3.49. Закладка *Advanced* в меню *Descriptive Statistics*.

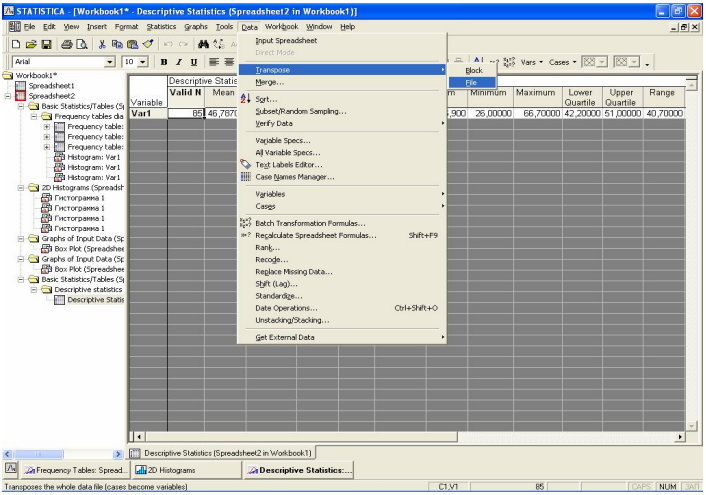


Рис. 3.50. Окно выбора меню транспонирования таблицы.

Variable	1 Var1
Valid N	85
Mean	46,78706
Median	46,60000
Mode	Multiple
Frequency	2
Sum	3976,900
Minimum	26,00000
Maximum	66,70000
Lower	42,20000
Upper	51,00000
Percentile	36,60000
Percentile	56,90000
Range	40,70000
Quartile	8,800000
Variance	64,07400
Std.Dev.	8,004623
Skewness	0,033423
Std.Err.	0,261153
Kurtosis	0,332543
Std.Err.	0,516756

Рис. 3.47. Основные характеристики распределения регионов России по значению показателя «Доля убыточных предприятий» в 2003г.

Приведем формулы полученных статистических характеристик:

Valid N – объем выборки (число единиц в совокупности).

Mean – средняя арифметическая:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i , \qquad (3.3.)$$

где x_i - значение признака у i-й единицы совокупности;

n – объем совокупности (Valid N).

Geometric mean – геометрическая средняя:

$$\tilde{x}_{геом} = \sqrt[n]{\prod_{i=1}^{n} x_i} . \qquad (3.4.)$$

Median – медиана:

$$M_e = \frac{x_{\frac{n}{2}} + x_{\frac{n}{2}+1}}{2}, \quad \text{если n — четное,} \qquad (3.5.)$$

$$M_e = x_{\frac{n+1}{2}}, \quad \text{если n — нечетное.} \qquad (3.6.)$$

Mode –мода (M_o) определяется непосредственно по исходным данным (запись в строке Multiple означает, что распределение имеет не одну моду).

Frequency – частота модального значения.

Sum – сумма значений признака в совокупности: $\sum_{i=1}^{n} x_i$.

Variance – дисперсия:

$$\sigma^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2, \qquad (3.7.)$$

где $\bar{x}$ - средняя арифметическая.

Standard deviation – среднее квадратическое (стандартное) отклонение:

$$\sigma = \sqrt{\sigma^2} \quad . \qquad (3.8.)$$

Standard (Standard error) – средняя (стандартная) ошибка выборки:

$$\mu = \frac{\sigma}{\sqrt{n}} . \qquad (3.9.)$$

Minimum – минимальное значение признака в совокупности: $x_{\min}$.

Maximum – максимальное значение признака в совокупности: $x_{\max}$.

Range – размах вариации: $R = x_{\max} - x_{\min}$. (3.10.)

Lower (Lower quartile) – нижний (первый) квартиль:

$$Q_1 = \frac{x_i + x_j}{2}, \qquad (3.11.)$$

$$i = floor\left[\frac{n+1}{4}\right], \qquad j = ceiling\left[\frac{n}{4}\right],$$

где *floor* — округление до ближайшего целого;
ceiling — округление до ближайшего большего.

Upper (Upper quartile) – верхний (третий) квартиль:

$$Q_3 = \frac{x_i + x_j}{2},$$

$$i = floor\left[\frac{3(n+1)}{4}\right],\ j = ceiling\left[\frac{3n}{4}\right].$$

Quartile (Interquartile range) – межквартильный размах: $Q_3 - Q_1$.

Skewness – асимметрия:

$$A_S = \frac{n\sum_{i=1}^{n}(x_i - \bar{x})^3}{(n-1)(n-2)\sigma^3}. \qquad (3.12.)$$

Std.err. (Standardized skewness) – стандартизованная асимметрия:

$$\sigma_{A_S} = \sqrt{\frac{6}{n}}; t_{A_S} = \frac{A_s}{\sigma_{A_S}}. \qquad (3.13.)$$

Kurtosis – коэффициент эксцесса (куртозис):

$$E_x = \frac{n(n+1)\sum_{i=1}^{n}(x_i - \bar{x})^4}{(n-1)(n-2)(n-3)\sigma^4} - \frac{3(n-1)^2}{(n-2)(n-3)}. \qquad (3.14.)$$

Std.err. (Standardized kurtosis) – стандартизованный куртозис:

$$\sigma_{E_X} = \sqrt{\frac{24}{n}}; t_{E_X} = \frac{E_x}{\sigma_{E_X}}. \qquad (3.15.)$$

Особо отметим для внимательного и заинтересованного читателя то, что кажущаяся громоздкость формул для A_s и E_x, легко объяснима. Введение дополнительных множителей на основе объема выборки n являются результатом успешной попытки известных американских статистиков Снедекора и Кохрана (Snedecor G.W. , Cochran W.G.)[3] учесть заметную на малых выборках смещенность выборочной дисперсии.

Рассмотрим, например, три выборки: $n_1 = 10$, $n_2 = 50$, $n_3 = 100$. Рассчитаем для них множители, содержащие n, по формулам коэффициента асимметрии в двух вариантах – канонической ($A_s^{'}$) и "по Снедекору" ($A_s^{''}$).

$$A_s^{'} = \frac{\mu_3}{\sigma^3} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^3}{n\sigma^3};\ A_s^{''} = \frac{n\sum_{i=1}^{n}(x_i - \bar{x})^3}{(n-1)(n-2)\sigma^3} \qquad (3.16.)$$

В первой формуле множитель равен $\frac{1}{n}$, во второй - $\frac{n}{(n-1)(n-2)}$.

Для первой выборки $\frac{1}{n} = \underline{0.1000}$, $\frac{n}{(n-1)(n-2)} = \underline{0,1389}$.

Для второй выборки $\frac{1}{n} = \underline{0.0200}$, $\frac{n}{(n-1)(n-2)} = \underline{0,0212}$.

Для третьей выборки $\frac{1}{n} = \underline{0,0100}$, $\frac{n}{(n-1)(n-2)} = \underline{0,0103}$.

Сходимость значений коэффициентов с увеличением объема выборки очевидна. Оценку поправки Снедекора для куртозиса оставляем за читателем.

[3] Snedecor G.W. and Cochran W.G. , Statistical Methods, Sixth edition, Ames, Iowa: Iowa State University Press, 1967.

Среди рассчитанных характеристик нет такого важного показателя вариации, как коэффициент вариации (принято рассчитывать в процентах):

$$V = \frac{\sigma}{\overline{x}} \times 100 \ (\%) \qquad (3.17.)$$

На основе этого показателя делается вывод об однородности или неоднородности совокупности по изучаемому признаку. Он легко может быть рассчитан, поскольку необходимые для его вычисления значения средней величины и стандартного отклонения имеются.

В пояснительной записке к лабораторной работе представленные выше статистические характеристики необходимо объединить в группы, в соответствии с решаемыми ими аналитическими задачами, и дать содержательную интерпретацию полученных значений, используя в качестве теоретической базы конспекты лекций и учебную литературу (см. Приложение 1).

Далее будут приведены формулы для расчета некоторых показателей при представлении данных в виде интервального вариационного ряда.

Взвешенная средняя арифметическая:

$$\overline{x} = \frac{\sum_{i=1}^{k} x_i' \times f_i}{\sum_{i=1}^{k} f_i}, \qquad (3.18.)$$

где f_i – абсолютные частоты, $\sum f_i = n$;

x'_i – середина интервала:

$$x_i' = \frac{x_i^{\min} + x_i^{\max}}{2}, \qquad (3.19.)$$

где $x_i^{\min}$ – нижняя граница i-го интервала;

$x_i^{\max}$ – верхняя граница i-го интервала.

Медиана:

$$M_e = x_{M_e}^{\min} + h \times \frac{\frac{1}{2}\sum_{i=1}^{k} f_i - F_{M_e-1}}{f_{M_e}}, \qquad (3.20.)$$

где $x_{M_e}^{\min}$ – нижняя граница медианного интервала;

h – величина группировочного интервала;

f_{M_e} – абсолютная частота медианного интервала;

F_{M_e-1} – накопленные (кумулятивные) частоты интервалов, предшествующих медианному.

Для расчёта квартилей в интервальном вариационном ряду распределения применяются следующие формулы:

$$Q_1 = x_{Q_1}^{\min} + h \times \frac{\frac{1}{4}\sum_{i=1}^{k} f_i - F_{Q_1-1}}{f_{Q_1}}, \qquad (3.21.)$$

$$Q_3 = x_{Q_3}^{\min} + h \times \frac{\frac{3}{4}\sum_{i=1}^{k} f_i - F_{Q_3-1}}{f_{Q_3}}, \qquad (3.22.)$$

где Q_1 и Q_3 – нижний и верхний квартили;

$x^{\min}_{Q_1}, x^{\min}_{Q_3}$ – нижние границы квартильных интервалов;

h – величина группировочного интервала;

f_{Q_1, Q_3} – абсолютные частоты квартильных интервалов;

F_{Q_1-1, Q_3-1} – накопленные (кумулятивные) частоты интервалов, предшествующих квартильным.

Мода:

$$M_o = x^{\min}_{M_o} + h \frac{f_{M_o} - f_{M_o-1}}{(f_{M_o} - f_{M_o-1}) + (f_{M_o} - f_{M_o+1})}, \qquad (3.23.)$$

где $x^{\min}_{M_o}$ – нижняя граница модального интервала;

f_{M_o} – частота модального интервала;

f_{M_o-1} – частота интервала перед модальным;

f_{M_o+1} – частота интервала после модального;

h – величина группировочного интервала.

Величина дисперсии по сгруппированным данным определяется:

$$\sigma^2 = \frac{\sum_{i=1}^{k} (x_i^{'} - \bar{x})^2 \times f_i}{\sum_{i=1}^{k} f_i}, \qquad (3.24.)$$

где $x_i^{'}$ – середина i-го интервала;

$\bar{x}$ – средняя арифметическая величина признака в изучаемой совокупности.

f_i – абсолютные частоты i-го интервала.

Далее предоставим пользователю возможность сравнить значения показателей, полученных в системе и рассчитанных вручную по сгруппированным данным (табл. 3.2).

Таблица 3.2

Сравнение статистических показателей, рассчитанных различными способами

№	Название показателя	Значение в ППП STATISTICA	Значение после ручного расчета
1	Средняя арифметическая		
2	Медиана		
3	Мода		
4	Дисперсия		
5	Верхний квартиль		
6	Нижний квартиль		
7			
8			

3.3. Сглаживание эмпирического распределения, проверка гипотезы о законе распределения

Процедура выравнивания, сглаживания анализируемого распределения заключается в замене эмпирических частот теоретическими, определяемыми по формуле теоретического распределения, но с учетом фактических значений переменной. На основе сопоставления эмпирических и теоретических частот рассчитываются критерии согласия, которые используются для проверки гипотезы о соответствии исследуемого распределения тому или иному типу теоретического распределении.

Построение модели эмпирического распределения, т.е. сглаживание его тем или иным теоретическим распределением, реализуется в меню *Statistics*/*Distribution Fitting* (рис.3.52.).

В качестве теоретической модели, возможно, использовать разные типы распределений. В окне процедуры они объединены в две группы:

- *Continuous Distributions* – непрерывные распределения;
- *Discrete* – дискретные распределения (рис. 3.53.)

Выбор конкретного типа модельного распределения осуществляется либо на основе содержательного анализа ранее полученных статистических характеристик вариационного ряда, либо исходя из самых общих соображений, опирающихся на визуальный анализ построенных графиков распределения. В практическом анализе обяза-

тельной является проверка соответствия изучаемого распределения нормальному закону распределения. Необходимость этого связана с тем, что условием применения значительного числа статистических характеристик и оценок является наличие нормального распределения.

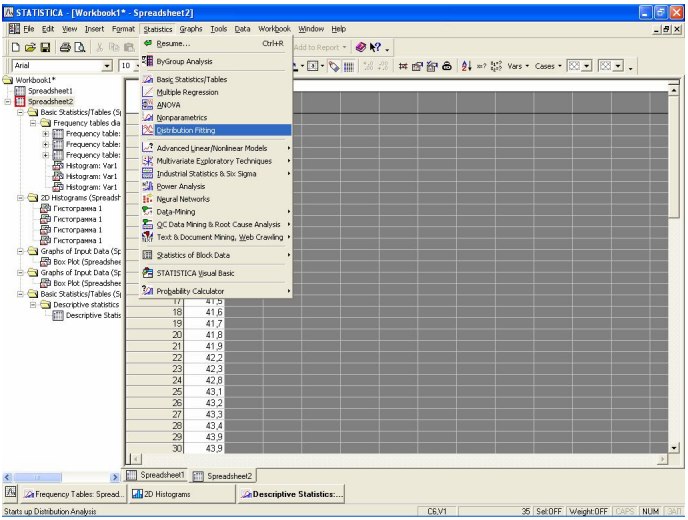


Рис. 3.52. Окно запуска процедуры сглаживания.

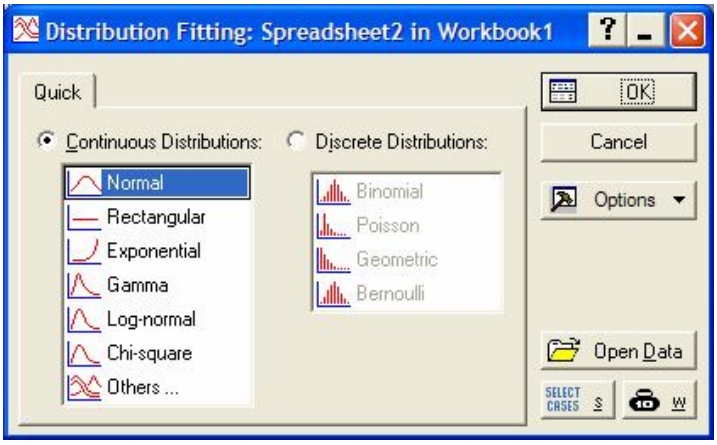


Рис. 3.53. Окно выбора теоретической модели меню *Distribution Fitting*.

После выбора теоретической модели распределения пользователю предлагается меню для ввода параметров сглаживания (рис. 3.54). Сначала вводится переменная (кнопка *Variable* = Var1), в подменю *Distribution* можно выбрать закон распределения. На закладке *Quick* находятся кнопки *Summary: Observed and expected distribution* (пересчет наблюдаемого и ожидаемого распределения) (кнопка *Summary* в правом углу меню аналогична) и *Plot of observed and expected distribution* (графическое представление наблюдаемого и ожидаемого распределения). Первая означает вывод расчетной таблицы теоретических и эмпирических частот с расчетом выбранных критериев согласия. Вторая кнопка обеспечивает вывод гистограммы эмпирического распределения с наложением на нее кривой теоретического распределения.

С помощью закладки *Parameters* устанавливаются показатели распределения (см. рис. 3.51.):

1. *Number of categories* – число интервалов. (В это поле вводится число интервалов из таблицы частот, признанной нами наиболее подходящей, то есть $k = 8$).
2. *Lower limit* – нижняя граница эмпирического распределения. (Из таблицы частот с числом интервалов $k = 8$ вводится нижняя граница первого интервала).
3. *Upper limit* – верхняя граница эмпирического распределения. (Из таблицы частот с числом интервалов $k = 8$ вводится верхняя граница последнего интервала).
4. *Mean* – среднее значение признака в совокупности.
5. *Variance* – дисперсия изучаемого признака.

Далее переходим к закладке *Options* (рис. 3.56.). Здесь можно задать расчет критериев согласия Колмогорова-Смирнова (*Kolmogorov-Smirnov test*) и критерия Пирсона (*Chi-square test/Combine categories*).

Критерий Колмогорова-Смирнова, строго говоря, не может быть использован в решаемой задаче (и во всех задачах подобного типа), так как в данном случае параметры модельного распределения определены по выборочным данным, что запрещено авторами критерия. Модельное распределение должно быть изначально полностью определено[4] [4, стр. 301-302]. Упомянутое ограничение, безусловно,

[4] Вадзинский Р., Статистические вычисления в среде Excel. Библиотека пользователя.- Спб.: Питер, 2008.- 608с.: ил. – (Серия «Библиотека пользователя»). Стр 308.
Вентцель Е.С., Теория вероятностей. - М.: Наука, 1969.-576с. Стр. 157-158.

имеет принципиальное значение, но оно полностью игнорируется в литературе по общей теории статистики.

Для проверки статистической гипотезы о законе распределения будем использовать критерий χ^2 - критерий Пирсона (*Chi-square test*) [8].

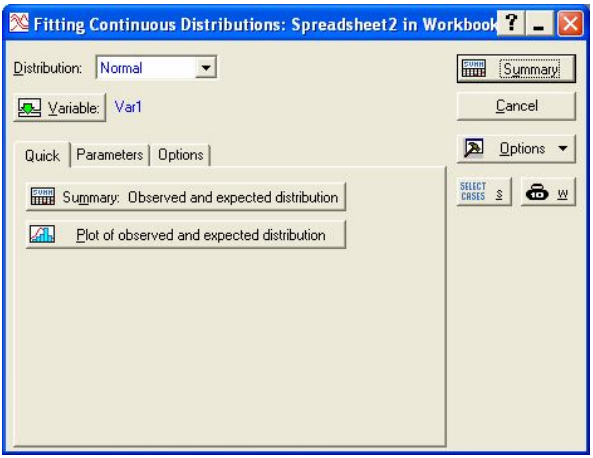


Рис. 3.54. Закладка *Quick* в меню *Distribution Fitting*.

Для решения задачи необходимо поставить метку на поле *Chi-square test/Combine categories*. Замечание в поле критерия (*if expected bin-frequency is less than or equal to 5, then combine with adjacent bins*) говорит о том, что автоматически реализуется известное ограничение на значения теоретических частот при вычислении критерия Пирсона, а именно $-f_i^{'} >= 5$ (некоторые авторы устанавливают нижнюю границу частот равной 7 или 10). Если частота эмпирического распределения, в каком либо интервале меньше, либо равна 5, то этот интервал объединяется с соседним. Следует иметь в виду, что упомянутое ограничение реализуется в программе не в явном виде. В выводимом на экран (на печать) протоколе результатов объединение интервалов не отображается.

В поле ввода *Graph* задаются опции для вывода графика (абсолютные, относительные, или накопленные частоты) (см. рис. 3.56.).

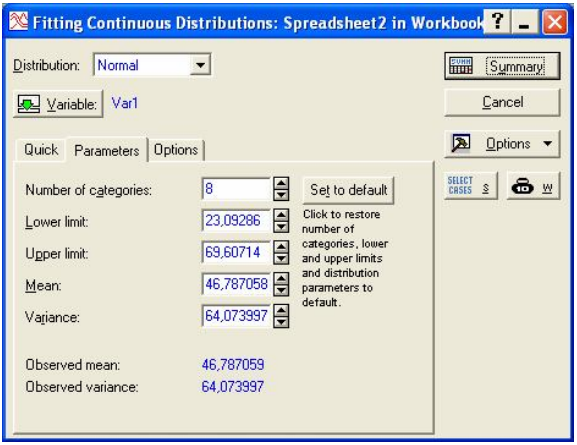


Рис. 3.55. Закладка *Parameters* в меню *Distribution Fitting*.

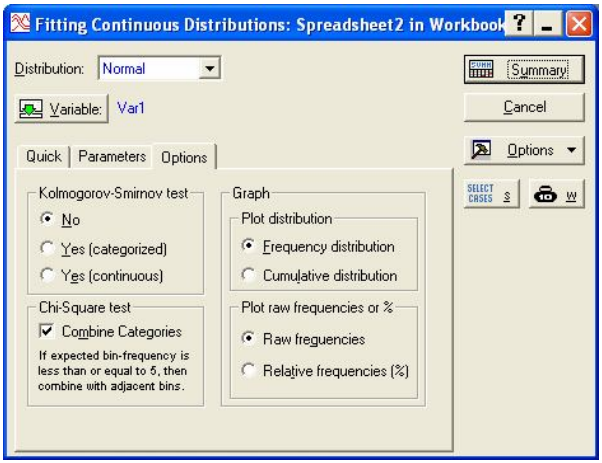


Рис. 3.56. Закладка Options в меню Distribution Fitting.

Chi-square test — позволяет решать задачу проверки гипотезы о законе распределения, результат оценки представляется в табличном виде (рис. 3.57., 3.58., 3.59.). Расчет критерия производится по следующей формуле:

$$\chi_0^2 = \sum_{i=1}^{k} \frac{(f_i - f_i^{'})^2}{f_i^{'}} , \qquad (3.24.)$$

где f_i – эмпирические абсолютные частоты (*Observed Frequency*);

$f_i^{'}$ – абсолютные частоты теоретического распределения (*Expected Frequency*);

k – число интервалов.

Variable: Var1, Distribution: Normal (Spreadsheet2 in Workbook1)
Chi-Square = 3,19542, df = 3 (adjusted) , p = 0,36247

Upper Boundary	Observed Frequency	Cumulative Observed	Percent Observed	Cumul. % Observed	Expected Frequency	Cumulative Expected	Percent Expected	Cumul. % Expected	Observed-Expected
<= 28,90715	2	2	2,35294	2,3529	1,08387	1,08387	1,27515	1,2751	0,91613
34,72143	3	5	3,52941	5,8824	4,51446	5,59833	5,31113	6,5863	-1,51446
40,53572	9	14	10,58824	16,4706	12,88161	18,47994	15,15483	21,7411	-3,88161
46,35000	27	41	31,76471	48,2353	22,16946	40,64940	26,08172	47,8228	4,83054
52,16429	26	67	30,58824	78,8235	23,02693	63,67634	27,09051	74,9133	2,97307
57,97857	12	79	14,11765	92,9412	14,43553	78,11186	16,98298	91,8963	-2,43553
63,79286	4	83	4,70588	97,6471	5,45892	83,57078	6,42226	98,3186	-1,45892
< Infinity	2	85	2,35294	100,0000	1,42922	85,00000	1,68143	100,0000	0,57078

Рис. 3.57. Проверка гипотезы о нормальном распределении переменной Var1.

Variable: Var1, Distribution: Log-normal (Spreadsheet2 in Workbook1)
Chi-Square = 4,34274, df = 2 (adjusted) , p = 0,11402

Upper Boundary	Observed Frequency	Cumulative Observed	Percent Observed	Cumul. % Observed	Expected Frequency	Cumulative Expected	Percent Expected	Cumul. % Expected	Observed-Expected
<= 28,90715	2	2	2,35294	2,3529	0,37507	0,37507	0,44126	0,4413	1,62493
34,72143	3	5	3,52941	5,8824	4,38126	4,75633	5,15442	5,5957	-1,38126
40,53572	9	14	10,58824	16,4706	15,27678	20,03311	17,97269	23,5684	-6,27678
46,35000	27	41	31,76471	48,2353	23,56555	43,59866	27,72417	51,2925	3,43445
52,16429	26	67	30,58824	78,8235	20,72815	64,32681	24,38606	75,6786	5,27185
57,97857	12	79	14,11765	92,9412	12,28966	76,61647	14,45843	90,1370	-0,28966
63,79286	4	83	4,70588	97,6471	5,50014	82,11661	6,47075	96,6078	-1,50014
< Infinity	2	85	2,35294	100,0000	2,88339	85,00000	3,39222	100,0000	-0,88339

Рис. 3.58. Проверка гипотезы о логарифмически нормальном распределении переменной Var1.

Variable: Var1, Distribution: Rectangular (Spreadsheet2 in Workbook1)
Chi-Square = 52,61178, df = 5, p = 0,00000

Upper Boundary	Observed Frequency	Cumulative Observed	Percent Observed	Cumul. % Observed	Expected Frequency	Cumulative Expected	Percent Expected	Cumul. % Expected	Observed-Expected
<= 28,90715	2	2	2,35294	2,3529	6,07143	6,07143	7,14286	7,1429	-4,07143
34,72143	3	5	3,52941	5,8824	12,14286	18,21429	14,28571	21,4286	-9,14286
40,53572	9	14	10,58824	16,4706	12,14286	30,35714	14,28571	35,7143	-3,14286
46,35000	27	41	31,76471	48,2353	12,14286	42,50000	14,28571	50,0000	14,85714
52,16429	26	67	30,58824	78,8235	12,14286	54,64286	14,28571	64,2857	13,85714
57,97857	12	79	14,11765	92,9412	12,14286	66,78571	14,28571	78,5714	-0,14286
63,79286	4	83	4,70588	97,6471	12,14286	78,92857	14,28571	92,8571	-8,14286
< Infinity	2	85	2,35294	100,0000	6,07143	85,00000	7,14286	100,0000	-4,07143

Рис. 3.59. Проверка гипотезы о прямоугольном распределении переменной Var1.

Проиллюстрируем работу *Distribution Fitting*, сгладив эмпирическое распределение переменной Var1 нормальным распределениям. Формулы, по которым рассчитывается плотность модельного распределения, а также формулы для расчета теоретических частот распределения могут быть легко найдены в общедоступной справочной и учебной литературе [7, 8]. В данном пособии приведены формулы для нормального распределения.

Функция нормального распределения:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\bar{x})^2}{2\sigma^2}}, x \in (-\infty; +\infty). \qquad (3.25.)$$

Плотность нормального распределения (Приложение 4):

$$\varphi(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}, \qquad (3.26.)$$

где x – значение изучаемого признака;

$\bar{x}$ – средняя арифметическая величина;

σ – среднее квадратическое отклонение изучаемого признака;

e, π – математические константы;

$t = \frac{x - \bar{x}}{\sigma}$ – нормированное отклонение.

Теоретические частоты нормального распределения рассчитываются по следующей формуле:

$$f_i' = \frac{Nh}{\sigma}\varphi(t_i), \qquad (3.27.)$$

где N – объем совокупности;

h - величина интервала.

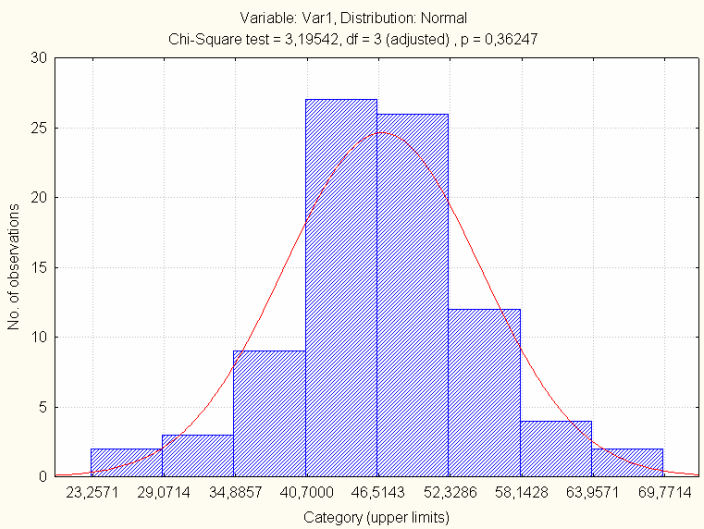


Рис. 3.60. Гистограмма и расчетная кривая нормального распределения для переменной Var1.

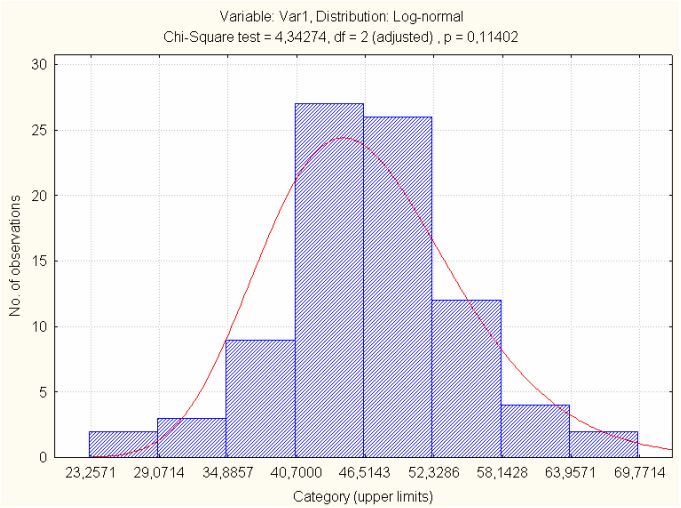


Рис. 3.61. Гистограмма и расчетная кривая логарифмически нормального распределения для переменной Var1.

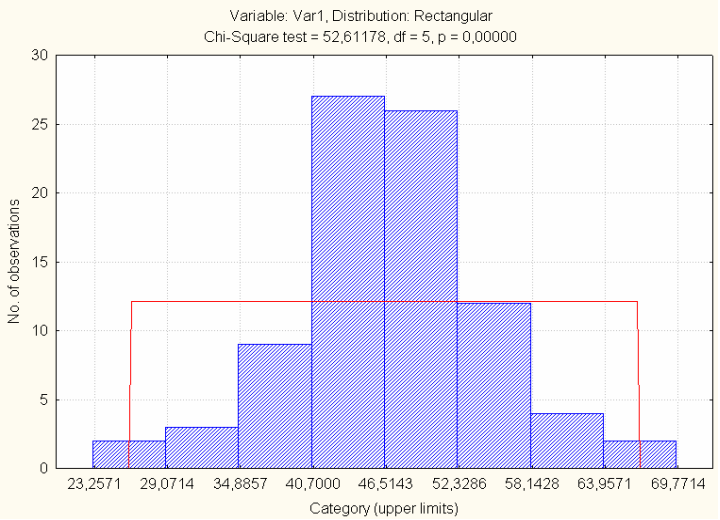


Рис. 3.62. Гистограмма и расчетная кривая прямоугольного распределения для переменной Var1.

В шапке таблиц и графиков находятся следующие показатели:
Chi-square test – расчетное значение критерия Пирсона,
d.f. (adjusted) – уточненное значение числа степеней свободы:

$$d.f.(r) = k - l - 1, \qquad (3.28.)$$

где k – число интервалов эмпирического распределения (вариационного ряда);
l – число параметров теоретического распределения, определяемых по опытным данным (например, в случае нормального закона распределения число оцениваемых по выборке параметров $l = 2$, математическое ожидание и дисперсия)
p – расчетный уровень значимости.

Таблица 3.3

Результаты решения задачи сглаживания

Тип распределения	Число степеней свободы r	Расчетное значение критерия χ_0^2	Табличное значение критерия $\chi^2{}_{\alpha,r}$	$P(\chi^2_{\alpha,r} \geq \chi^2_0)$ (расчетное значение уровня значимости)
нормальное	3	3,19542	7,815	0,36247
логнормальное	2	4,34274	5,991	0,11402
прямоугольное	5	52,61178	11,07	0,00000

Принятие решения о справедливости гипотезы о законе распределения можно осуществить, ориентируясь на эмпирическое значение критерия χ_0^2, либо на расчетное значение вероятности (расчетный уровень значимости) $P(\chi^2_{\alpha,r} \geq \chi^2_0)$. Первое сравнивается с табличным значением $\chi^2{}_{\alpha,r}$, второе с принятым уровнем значимости (примем $\alpha = 0{,}05$). Окончательные выводы по проверке гипотез о законе распределения:

1. Так как

$\chi_0^2 = 3{,}19542 < \chi^2_{0.05;3} = 7{,}815$ и

$P(\chi^2{}_{0{,}05,3} \geq \chi^2_0) = 0{,}36247 > \alpha = 0{,}05,$

то гипотеза о нормальном распределении переменной Var1 не противоречит статистическим данным.

2. Так как

$\chi_0^2 = 4{,}34274 < \chi^2_{0.05;2} = 5{,}991$ и

$P(\chi^2{}_{0{,}05,2} \geq \chi^2_0) = 0{,}11402 > \alpha = 0{,}05,$

то гипотеза о логнормальном распределении переменной Var1 также не противоречит статистическим данным.

3. Так как

$\chi_0^2 = 44{,}03609 > \chi^2_{0.05;5} = 11{,}07$ и

$P(\chi^2{}_{0{,}05,5} \geq \chi^2_0) = 0{,}00000 < \alpha = 0{,}05,$

то гипотеза о прямоугольном распределении переменной Var1 отвергается на 0,00000 уровне значимости. Очевидно, что равномерное распределение взято в качестве модели исключительно в иллюстративных

целях для «контрастного» проявления механизма работы критерия Пирсона.

При решении задачи сглаживания вручную автору понадобиться заполнить следующие таблицы.

Таблица 3.4

Таблица для расчета критерия χ_0^2 вручную

№	$x_i^{\min}$	$x_i^{\max}$	x_i'	t_i	$\varphi_i(t)$	f_i	χ_{i0}^2

Таблица 3.5

Проверка гипотезы о нормальном законе распределения вручную

Тип распределения	Число степеней свободы r	Расчетное значение критерия χ_0^2	Табличное значение критерия $\chi^2_{\alpha,r}$
нормальное			

Для графического изображения сглаживания целесообразно воспользоваться полем на рис. 3.63.

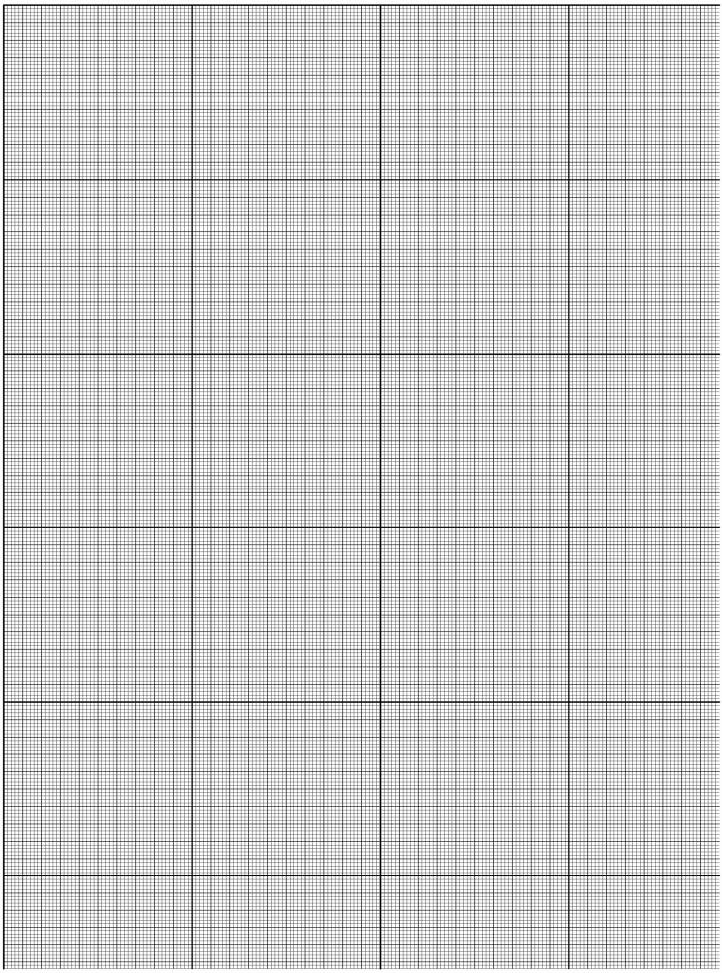


Рис. 3.63. Поле масштабно-координатной бумаги.

4. Выборочное наблюдение

Любую статистическую акцию, связанную с выборочным наблюдением, можно в общем смысле разделить на два этапа:

- собственно проведение выборочного статистического наблюдения как этапа получения данных;
- обработка, анализ и интерпретация полученной информации, формулировка окончательных выводов.

ППП STATISTICA может быть эффективно применен как на первом этапе планирования и проведения выборочного наблюдения, так и на втором этапе количественной обработки его результатов.

Для наглядной иллюстрации проявления эффекта случайной ошибки репрезентативности при выполнении лабораторной работы рекомендуется сформировать несколько независимых малых выборок равного объема (соблюдается принцип – «при прочих равных условиях») и одну выборку большого объема. Конкретные значения объема выборок определяются в индивидуальных заданиях.

4.1. Определение объема выборки. Формирование выборочной совокупности

Важным вопросом подготовки выборочного наблюдения является определение объема выборочной совокупности, необходимой и достаточной для оценки тех или иных свойств генеральной совокупности. В практике экономико-статистических исследований, как правило, используется процедура бесповторного отбора единиц в выборочную совокупность. Первым этапом подготовки выборочного наблюдения является расчет объема выборки. Расчет, как правило, проводится по следующей формуле:

$$\mathrm{n} = \frac{t^2 \times \sigma^2 \times N}{\Delta^2 \times N + t^2 \times \sigma^2}\ , \qquad (4.1)$$

где N – объем генеральной совокупности;

t – параметр нормального распределения; находится по таблицам интегральной функции нормального распределения в соответствии с заданным уровнем доверительной вероятности;

σ – среднее квадратическое отклонение в генеральной совокупности; его величина берется на основе прошлого опыта или пилотажного исследования, при отсутствии таковых, как 1/6 (1/5) размаха вариации; Студентам рекомендуем взять значение σ из первой лабораторной работы;

Δ – предельная ошибка выборки; устанавливает точность результатов выборочного наблюдения.

В реальных условиях значение предельной ошибки выборки устанавливается экспертным путем, исходя из требований к точности результатов выборочного наблюдения. При определении величины предельной ошибки следует учитывать то, что уменьшение величины ошибки на порядок ведет к увеличению объема выборки на два порядка. В практических исследованиях, как правило, расчет объема выборки проводят многократно, с учетом разных значений ошибки.

В выполняемой лабораторной работе предельная ошибка выборки принимается равной определенной доле генеральной средней (задается в индивидуальном задании в процентах).

ВНИМАНИЕ!!! При выполнении 2-ой лабораторной работы на основе данных 1-ой работы следует использовать не ранжированную совокупность.

Первоначальный рабочий лист делаем активным, для этого щелкаем по нему правой кнопкой мыши и выбираем функцию *Use as Active Input* (рис. 4.1.). Теперь этот рабочий лист подсвечивается рамкой красного цвета, и все процедуры выполняются с его данными.

Для формирования случайной бесповторной выборки необходимо воспользоваться в меню *Data* (данные) процедурой *Subset/Random Sampling* (подмножество/случайный выбор). Для запуска процедуры требуется обязательно находиться на рабочем листе с исходными данными и этот рабочий лист должен быть активным (рис. 4.2., 4.3.).

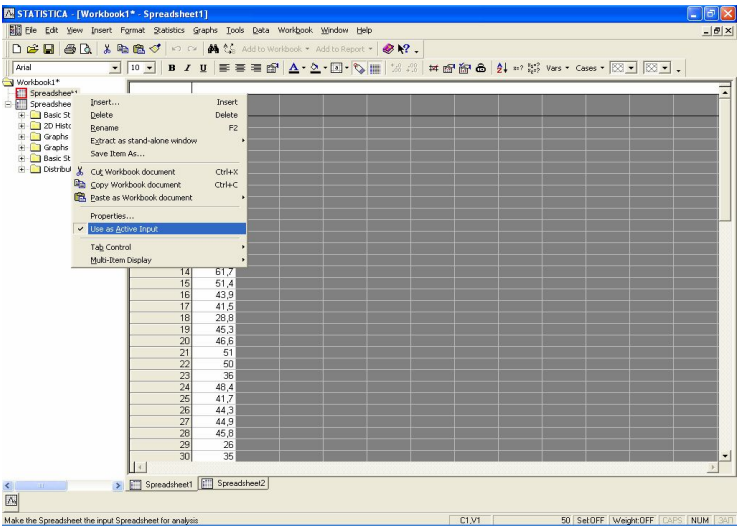


Рис. 4.1. Функция задания активного рабочего листа.

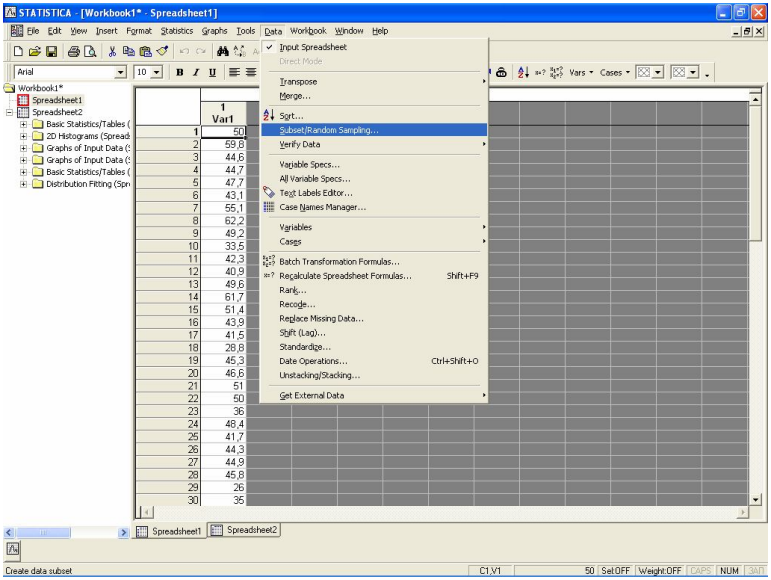


Рис. 4.2. Окно выбора процедуры выборки.

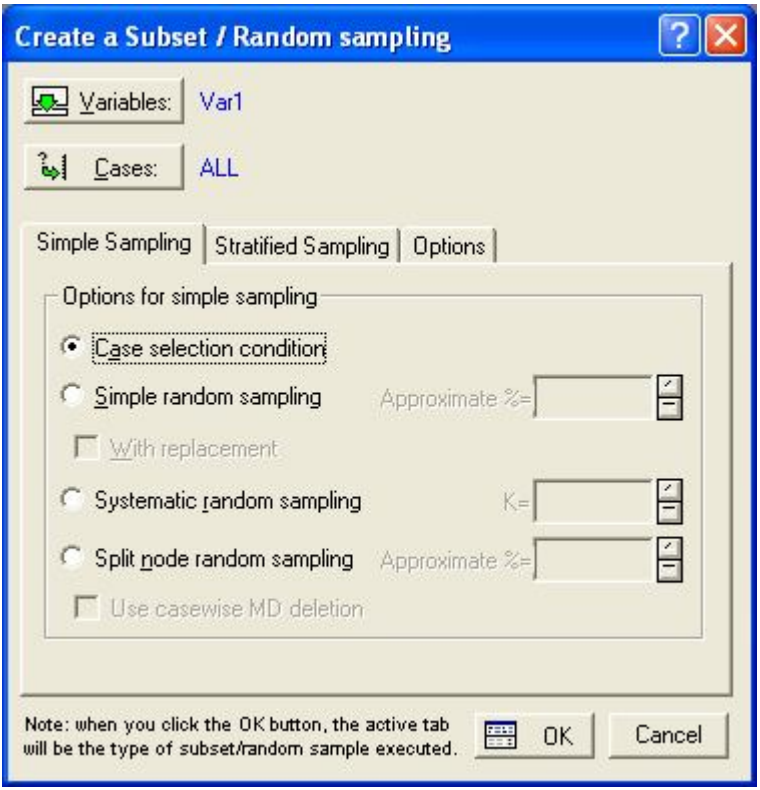


Рис. 4.3. Вид закладки *Simple Sampling* меню создания выборки.

В появившемся окне (рис. 3.4.) выбираем переменную (кнопка *Variables* = Var1), в этом случае исходные данные выступают в качестве генеральной совокупности. Кнопка *Cases* используется тогда, когда необходимо организовать не простой случайный отбор элементов, а задать определенные условия отбора, например, отобрать единицы под конкретными номерами или со значениями больше/меньше определенного. Эта кнопка работает, если на закладке *Simple Sampling* (простой отбор) в подменю *Options for simple sampling* (опции построения простой случайной выборки) стоит метка в поле *Case selection condition* (использовать условия выбора ячеек). Если же метка стоит в дру-

гом поле, то кнопку *Cases* нажать нельзя. Также на этой закладке находятся следующие поля.

Поле *Systematic random sampling* (систематический случайный выбор). Эта процедура позволяет реализовать выборочное наблюдение по схеме механической выборки. При выборе этой процедуры задается число интервалов (k =) и программа выбирает элементы через равный интервал.

Поле *Split node random sampling* позволяет получить из генеральной совокупности две выборки, при этом примерный процент или размер одной из выборок задается в поле *Approximate N (%)*.

Здесь необходимо отметить, что система предлагает два варианта создания выборок. В первом случае задается доля элементов (в процентах), которые необходимо выбрать из генеральной совокупности (этот вариант выбран по умолчанию). Во втором случае задается число элементов (собственно объем выборки), которое необходимо выбрать из совокупности (выбирается с помощью закладки *Options*).

В соответствии с программой расчетной работы нам необходимо создать простые случайные выборки, для чего используем поле *Simple random sampling.*

При этом очевидно, что в соответствии с заданием необходимо выбрать точное число элементов, то есть создать выборку строго определенного объема.

Поэтому сначала необходимо воспользоваться закладкой Options (рис. 4.4.). В поле *Copy formatting to new Spreadsheet cases* (копировать формат наблюдений в новую таблицу) метку ставить не надо. Поле *Use case selection condition* соответствует одноименному на закладке *Simple Sampling*. Поле *Use Diehard-certified random number generator (note: this algorithm is slower)* означает возможность использования генератора случайных чисел, в скобках предупреждение о том, что эта процедура более медленная.

Поля *Calculate based on percentage of cases* и *Calculate based on approximate N* предлагают соответственно задать формат выбора либо в процентах от числа ячеек, либо в виде объема выборки. Нас интересует второй вариант.

Также необходимо поставить метку в поле *Use spreadsheet case weights as case multipliers in random sampling* – это позволит получать меньше повторов в выборке (см. рис. 4.4.).

Закладка *Stratified Sampling* используется для построения стратифицированной выборки.

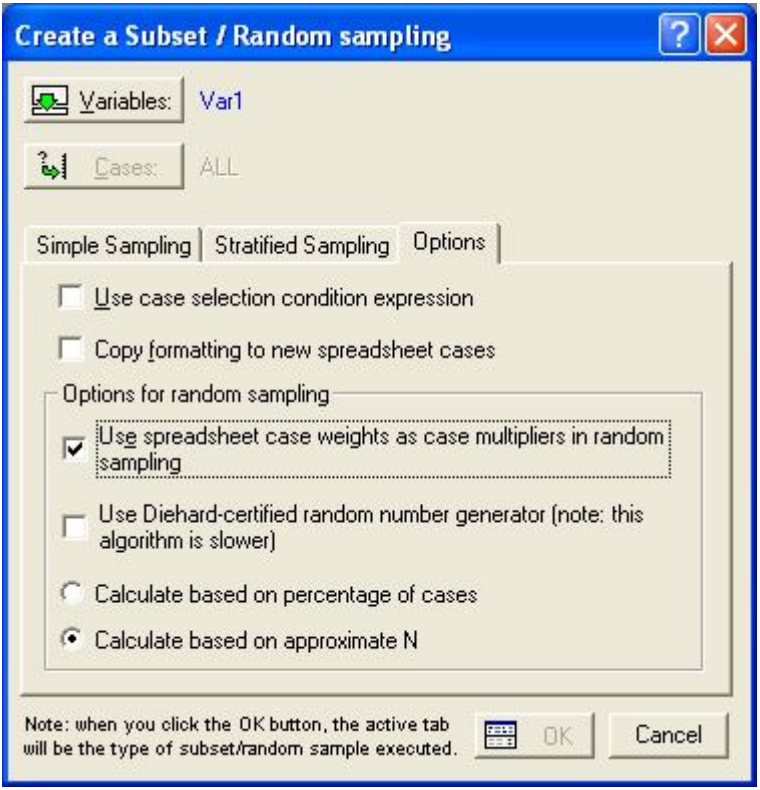


Рис. 4.4. Вид закладки *Options* меню создания выборки.

Далее на закладке *Simple Sampling* ставим метку в поле *Simple random sampling* и вводим необходимый объем выборки в поле *Approximate N* (рис. 4.5.)

Но STATISTICA предлагает задать примерное число (анг. Approximate) элементов и формирует выборки различного объема, но не превышающего заданного пользователем. Дело в том, что в STATISTICA задан определенный алгоритм бесповторной выборки: программа системы отбирает определенное количество элементов, ориентируясь на заданный исследователем объем выборки. Однако система сканирует совокупность только один раз. И если необходимое число элементов не найдено, то система создает выборку только из выбранных элементов.

Для устранения этого необходимо поставить метку в поле *With replacement*, что означает «выбор с возвращением». Это процедура означает, что программа обследует совокупность, выбирает определенное число элементов, и, если объем выборки не совпадет с заданным пользователем в поле *Approximate N*, то программа обследует совокупность заново, чтобы добрать необходимое число элементов.

Далее нажимаем кнопку *OK*.

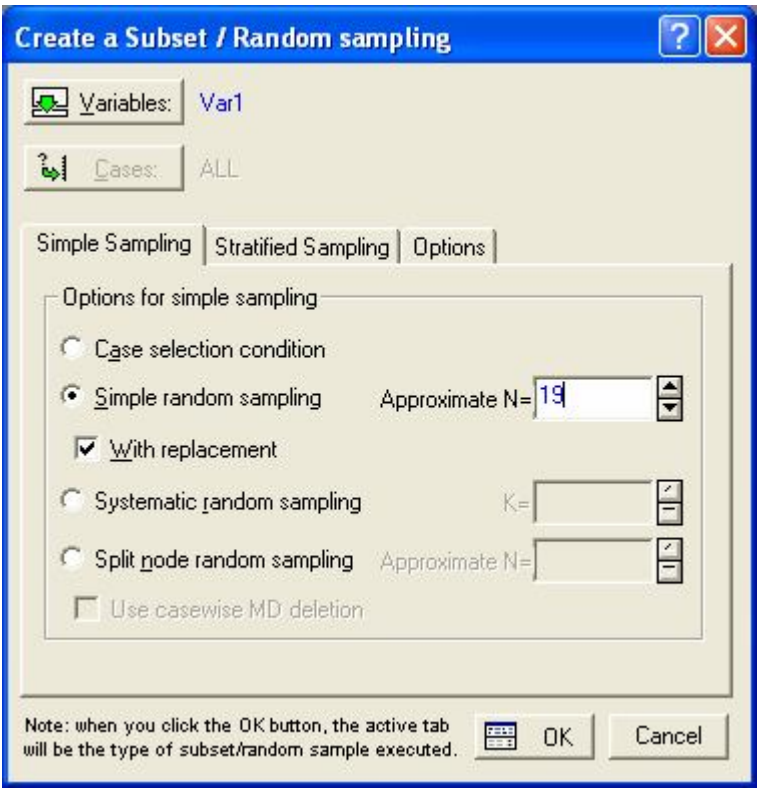


Рис. 4.5. Вид закладки *Simple Sampling* меню создания выборки с заданными условиями построения простой случайной выборки.

Необходимо особо отметить, что в столбце с исходными данными не должно быть пустых ячеек (пропущенных данных), иначе система может выбрать их как элементы формируемой выборки.

Появляется окно с новой электронной таблицей, в которой отображается сформированная выборочная совокупность (рис.4.6.).

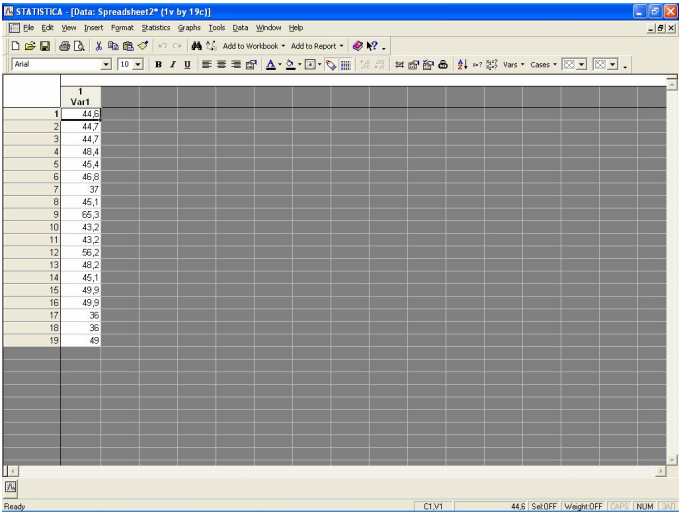


Рис. 4.6. Окно с новой рабочей таблицей и сформированной выборкой.

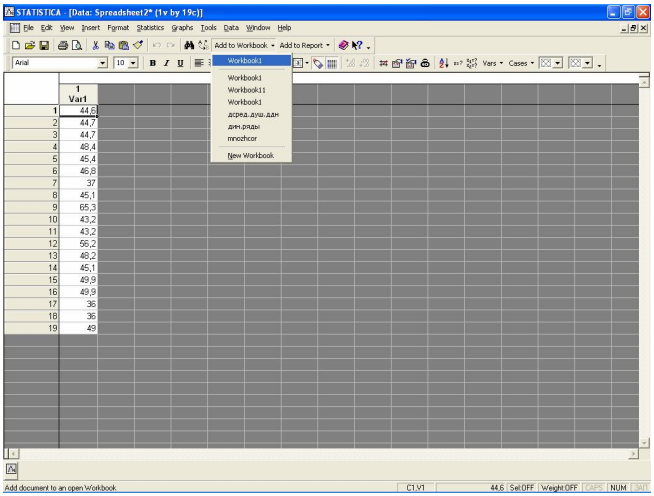


Рис. 4.7. Окно с выбранной функцией добавления в рабочую книгу.

Полученную выборку следует перенести в рабочую книгу, для чего воспользуемся кнопкой *Add to Workbook* и выберем нужную рабочую книгу Workbook1 (рис. 4.7.).

Выборка помещается в следующий по счету рабочий лист и отображается в дереве рабочей книги в левой стороне экрана. При этом данный рабочий лист становится активным (помечается красным цветом) другими словами, любая запущенная вновь процедура будет выполняться на базе данных, размещенных в нем.

Для создания необходимого числа выборок, предусмотренных программой расчетной работы, необходимо активизировать рабочий лист с исходными данными (генеральной совокупностью) (см. рис. 4.1.). Помимо этого для создания выборок нужно находиться на рабочем листе с исходными данными. Эту процедуру необходимо проделывать после создания каждой выборки.

Процедура формирования выборки повторяется необходимое число раз (в нашем примере - 6) и в итоге мы имеем требуемое число независимых случайных выборок. Во избежание путаницы следует переименовать рабочие листы. Таким образом, первый этап выборочного статистического наблюдения завершен: мы имеем несколько выборочных совокупностей для последующего анализа.

Для удобства проведения анализа рекомендуется разместить выборки на одном рабочем листе. С этой целью выбирается лист с генеральной совокупностью (исходными данными), затем создаем там необходимое число новых переменных (в нашем примере 6). После этого копируем данные из отдельных листов с выборками в новые столбцы нашего рабочего листа. Процедуры создания новых переменных и копирования данных описаны в разделе 2.4.

Результатом должен стать рабочий лист с пятью малыми выборками, одной выборкой большого объема и исходными данными (напоминаем, что система не учитывает незаполненные ячейки) (рис. 4.8.). Именно этот рабочий лист следует сделать активным на данном этапе. Необходимо напомнить, что все запускаемые процедуры используют данные активного рабочего листа. Однако если анализ уже запущен, то возможно выбрать другой активный лист и начать другую процедуру (или ту же). При этом анализ, запущенный ранее будет все равно работать с данными листа, который был активен при его запуске.

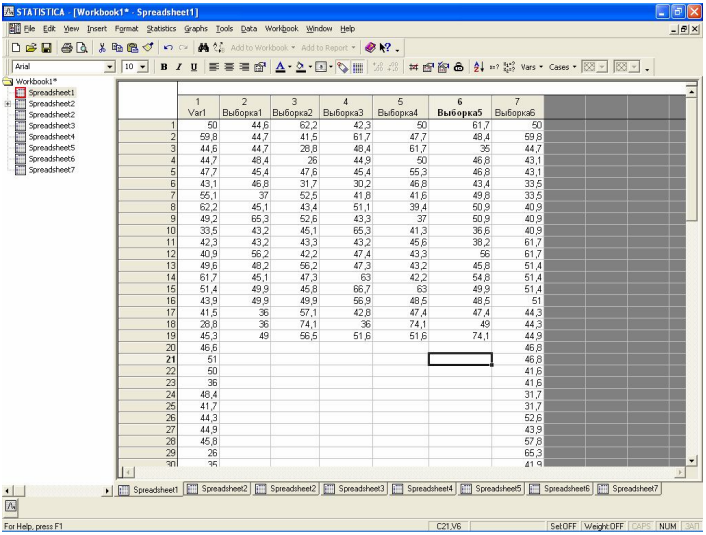


Рис. 4.8. Рабочий лист с исходными данными и выборками.

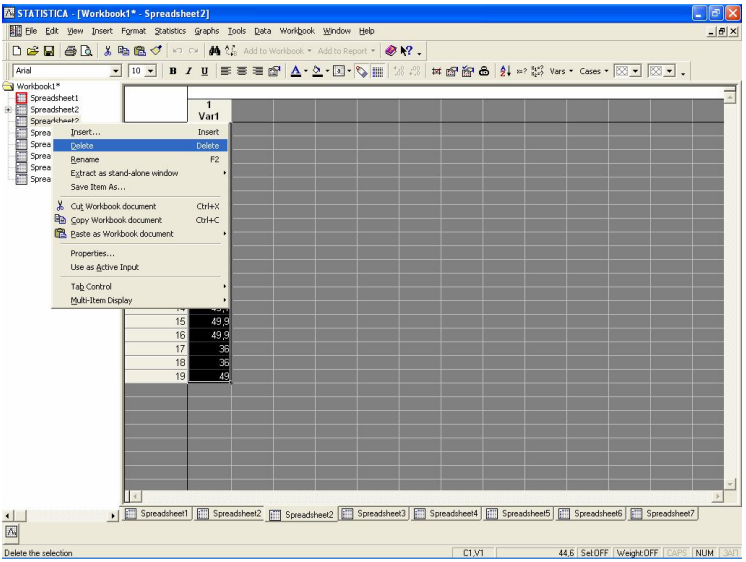


Рис. 4.9. Функция удаления элементов рабочей книги.

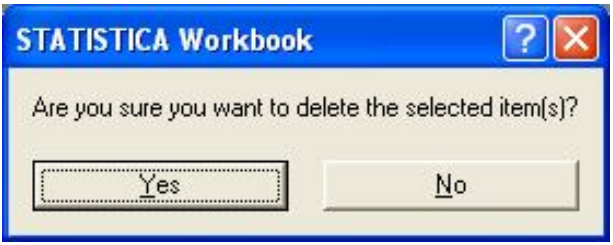


Рис. 4.10. Подтверждение удаления элемента.

Ненужные рабочие листы рекомендуется удалить, так как теперь выборки сосредоточены в одном месте. С этой целью в дереве рабочей книги достаточно щелкнуть на нужном объекте правой кнопкой мыши и выбрать функцию *Delete*, подтвердив ее (рис. 4.9., 4.10.).

4.2. Статистическая обработка результатов выборочного наблюдения

Обычно обработка выборочных данных предполагает расчет основных статистических характеристик выборки, величины ошибки выборки и, затем, вероятностную оценку параметров генеральной совокупности, и проверку гипотез о значениях этих параметров. В специализированных статистических программах эти расчеты объединены в одной процедуре.

Для обработки выборочных данных используем меню *Statistics* и в нем процедуру *Basic Statistics/Tables*. Далее выбираем *t-test, single sample*, что означает расчет t-критерия отдельно для каждой выборки (рис. 4.11.).

В меню выбора параметров расчета (рис.4.12.) воспользуемся кнопкой *Variables* для выбора переменной, в данном случае полезно выбрать сразу все переменные, соответствующие сформированным выборкам (рис. 4.13.).

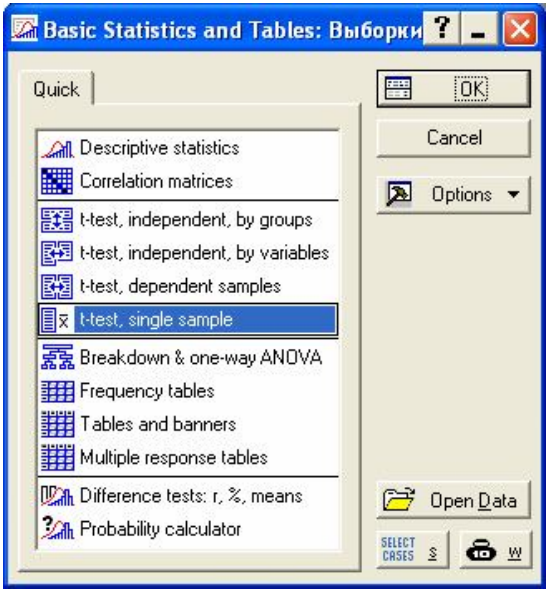


Рис. 4.11. Окно меню *Basic Statistics and Tables.*

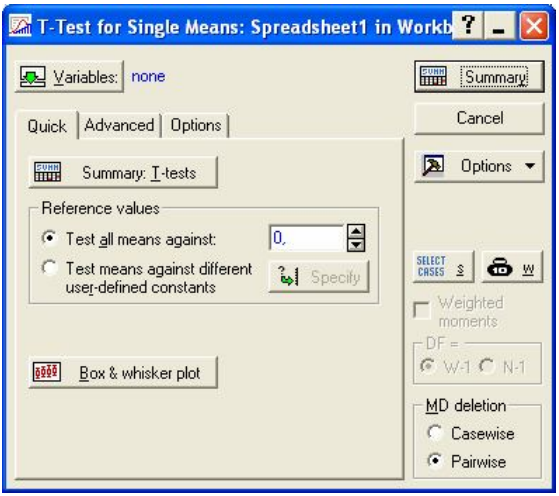


Рис. 4.12. Меню выбора параметров расчета t-критерия.

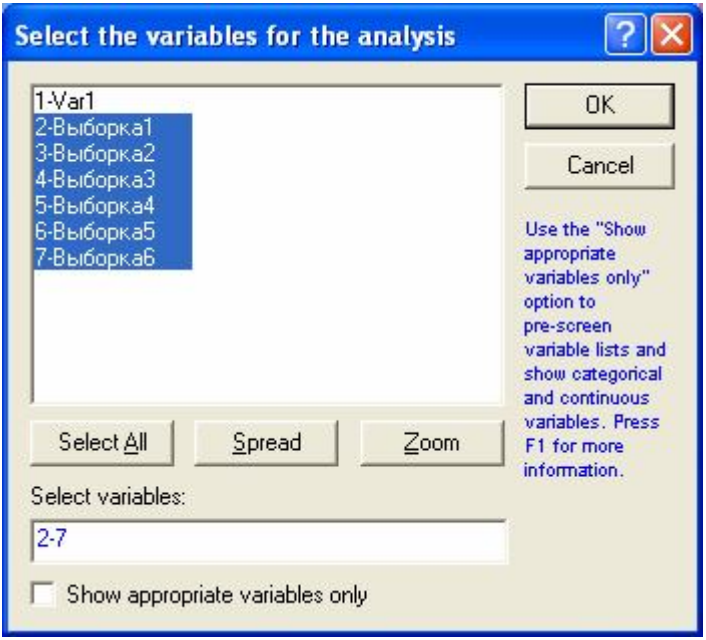


Рис. 4.13. Окно выбора переменных.

Кнопки *Summary: T-tests* и *Summary* запускают расчет *t*-критериев для проверки гипотезы о значении генеральной средней. В меню *Reference values* задается значения генеральной средней, относительно которой и проверяется гипотеза. При этом если метка стоит на поле *Test all means against*, проверка гипотезы о генеральной средней по данным всех выборок осуществляется с учетом одного и того же числа. Если же поставить метку на поле *Test means against different user-defined constant*, то с помощью кнопки *Specify*, можно задать проверяемое значение генеральной средней для каждой выборки отдельно. Нас интересует первый вариант. (Кнопка *Box & whisker plot* строит уже известное графическое изображение, но по выборочным данным.).

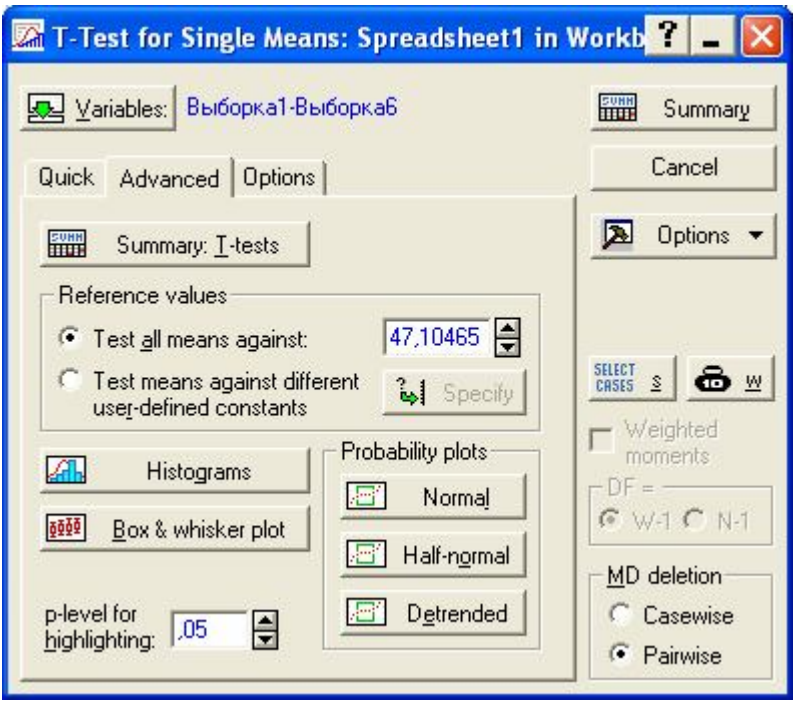


Рис. 4.14. Выбор дополнительных параметров расчета t-критерия.

Далее переходим к закладке *Advanced.* Пользователю предоставляется возможность построить гистограммы, диаграммы размаха, а так же вероятностные графики. В поле *p-level for highlighting (*заданный уровень значимости) устанавливается принятый при проверке гипотезы уровень значимости критерия. В нашем примере он равен 0,05 ($\alpha = 0.05$), т.к. доверительная вероятность равна 0.95 ($P = 0.95$) (рис. 4.14.).

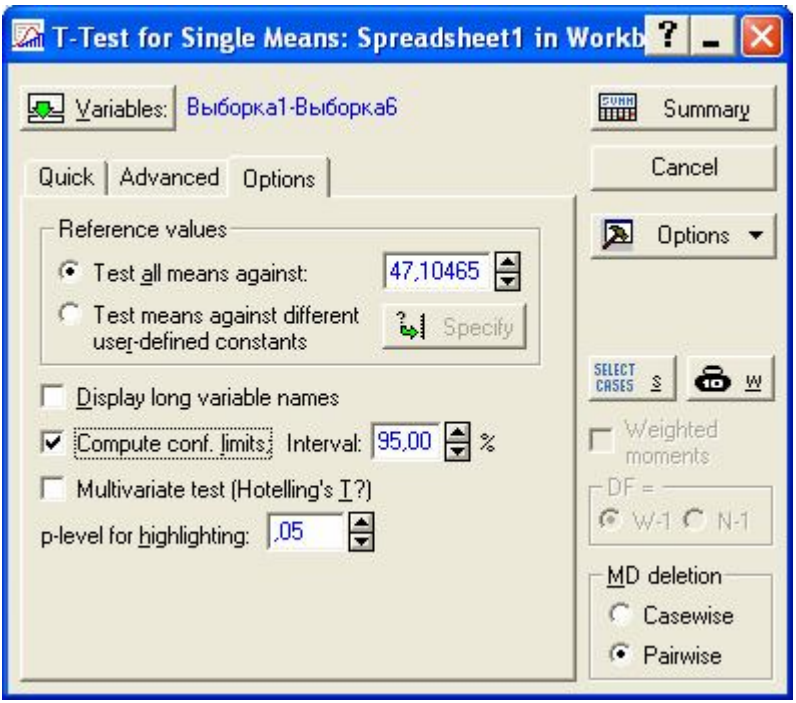


Рис. 4.15. Выбор дополнительных параметров расчета t-критерия.

На закладке *Options* ставим метку в поле *Compute conf.limits; Interval* для расчета доверительных интервалов с заданной вероятностью (95%). Поле *Display long variable names* позволяет в отчете воспроизводить длинные имена переменных (рис. 4.15.). Нажимаем кнопку *Summary*. Результаты анализа выводятся на экран в виде следующей таблицы:

Test of means against reference constant (value) (Spreadsheet1 in Workbook1)

Variable	Mean	Std.Dv.	N	Std.Err.	Confidence -95,000%	Confidence +95,000%	Reference Constant	t-value	df	p
Выборка1	46,24737	6,79063	19	1,557877	42,97439	49,52035	47,10465	-0,550288	18	0,588886
Выборка2	47,56842	11,54220	19	2,647961	42,00526	53,13158	47,10465	0,175143	18	0,862923
Выборка3	48,91053	9,91357	19	2,274328	44,13234	53,68871	47,10465	0,794026	18	0,437524
Выборка4	48,93158	9,18677	19	2,107589	44,50370	53,35946	47,10465	0,866833	18	0,397442
Выборка5	49,15789	8,82857	19	2,025413	44,90266	53,41313	47,10465	1,013741	18	0,324140
Выборка6	48,34651	10,01942	43	1,527947	45,26299	51,43003	47,10465	0,812765	42	0,420936

Рис. 4.16. Результаты обработки выборочных данных.

В первом столбце (*Variable*) представлены имена переменных (выборок).

Mean – значения выборочных средних.

Std. Dev. – значения среднего квадратического (стандартного) отклонения.

N – объем выборки.

Std.Err. – средняя ошибка выборки.

Confidence -95,000% – нижняя граница доверительного интервала при вероятности 95%.

Confidence +95,000% – верхняя граница доверительного интервала при вероятности 95%.

Reference – гипотетическое значение генеральной средней (в нашем примере это значение известно из первой работы).

t-value – расчетное значение *t*-критерия для проверки гипотезы о значении генеральной средней.

df – число степеней свободы (определяется как n – 1).

p – расчетный уровень значимости *t*-критерия.

Таким образом, по данным каждой выборки рассчитаны: среднее значение анализируемого показателя, стандартное отклонение и величина средней ошибки выборки. Эти результаты позволяют, с учетом заданной доверительной вероятности (в примере 95%), определить границы доверительных интервалов для генеральной средней (графы: *Confidence -95,000%* и *Confidence +95,000%* на рис. 4.16.). Доверительный интервал для неизвестной генеральной средней определяется:

$$\widetilde{x} - \Delta \le \overline{x} \le \widetilde{x} + \Delta ,$$

где $\overline{x}$ — генеральная средняя;

$\widetilde{x}$ — выборочная средняя;

Δ — предельная ошибка выборки.

Предельная ошибка выборки вычисляется по формуле:

$$\Delta = t \times \mu , \qquad (4.2.)$$

где t – параметр нормального распределения (для малых выборок – распределения Стьюдента);

μ - средняя ошибка выборки, определяемая как:

$$\mu = \sqrt{\frac{\sigma^2}{n}}\,, \qquad (4.3.)$$

где n – объем выборки;

σ^2 - выборочная дисперсия.

Для выборки 1, например, рассчитанный доверительный интервал:

$$42{,}97439 \le \bar{x} \le 49{,}52035\,,$$

что означает: с вероятностью 95% можно утверждать, что в среднем по России показатель доли убыточных организаций в 2003г. находился в указанных пределах. Как видим, минимальная величина средней ошибки выборки соответствует выборке большего объема (n=43) и доверительный интервал по результатам этой выборки:

$$45{,}26299 \le \bar{x} \le 51{,}43003\,.$$

4.3. Проверка статистических гипотез о значении генеральной средней и о равенстве двух генеральных средних

Наряду с определением доверительного интервала для неизвестной генеральной средней пользователю предоставляется возможность проверки простой гипотезы о значении генеральной средней (понятия простой и сложной гипотез рассматриваются в курсе теории вероятностей). В демонстрационном примере проверяется гипотеза:

$$H_{0:}\ \text{Mean} = 47{,}10465,$$

где 47,10465 – точное значение генеральной средней (берется из первой расчетной работы). Гипотеза проверяется с помощью t-критерия, который рассчитывается по следующей формуле:

$$t_{расч} = \frac{|\widetilde{x} - \overline{x}|}{\mu}. \qquad (4.4.)$$

Отметим, что при анализе выборок системой, в числителе формулы t-критерия не учитывается знак модуля.

Расчетное значение критерия сравнивается с табличным и если соблюдается неравенство:

$$t_{расч} \le t_{табл},$$

то гипотеза о значении генеральной средней принимается. В ППП STATISTICA предусмотрена удобная сервисная функция: если испытуемая гипотеза отвергается, то результаты расчета t-критерия высвечиваются в таблице красным цветом.

Вывод о результатах проверки гипотезы можно сделать также через сопоставление расчетного уровня значимости P с принятым исследователем α (обычно задается α = 0.05). Гипотеза принимается при условии, что $P > \alpha$.

Рассмотрим оценку гипотезы на примере Выборки 1:

$$t_{расч} = \frac{|46{,}24737 - 47{,}10465|}{1{,}557877} = 0{,}550288,$$

Теоретическое значение критерия берется из таблиц распределения Стьюдента – $t_{табл} = 2{,}101$.

Таким образом, имеем:

$$t_{расч} = 0{,}550288 < t_{табл} = 2{,}101.$$

Вывод очевиден – гипотеза о значении генеральной средней не отвергается. Тот же результат получим, сравнивая расчетный уровень значимости (P) с установленным (α):

$$P = 0{,}58886 > \alpha = 0{,}05.$$

Для более полного уяснения механизма проверки гипотезы о значении генеральной средней рекомендуем обратиться к таблице на рис. 4.16 и провести проверку относительно значения генеральной средней, выходящего за границы доверительного интервала.

Далее следуют таблицы для самостоятельного заполнения по результатам выборочного наблюдения и для проверки гипотезы о значении генеральной средней.

Таблица 4.1

Результаты выборочного наблюдения

Выборка	Выборочная средняя	Нижняя граница доверительного интервала	Верхняя граница доверительного интервала	Число степеней свободы	Расчетный уровень значимости

Таблица 4.2

Проверка гипотезы о значении генеральной средней

Выборка	Расчетное значение *t*-критерия	Табличное значение *t*-критерия	Результат

Описанная процедура испытания статистических гипотез позволяет провести оценку существенности разности двух выборочных средних. Если разность между средними величинами статистически значима, это означает, что различие вызвано неслучайными факторами, или выборки не принадлежат одной генеральной совокупности. Иначе эта задача формулируется как проверка статистической гипотезы о равенстве двух средних:

$$\mathrm{H}_0 : \bar{x}_1 = \bar{x}_2 .$$

В литературе встречается тождественная формулировка гипотезы, а именно:

$$H_0 : \bar{x}_1 - \bar{x}_2 = 0 .$$

В нашем примере содержательно гипотеза формулируется следующим образом: взяты выборки из одной или из разных генеральных совокупностей? В контексте решаемой задачи ответ очевиден – выборки взяты из одной и той же совокупности. Но следует обратить особое внимание на проявление эффекта случайной ошибки репрезентативности. Реализация процедуры проверки гипотезы может дать, в редких случаях, парадоксальный результат, а именно, показать на основе *t*-критерия, что выборки как бы взяты из разных генеральных совокупностей с разными значениями средних величин. С дидактической точки зрения такой результат весьма полезен для понимания существа статистических выводов и степени их условности. Для демонстрации этого эффекта рекомендуется взять такие две выборки, из ранее полученных, для которых: $\tilde{x}_i - \tilde{x}_k = \text{Max}$.

В рассматриваемом примере - это выборки 1 и 5. Вычисления проводятся с помощью программы меню *Statistics/ Basic Statistics/Tables, процедура t-test, independent, by variables* (*t*-критерий для двух независимых выборок) (рис. 4.17.).

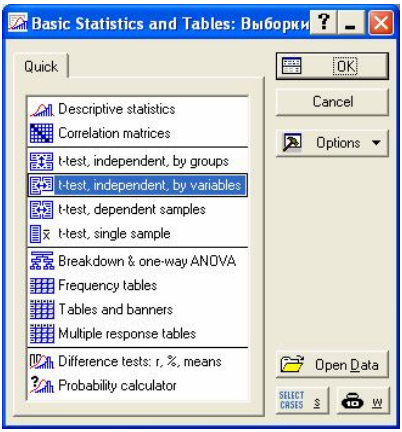


Рис. 4.17. Меню Basic Statistics and Tables.

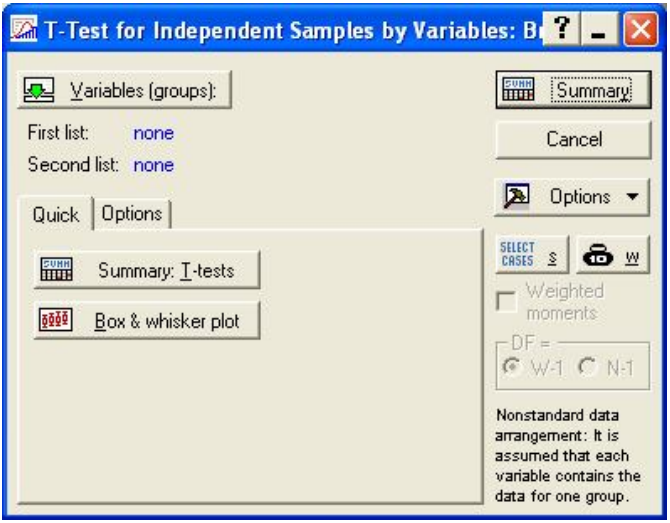


Рис. 4.18. Окно для расчета двухвыборочного t-критерия.

В отличие от предыдущей задачи необходимо одновременно ввести две выборки с помощью кнопки *Variables (groups)* (рис. 4.18, 4.19.).

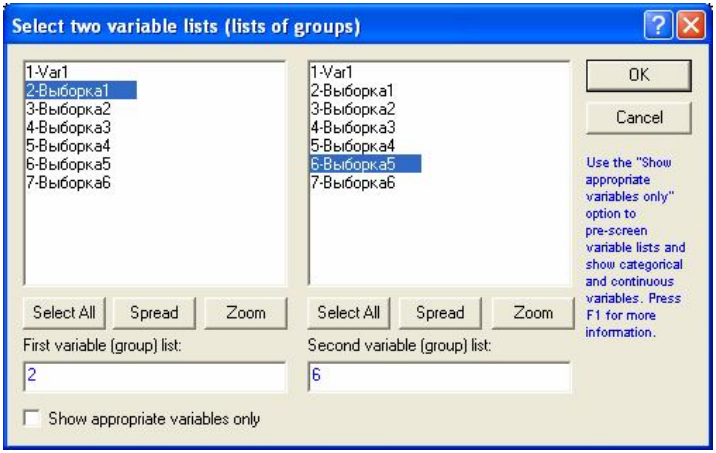


Рис. 4.19. Окно выбора переменных.

На закладке *Quick* находятся уже знакомые нам кнопки *Summary: T-tests Box & whisker plot*. Но в данном случае потребуется закладка *Options*. Необходимо задать уровень значимости (*p-level for highlighting*), в нашем примере, как было установлено выше, $\alpha = 0{,}05$. Дополнительно можно поставить метку в поле для отображения в результатах анализа длинных имен переменных (*Display long variables names*) и включения в отчет расчета определенных критериев (поля *Levene's test* и *Brown & Forsythe test* – критерии, разработанные для выборочного наблюдения) (рис. 4.20.).

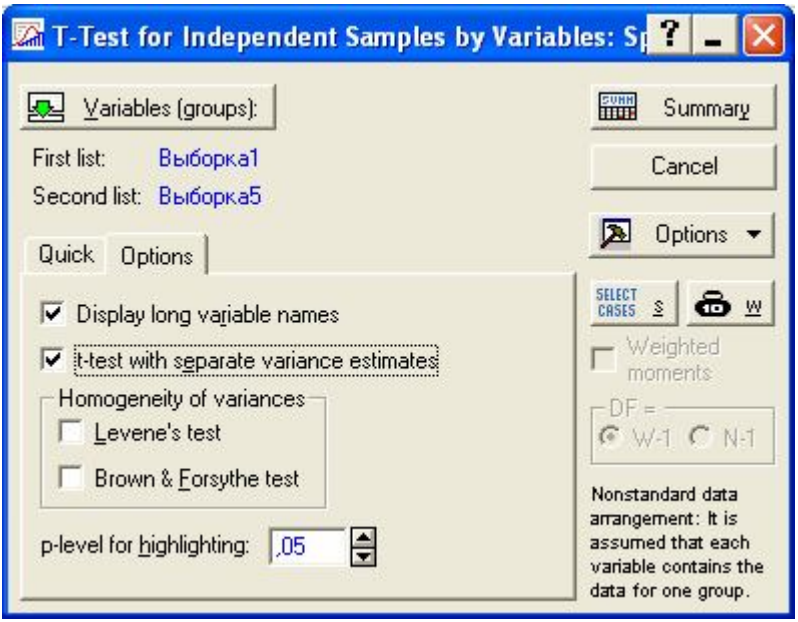


Рис. 4.20. Дополнительные параметры расчета t-критерия.

Если метку в поле *t-test with separate variance estimates* не ставить, это означает, что задача решается в условиях предположения о неизвестных, но равных дисперсиях.

Результаты анализа выводятся в виде строки, что не очень удобно. Чтобы представить результаты в виде столбца, используем процедуру транспонирования: *Data/Transpose/File* (см. раздел 3.2.).

Group 1 vs. Group 2	1 Выборка1 vs. Выборка5
Mean	46,2473684
Mean	49,1578947
t-value	-1,13903978
df	36
p	0,262209471
Valid N	19
Valid N	19
Std.Dev.	6,79062822
Std.Dev.	8,82857204
F-ratio	1,69028922
p	0,27482827

Рис. 4.21. Результаты расчета t-критерия при условии равных дисперсий.

В полученной таблице рассчитаны следующие показатели:

Mean - выборочные средние

$$\widetilde{x}_j = \frac{1}{n_j}\sum_{i=1}^{n} x_{ij} \qquad \text{j = 1,2 ,} \qquad (4.5.)$$

где x_{ij} - i – й элемент j – ой выборки ($i = 1,\ldots,n$, $j = 1,2$)

t-value – *t*-критерий, необходимый для оценки существенности разности двух средних

$$t_{расч} = \frac{\widetilde{x}_1 - \widetilde{x}_2 - \Delta}{\overline{\sigma}_p \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \ , \qquad (4.6.)$$

где $\widetilde{x}_1$ — выборочная средняя первой выборки;

$\widetilde{x}_2$ — выборочная средняя второй выборки;

Δ — гипотетическая разность между генеральными средними, которая в контексте проверяемой нулевой гипотезы принимается равной 0 ($\Delta = 0$). Формула принимает вид:

$$t_{расч} = \frac{\widetilde{x}_1 - \widetilde{x}_2}{\overline{\sigma}_p \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}. \qquad (4.7.)$$

df – число степеней свободы, равное $(n_1 - 1) + (n_2 - 1)$,

где n_1 - объём первой выборки; n_2 - объём второй выборки.

p – расчетный уровень значимости t-критерия;

Valid N – объем выборки;

Std.Dev. - среднее квадратическое отклонение:

$$\widetilde{\sigma} = \sqrt{\widetilde{\sigma}^2} = \sqrt{\frac{1}{n_{j-1}} \sum_{i=1}^{n_j} (x_{ij} - \widetilde{x}_j)^2}, \quad j = 1,2. \qquad (4.8.)$$

Среднее квадратическое отклонение двух оцениваемых выборок:

$$\overline{\sigma}_p = \sqrt{\frac{(n_1 - 1) \cdot \widetilde{\sigma}_1^2 + (n_2 - 1) \cdot \widetilde{\sigma}_2^2}{n_1 + n_2 - 2}}, \qquad (4.9.)$$

где $\widetilde{\sigma}_1^2$ — дисперсия первой выборки;

$\widetilde{\sigma}_2^2$ — дисперсия второй выборки.

F-ratio – *F*-критерий (дисперсионное отношение), используемый для оценки существенности различия значений двух дисперсий:

$$F = \left(\frac{\tilde{\sigma}_1^2}{\tilde{\sigma}_2^2} \right). \qquad (4.10.)$$

p – расчетный уровень значимости F-критерия.

В том случае, если активизируется поле t-test with separate variance estimates, то задача решается в предположении неизвестных и не равных дисперсий. Результаты выводятся по следующей форме (см. рис. 4.24, обращаем внимание на то, что в таблице результатов появляются три новые строчки:

t-separ – расчетное значение t-критерия с учетом различных дисперсий. Очевидно, что в нашем примере оно не изменяется.

df - число степеней свободы *t-критерия* при условии неравных дисперсий определяется по следующим формулам:
если $n_1 \neq n_2$

$$df = m = \frac{\left(\frac{\tilde{\sigma}_1^2}{n_1} + \frac{\tilde{\sigma}_2^2}{n_2} \right)^2}{\frac{\left(\frac{\tilde{\sigma}_1^2}{n_1} \right)^2}{n_1 + 1} + \frac{\left(\frac{\tilde{\sigma}_2^2}{n_2} \right)^2}{n_2 + 1}} - 2, \qquad (4.11.)$$

и, если $n_1 = n_2$

$$m = n - 1 + \frac{2n - 2}{\frac{\tilde{\sigma}_1^2}{\tilde{\sigma}_2^2} + \frac{\tilde{\sigma}_2^2}{\tilde{\sigma}_1^2}}.$$

Расчетное значение m округляется до целого значения в силу того, что число степеней свободы есть целое число по определению.

p – расчетный уровень значимости t-критерия при условии неизвестных и неравных дисперсий.

Group 1 vs. Group 2	1 Выборка1 vs. Выборка5
Mean	46,2473684
Mean	49,1578947
t-value	-1,13903978
df	36
p	0,262209471
t separ.	-1,13903978
df	33,7762994
p	0,262700782
Valid N	19
Valid N	19
Std.Dev.	6,79062822
Std.Dev.	8,82857204
F-ratio	1,69028922
p	0,27482827

Рис. 4.22. Результаты расчета t-критерия при условии неравных дисперсий.

Гипотеза принимается, если $\left|t_{\text{ðàñ÷}}\right| \le t_{\text{òàáë}}$. В нашем примере

$$t_{расч} = \frac{46{,}247 - 49{,}158}{7{,}876 \cdot \sqrt{\frac{1}{19} + \frac{1}{19}}} = -1{,}139 .$$

Табличное значение t-критерия равно $t_{36\,;\,0.05} = 2{,}028$ (уровень значимости – 0,05, число степеней свободы - 36).

Таким образом, $t_{расч} = 1{,}139 < t_{табл} = 2{,}028$, следовательно, испытуемая гипотеза принимается.

Аналогичный вывод можно получить на основе сравнения расчетного и принятого уровней значимости:

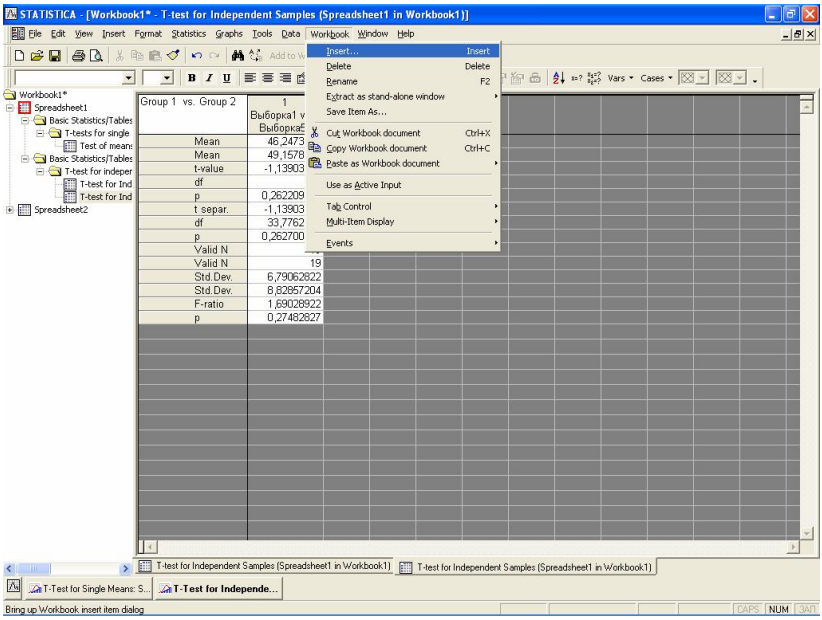
$$p = 0{,}2627 > \alpha = 0{,}05 .$$

4.4. Графическое представление результатов выборочного наблюдения

Для наглядного и компактного представления результатов проведенного выборочного наблюдения воспользуемся графическими возможностями ППП STATISTICA. Весьма существенным, с дидактической точки зрения, является то, что последовательное выполнение рассматриваемых лабораторных работ, дает возможность наглядного сравнения результатов выборочного и сплошного наблюдений. Вполне очевидно, что, по определению, такое сравнение исключено в реальных практических условиях.

Прежде чем приступить к построению графика, потребуется выполнить ряд дополнительных преобразований.

Необходимо создать новую рабочую книгу. Для этого воспользуемся меню *Workbook/Insert* (рис. 4.23.).



4.23. Запуск меню вставки нового документа в рабочую книгу.

В появившемся окне выбираем пункт STATISTICA document и нажимаем *OK* (рис 4.24.).

Рис. 4.24. Меню выбора нового документа.

В новом окне ставим метку в поле *Create new* (создать новый) и выбираем пункт *Spreadsheet* (рис. 4.25.) .

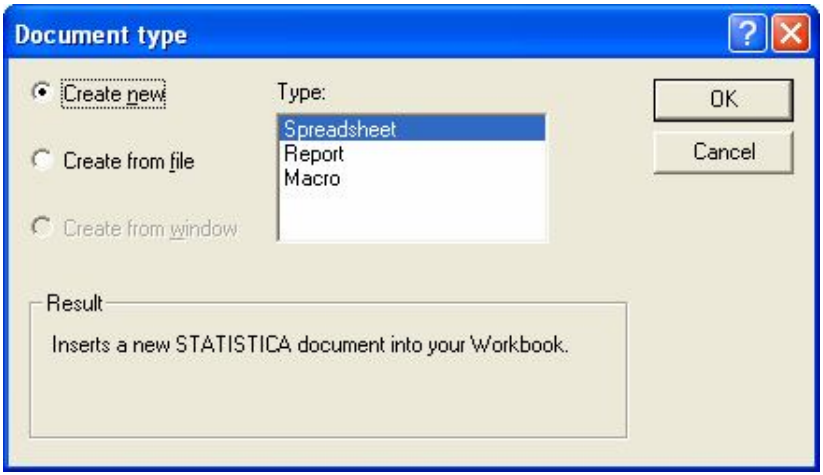


Рис. 4.25. Меню создания нового рабочего листа.

Система создает электронную таблицу с десятью графами (рис. 4.26.). Этот рабочий лист делаем активным.

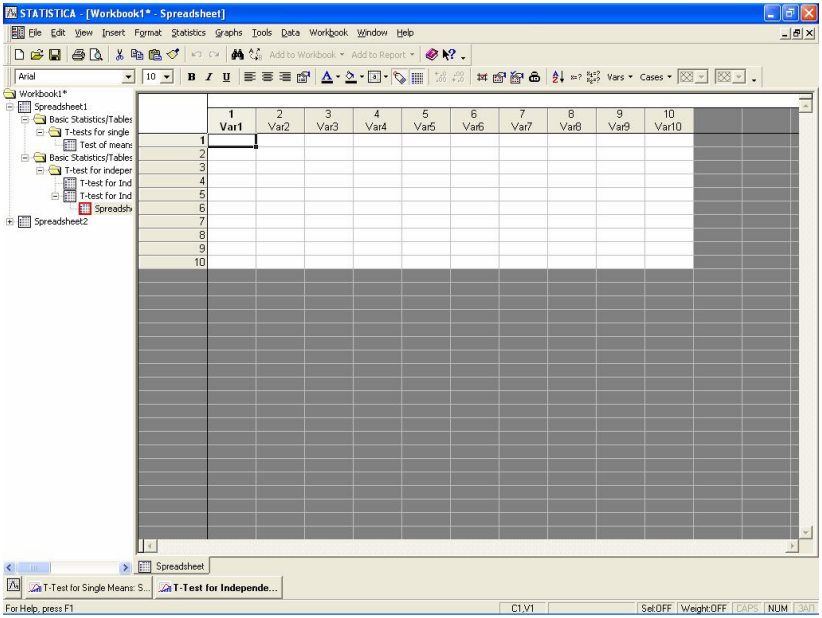


Рис. 4.26. Вид пустого рабочего листа размером 10x10.

Нам необходимо добавить еще 9 переменных, поскольку требуется задать значения следующих характеристик: генеральной средней (ген.ср.), шести выборочных средних (выб.ср. 1-6) и нижние (н.г. 1-6) и верхние (в.г. 1-6) границы доверительных интервалов, рассчитанные по каждой выборке. Записи по генеральной средней, а также по каждой выборке делаются на разных строчках (рис. 4.27.).

	1 ген.ср	2 выб.ср.1	3 н.г.1	4 в.г.1	5 выб.ср.2	6 н.г.2	7 в.г.2	8 выб.ср.3	9 н.г.3	10 в.г.3	11 выб.ср.4	12 н.г.4
1	47,10465											
2		46,24737	42,97439	49,52035								
3					47,56842	42,00526	53,13158					
4								48,91053	44,13234	53,68871		
5											48,93158	44,50370
6												
7												

Рис. 4.27. Рабочая таблица для построения диаграмм размаха.

Далее выбираем меню *Graphs/2D Graphs/Range Plots* (диаграммы размаха) (рис. 4.29.).

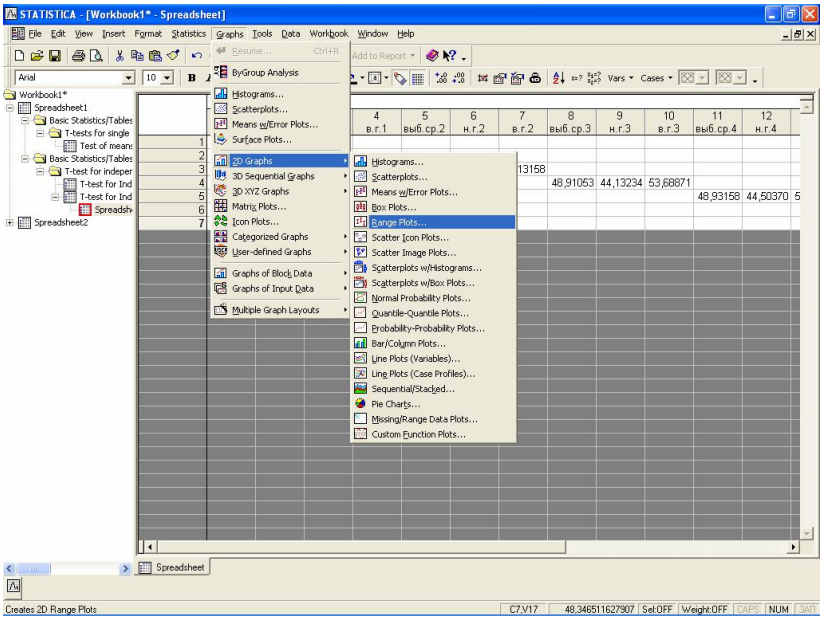


Рис. 4.28. Запуск меню для построения диаграмм размаха.

В появившемся диалоговом окне выбираем тип диаграммы *Whiskers*, чтобы явно показать положение доверительных интервалов, и значений выборочных средних относительно значения генеральной средней. В поле *Mode (range values*) можно задать построение диаграммы в абсолютных величинах *(Absolute*) или в процентах к середине интервала (*Relative to Mid-point*) (см. рис. 4.29.).

Далее нажимаем на кнопку *Variables*, в появившемся окне выбора переменных в поле *Mid-point* выбираются переменные со значением генеральной и выборочных средних. В поле *Max* заносятся переменные: генеральная средняя, верхние границы доверительных интервалов. В поле *Min* – генеральная средняя и нижние границы доверительных интервалов. Далее нажимаем *OK* (рис. 4.30.).

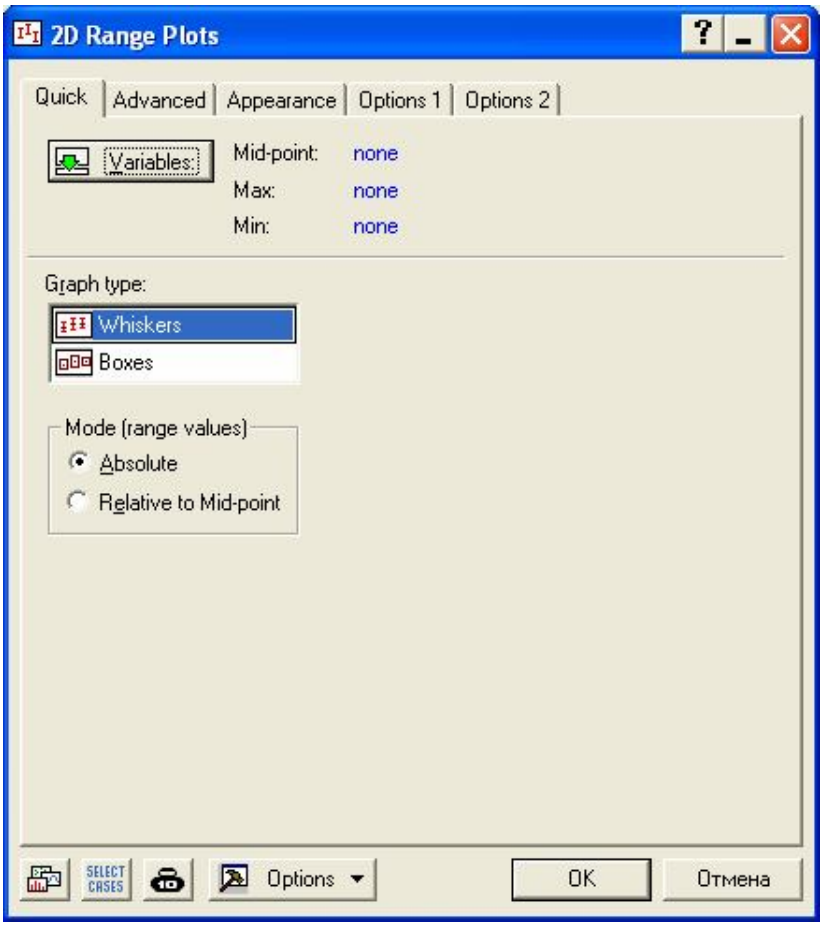


Рис. 4.29. Меню построения диаграмм размаха.

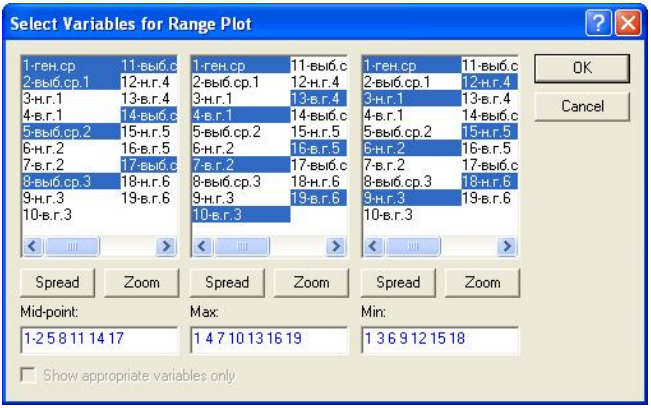


Рис. 4.30. Окно выбора переменных.

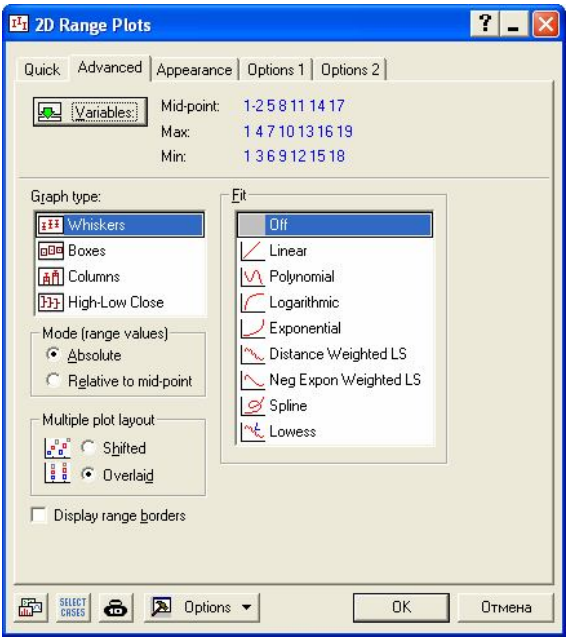


Рис. 4.31. Дополнительные настройки диаграммы размаха.

На закладке *Advanced* в поле *Multiple plot layout* можно выбрать способ построения диаграммы размаха: каждая диаграмма на графике располагается отдельно (*Shifted*), либо они накладываются друг на друга (*Overlaid)*. Выбираем второй вариант (рис. 4.31.). Так же на закладке *Option 1* можно задать название графика в поле *Custom Title* (рис. 4.32.). После этого нажимаем кнопку *OK*.

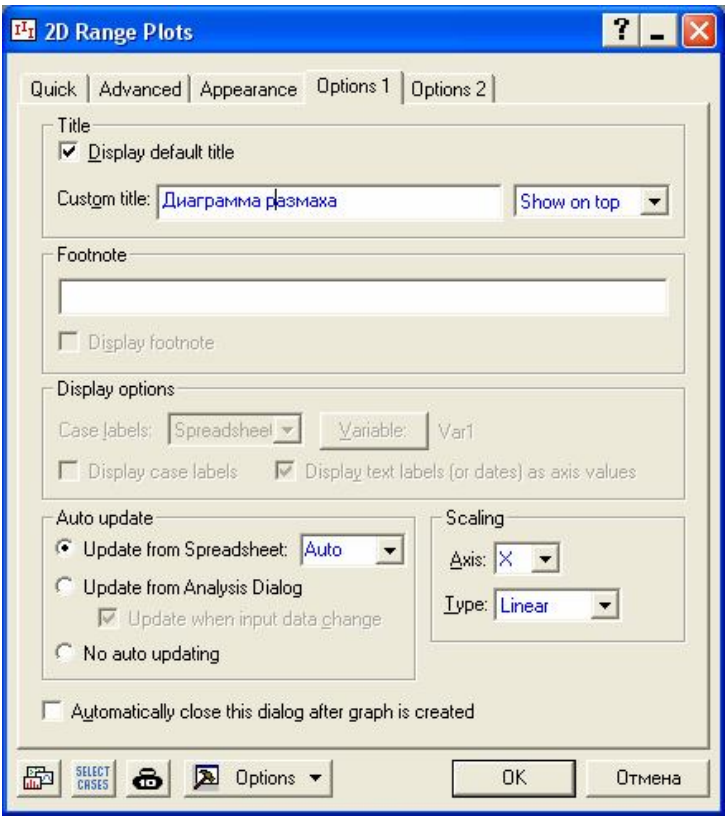


Рис. 4.32. Закладка Options в настройках диаграммы размаха.

Итогом являются графики следующего вида:

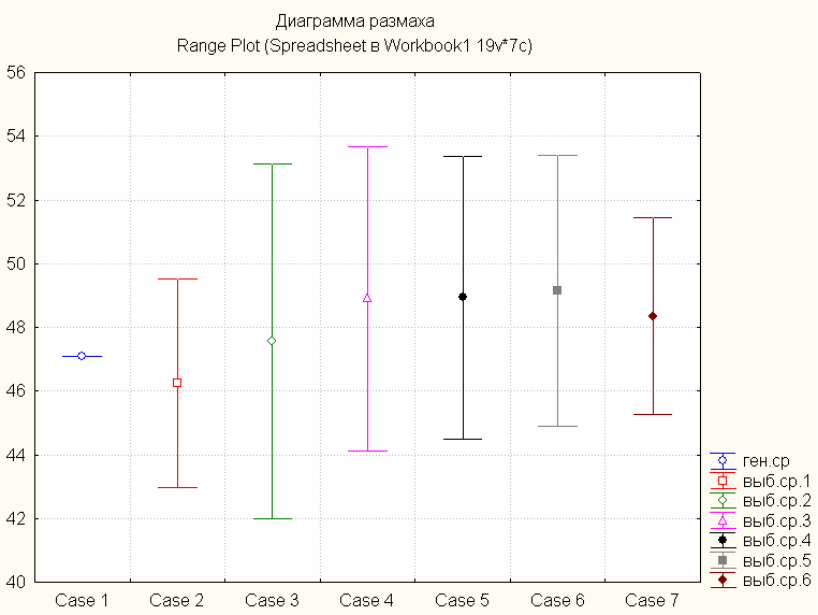


Рис. 4.33. Графическое сравнение результатов сплошного и выборочного наблюдения

График наглядно показывает, что доверительные интервалы, построенные по всем выборкам, накрывают генеральную среднюю, что естественно. Если бы, какой либо доверительный интервал, рассчитанный по результатам выборки, не включал в себя значение генеральной средней, то в реальных условиях, это означало бы получение ошибочного вывода на основе выборки.

Диаграмма наглядно демонстрирует возможный результат выборочного зондирования исследуемой генеральной совокупности и убедительно иллюстрирует объективную неоднозначность выводов, формулируемых на основе выборочных данных.

Заключение

Данное учебное пособие, во-первых, знакомит с необходимыми общесистемными приемами работы в среде пакета прикладных программ STATISTICA, владение которыми позволит пользователю непосредственно приступить к работе со специализированными вычислительными программами. Во-вторых, позволяет освоить конкретные методы статистического анализа. На сквозном числовом примере достаточно подробно рассмотрен блок программ, обеспечивающий решение таких популярных и важных в статистической практике задач, как анализ эмпирических распределений и оценка параметров генеральной совокупности на основе выборочных данных.

Следующие выпуски пособия будут посвящены возможностям реализации в рамках ППП STATISTICA статистических методов исследования связей и зависимостей, а также анализа временных рядов и прогнозирования.

Список использованных источников

1. **Боровиков В. П.**, STATISTICA. Искусство анализа данных на компьютере: для профессионалов / В. П. Боровиков. – 2-е изд. – СПб. : – 2003. – 688 с.

2. **Венецкий И. Г.**, Основные математико-статистические понятия и формулы в экономическом анализе. Справочник / И. Г. Венецкий, В. И. Венецкая. – 2-е изд., перераб. и доп. – М. : Статистика, 1979. – 477 с.

3. **Ефимова М.Р.**, Общая теория статистики: учеб. / М. Р. Ефимова, Е. В. Петрова, В. Н. Румянцев. – М. : ИНФРА-М, 2002. – 416 с.

4. **Закс Л.**, Статистическое оценивание: Пер. с нем / Л. Закс. – М. : Статистика, 1976. – 597 с.

5. **Куприенко Н. В.**, Статистика. Методы анализа распределений. Выборочное наблюдение: учеб. пособие / Н. В. Куприенко, О. А. Пономарева. – СПб. : Издательство СПбГПУ, 2005. – 128 с.

6. Статистика для менеджеров. 4-е изд.: Пер. с англ / Д. Левин и др. – М. : Изд. дом «Вильямс», 2004. – 1312 с.

7. **Сигел Э.** Практическая бизнес-статистика : Пер. с англ / Э. Сигел. – М. : Изд. дом «Вильямс», 2002. – 1056 с.8. Статистика.: Учебник / Под ред. проф. И.И. Елисеевой. М.: ООО “ВИТРЭМ”, 2005. С. 448.

9. Теория статистики. : учеб. / Под ред. Р. А. Шмойловой. – М. : Финансы и статистика, 2005. – 560 с.

10. Теория статистики: учеб. / Под ред. проф. Г. Л. Громыко. – 2-е изд., перераб. и доп. – М. : ИНФРА-М, 2006. – 476 с.

11. Abridged from Table 12 of Biometrika Tables for Statisticians, Vol. 1, edited by E.S. Pearson and H.O. Hartley (London: Cambridge University Press, 1962).

12. M. Merrington and C.M. Thompson, Tables of Percentage Points of the Inverted Beta (D-Distribution, Biometrika 33 (1943): 73-88. Reproduced by permission of the Biometrika Trustees.

13. Регионы России. Социально-экономические показатели. 2006: Стат.сб. М., 2004. С. 981.

14. www.statsoft.ru (сайт компании StatSoft Russia – документация по ППП STATISTICA).

15. www.exponenta.ru (примеры решения практических задач в ППП STATISTICA).

Приложение 1

Лабораторная работа №1
«Статистический анализ эмпирических распределений»

Общие сведения

Целью работы является освоение методики и приобретение практических навыков анализа распределений, включающего расчет основных статистических характеристик, графическое и табличное представление рядов распределения, аппроксимацию эмпирического распределения, подбор модельного распределения с использованием критериев согласия. Лабораторная работа может быть выполнена на основе фактических материалов, публикуемых в официальных статистических изданиях или иных источниках, либо с использованием условных данных (случайных чисел), моделируемых с помощью специальной программы. Исходные данные могут быть предложены преподавателем или выбраны студентом, исходя из области его интересов.

Требования, предъявляемые к работе

В каждом разделе лабораторной работы должны быть кратко изложены основные теоретические положения по соответствующим проблемам. В разделах, посвященных расчету и анализу статистических характеристик, необходимо привести формулы. Основной акцент в работе следует сделать на содержательную интерпретацию полученных результатов. Работа должна быть оформлена в соответствии с требованиями ГОСТ.

Структура работы

Введение

Во введении целесообразно раскрыть понятие ряда распределения, цели изучения рядов распределения, дать характеристику исходных данных, с указанием источника информации.

1.Табличное и графическое представление вариационного ряда

Первый шаг - ранжирование исходных данных, определение наличия выбросов и работа с ними.

Табличное представление вариационного ряда предполагает определение числа групп, величины группировочного интервала, поэтапное построение вариационного ряда (с разным числом групп), при этом следует стремиться к одновершинному распределению и отсутствию нулевых и малонаполненных групп. В окончательном варианте

таблицы должны содержаться частоты, частости, накопленные частоты и частости; при использовании неравных интервалов – показатели плотности распределения.

Графически вариационный ряд необходимо представить в виде полигона, гистограммы и кумуляты распределения.

2.Характеристика центральной тенденции распределения

Расчет и анализ показателей центра распределения: среднего арифметического значения, моды и медианы.

3. Оценка вариации изучаемого признака

Расчет и анализ следующих показателей: размах вариации, дисперсия, среднее квадратическое отклонение, коэффициент вариации.

4. Характеристика структуры распределения

Расчет и анализ показателей: медиана, квартили, децили.

5. Характеристика формы распределения

Расчет и анализ показателей: коэффициент асимметрии, коэффициент эксцесса (куртозис).

6. Сглаживание эмпирического распределения. Проверка гипотезы о законе распределения

Расчет и анализ значений критерия согласия Пирсона для оценки соответствия эмпирического распределения некоторым типам теоретических распределений (определяются исходя из свойств распределения анализируемой совокупности), графическое представление сглаживания эмпирического распределения кривыми теоретических распределений.

Заключение

В заключении по результатам анализа необходимо дать общую характеристику изучаемого распределения.

Лабораторная работа №2 «Проведение выборочного наблюдения»

Общие сведения

Целью лабораторной работы является освоение методики организации и проведения выборочного наблюдения; статистических методов и методов компьютерной обработки полученной информации; методов оценки параметров генеральной совокупности на основе выборочных данных. Лабораторная работа может быть выполнена на основе исходных данных 1-й лабораторной работы или иных фактических материалов, публикуемых в официальных статистических изданиях и других источниках, либо с использованием условных данных (случайных чисел), моделируемых с помощью специальной программы. Исходные данные могут быть предложены преподавателем или выбраны студентом, исходя из области его интересов.

Требования, предъявляемые к работе

В каждом разделе лабораторной работы должны быть кратко изложены основные теоретические положения по соответствующим проблемам. В разделах, содержащих расчет и анализ статистических показателей, необходимо привести соответствующие формулы. Основной акцент в работе следует сделать на содержательную интерпретацию полученных результатов. Работа должна быть оформлена в соответствии с требованиями ГОСТ.

Структура работы

Введение

Во введении необходимо раскрыть само понятие выборочного наблюдения, как важнейшего вида не сплошного статистического наблюдения, его преимущества и области применения; перечислить виды выборки и способы отбора единиц в выборочную совокупность; дать характеристику исходных данных, указав источник информации.

1. Расчет необходимого объема выборочной совокупности

Дать понятие ошибки выборки, факторов, определяющих ее величину. Задавая разные значения ошибки репрезентативности, рассчитать необходимый объем выборки. Величина ошибки устанавливается по результатам теоретического анализа объекта исследования. Сделать выводы и окончательный выбор соотношения: ошибка/объем выборки.

2. Формирование выборочных совокупностей и обработка выборочных данных

Методом случайного бесповторного отбора (используя ППП «STATISTICA») сформировать 5 малых и одну большую выборки. Для каждой совокупности рассчитать основные статистические харак-

теристики (см. раздел 4 данного пособия), сравнить полученные результаты, сделать выводы.

3. Распространение результатов выборочного наблюдения на генеральную совокупность

Продемонстрировать расчет доверительных интервалов для генеральной средней (с вероятностью 90% , 95% или 99%, по указанию преподавателя). Прокомментировать полученные результаты. Провести сравнительный анализ результатов, полученных по выборкам не равного объема; объяснить наблюдаемые различия в результатах по выборкам равного объема. Графически представить результаты выборочного наблюдения.

4. Проверка статистических гипотез о значении генеральной средней и о равенстве двух выборочных средних

В заключительной части работы следует продемонстрировать методику проверки статистической гипотезы о значении генеральной средней, используя значение средней величины, по данным первой лабораторной работы (если вторая работа выполняется на оригинальных исходных данных, следует использовать приближенную величину); проверку гипотезы о равенстве двух средних (о принадлежности двух выборок одной генеральной совокупности) осуществить на основе двух выборок, у которых разность в средних величинах максимальна. Сделать выводы.

Приложение 3

Федеральное агентство по образованию
Государственное образовательное учреждение высшего
профессионального образования
«Санкт-Петербургский государственный
политехнический университет»

Факультет экономики и менеджмента

Кафедра «Предпринимательство и коммерция»

Лабораторная работа №1
по дисциплине «Статистика»
на тему «Анализ эмпирического распределения»

Выполнил: студент _____ группы

_____
(подпись)

Иванов И. И.
(Фамилия И.О.)

Принял: _____
(должность, ученая степень)

_____
(подпись)

_____
(Фамилия И.О.)

_____
(Дата)

Санкт-Петербург
200_

Таблица значений функции[5] $f(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}$

(плотность нормального распределения)[6]

t	*0*	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*
0.0	3989	3989	3989	3988	3986	3984	3982	3980	3977	3973
0.1	3970	3965	3961	3956	3951	3945	3939	3932	3925	3918
0.2	3910	3902	3894	3885	3876	3867	3857	3847	3836	3825
0.3	3814	3802	3790	3778	3765	3752	3739	3725	3712	3697
0.4	3683	3668	3653	3637	3621	3605	3589	3572	3555	3538
0.5	3521	3503	3485	3467	3448	3429	3410	3391	3372	3352
0.6	3332	3312	3292	3271	3251	3230	3209	3187	3166	3144
0.7	3123	3101	3079	3056	3034	3011	2989	2966	2943	2920
0.8	2897	2874	2850	2827	2803	2780	2756	2732	2709	2685
0.9	2661	2637	2613	2589	2565	2541	2516	2492	2468	2444
1.0	2420	2396	2371	2347	2323	2299	2275	2251	2227	2203
1.1	2179	2155	2131	2107	2083	2059	2036	2012	1989	1965
1.2	1942	1919	1895	1872	1849	1826	1804	1781	1758	1736
1.3	1714	1691	1669	1647	1626	1604	1582	1561	1539	1518
1.4	1497	1476	1456	1435	1415	1394	1374	1354	1334	1315
1.5	1295	1276	1257	1238	1219	1200	1182	1163	1145	1127
1.6	1109	1092	1074	1057	1040	1023	1006	0989	0973	0957
1.7	0940	0925	0909	0893	0878	0863	0848	0833	0818	0804
1.8	0790	0775	0761	0748	0734	0721	0707	0694	0681	0669
1.9	0656	0644	0632	0620	0608	0596	0596	0584	0573	0562
2.0	0540	0529	0519	0508	0498	0488	0478	0468	0449	0449
2.1	0440	0431	0422	0413	0404	0396	0387	0379	0371	0363
2.2	0355	0347	0339	0332	0325	0317	0310	0303	0297	0290
2.3	0283	0277	0270	0264	0258	0252	0246	0241	0235	0229
2.4	0224	0219	0213	0203	0203	0198	0194	0189	0184	0180
2.5	0175	0171	0167	0163	0158	0154	0151	0143	0139	0147
2.6	0136	0132	0129	0126	0122	0119	0116	0113	0110	0107
2.7	0104	0101	0099	0096	0093	0091	0088	0086	0084	0081
2.8	0079	0077	0075	0073	0071	0069	0067	0065	0063	0061
2.9	0060	0058	0056	0055	0053	0051	0050	0048	0047	0046
3.0	0044	0043	0042	0040	0039	0038	0037	0036	0035	0034
4.0	0001	0001	0001	0000	0000	0000	0000	0000	0000	0000

[5] Все значения функции умножены на 10 000

[6] Источник: Венецкий И.Г., Венецкая В.И. Основные математико-статистические понятия и формулы в экономическом анализе : справ. –2-е изд., перераб. и доп. – М. : Статистика, 1979. – 447 с.

Приложение 5

Таблица критических значений ***t*** - критерия *Student'a*[7]

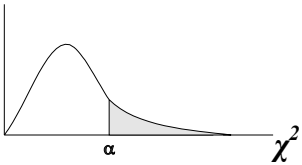
1 - односторонний критерий , 2 - двусторонний критерий

1	*.50*	*.20*	*.10*	*.05*	*.02*	*.01*	*.005*	*.002*	*.001*
2	*.25*	*.10*	*.05*	*.025*	*.01*	*.005*	*.0025*	*.001*	*.0005*
1	1.000	3.078	6.314	12.70	31.82	63.63	127.3	318.3	636.6
2	.816	1.866	2.920	4.303	6.965	9.925	14.08	22.32	31.59
3	.765	1.638	2.353	3.182	4.541	5.841	7.453	10.21	12.92
4	.741	1.533	2.132	2.776	3.747	4.604	5.598	7.173	8.610
5	.727	1.440	1.943	2.447	3.143	3.707	4.317	5.208	5.959
6	.718	1.440	1.943	2.447	3.143	3.707	4.317	5.208	5.959
7	.711	1.415	1.895	2.365	2.998	3.499	4.020	4.785	5.408
8	.706	1.397	1.860	2.306	2.896	3.355	3.833	4.501	5.041
9	.703	1.383	1.833	2.262	2.821	3.250	3.690	4.297	4.781
10	.700	1.372	1.812	2.228	2.764	3.169	3.581	4.144	4.537
11	.697	1.363	1.796	2.201	2.718	3.106	3.497	4.025	4.437
12	.695	1.356	1.782	2.179	2.681	3.055	3.428	3.930	4.318
13	.694	1.350	1.771	2.160	2.650	3.012	3.372	3.852	4.221
14	.692	1.345	1.761	2.145	2.624	2.977	3.326	3.787	4.140
15	.691	1.341	1.753	2.131	2.602	2.947	3.286	3.733	4.073
16	.690	1.337	1.746	2.120	2.583	2.921	3.252	3.686	4.015
17	.689	1.333	1.740	2.110	2.576	2.898	3.222	3.646	3.965
18	.688	1.330	1.734	2.101	2.552	2.878	3.197	3.610	3.922
19	.688	1.328	1.729	2.093	2.539	2.861	3.174	3.579	3.883
20	.687	1.325	1.725	2.086	2.528	2.845	3.153	3.552	3.850
21	.686	1.323	1.721	2.080	2.518	2.831	3.135	3.257	3.189
22	.686	1.321	1.717	2.074	2.508	2.819	3.119	3.505	3.792
23	.685	1.319	1.714	2.069	2.500	2.807	3.104	3.485	3.767
24	.685	1.318	1.711	2.064	2.492	2.797	3.091	3.467	3.745
25	.684	1.316	1.708	2.060	2.485	2.787	3.078	3.450	3.725
26	.684	1.315	1.706	2.056	2.479	2.779	3.067	3.435	3.707
27	.684	1.314	1.703	2.052	2.473	2.771	3.057	3.421	3.690
28	.683	1.313	1.701	2.048	2.467	2.763	3.047	3.408	3.674
29	.683	1.311	1.699	2.045	2.462	2.756	3.038	3.396	3.659
30	.683	1.310	1.697	2.042	2.457	2.750	3.030	3.385	3.646
40	.681	1.303	1.684	2.021	2.423	2.704	2.971	3.307	3.551
60	.679	1.296	1.671	2.000	2.390	2.660	2.915	3.232	3.460
120	.677	1.289	1.658	1.980	2.358	2.617	2.860	3.160	3.373
∞	.674	1.282	1.645	1.960	2.326	2.576	2.807	3.090	3.291

[7] Источник: Abridged from Table 12 // Biometrika Tables for Statisticians 1962. Vol. 1 / Ed. E.S. Pearson, H.O. Hartley. London: Cambridge University Press, 1962.

Приложение 6

Таблица критических значений χ^2
в зависимости от числа степеней свободы (r)
и уровня значимости (α)



r \ α	.500	.250	.100	.050	.025	.010	.005	.001
1	0.455	1.323	2.706	3.841	5.024	6.635	7.879	10.83
2	1.386	2.773	4.605	5.991	7.378	9.210	10.60	13.82
3	2.366	4.108	6.251	7.815	9.348	11.34	12.84	16.27
4	3.357	5.385	7.779	9.488	11.14	13.28	14.86	18.47
5	4.351	6.626	9.236	11.07	12.83	15.09	16.75	20.52
6	5.348	7.841	10.64	12.59	14.45	16.81	18.55	22.46
7	6.346	9.037	12.02	14.07	16.01	18.48	20.28	24.32
8	7.344	10.22	13.36	15.51	17.53	20.09	21.96	26.12
9	8.343	11.39	14.68	16.92	19.02	21.67	23.59	27.88
10	9.342	12.55	15.99	18.31	20.48	28.21	25.19	29.59
11	10.34	13.70	17.28	19.68	21.92	24.72	26.76	31.26
12	11.34	14.85	18.55	21.03	23.34	26.22	28.30	32.91
13	12.34	15.98	19.81	22.36	24.74	27.79	29.82	34.53
14	13.34	17.12	21.06	23.68	26.12	29.14	31.32	36.14
15	14.34	18.25	22.31	25.00	27.49	30.58	32.80	37.70

Приложение 6 (продолжение)

r \ α	.500	.250	.100	.050	.025	.010	.005	.001
16	15.34	19.37	23.54	26.30	28.85	32.00	34.27	39.25
17	16.34	20.49	24.77	27.59	30.19	33.41	35.72	40.79
18	17.34	21.60	25.99	28.87	31.53	34.81	37.16	42.31
19	18.34	22.72	27.20	30.14	32.85	36.19	38.58	43.82
20	19.34	23.83	28.41	31.41	34.17	37.57	40.00	45.32
21	20.34	24.93	29.62	33.67	35.48	38.93	41.40	46.80
22	21.34	26.04	30.81	33.92	36.78	40.29	42.80	48.27
23	22.34	27.14	32.01	35.18	38.08	41.64	44.18	49.73
24	23.34	28.24	33.20	36.42	39.36	42.98	45.56	51.18
25	24.34	29.34	34.38	37.65	40.65	44.31	46.93	52.62
26	25.34	30.43	35.56	38.89	41.92	45.64	48.29	54.05
27	26.34	31.51	36.74	40.11	43.19	46.96	49.64	55.48
28	27.34	32.62	37.92	41.34	44.46	48.28	50.99	56.89
29	28.34	33.71	39.09	42.56	45.72	49.59	52.34	58.30
30	29.34	34.80	40.26	43.77	46.98	50.89	53.67	59.70
40	39.34	45.62	51.81	55.76	59.34	63.69	66.77	73.40
50	49.33	56.33	63.17	67.50	71.42	76.15	79.49	86.66
60	59.33	66.98	74.40	79.08	83.30	88.38	91.95	99.61
70	69.33	77.58	85.53	90.53	95.02	100.4	104.2	112.3
80	79.33	88.13	96.58	101.9	106.6	112.3	116.3	124.8
90	89.33	98.65	107.6	113.1	118.1	124.1	128.3	137.2
100	99.33	109.1	118.5	124.3	129.6	135.8	140.2	149.4

Источник: Abridged from Table 8 of Biometrika Tables for Statisticians, Vol. 1, edited E.S. Pearson and H.O. Hartley (London: Cambridge University Press, 1962).

Приложение 7

Таблица значений F-распределения[8], $\alpha = 0.05$

	Число степеней свободы								
	1	2	3	4	5	6	7	8	9
1	161.4	199.5	215.7	224.6	230.2	234.0	236.8	238.9	240.5
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10
7	5.59	4.74	4.35	4.12	3.97.	3.87	3.79	3.73	3.68
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18
10	4.96	4.10	3.71	3.48	3.33.	3.22	3.14	3.07	3.02
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2:39
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04
120	3.92	3.07	2.68	2.45	2.29	2.17	2.09	2.02	1.96
∞	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88

Приложение 7 (продолжение)

	Число степеней свободы									
	10	12	15	20	24	30	40	60	120	∞
1	241.9	243.9	245.9	248.0	249.1	250.1	251.1	252.2	253.3	254.3
2	19.40	19.41	19.43	19.45	19.45	19.46	19.47	19.48	19.49	19.50
3	8.79	8.74	8.70	8.66	8.64	8.62	8.59	8.57	8.55	8.53
4	5.96	5.91	5.86	5.80	5.77	5.75	5.72	5.69	5.66	5.63
5	4.74	4.68	4.62	4.56	4.53	4.50	4.46	4.43	4.40	4.36
6	4.06	4.00	3.94	3.87	3.84	3.81	3.77	3.74	3.70	3.67
7	3.64	3.57	3.51	3.44	3.41	3.38	3.34	3.30	3.27	3.23
8	3.35	3.28	3.22	3.15	3.12	3.08	3.04	3.01	2.97	2.93
9	3.14	3.07	3.01	2.94	2.90	2.86	2.83	2.79	2.75	2.71
10	2.98	2.91	2.85	2.77	2.74	2.70	2.66	2.62	2.58	2.54
11	2.85	2.79	2.72	2.65	2.61	2.57	2.53	2.49	2.45	2.40
12	2.75	2.69	2.62	2.54	2.51	2.47	2.43	2.38	2.34	2.30
13	2.67	2.60	2.53	2.46	2.42	2.38	2.34	2.30	2.25	2.21
14	2.60	2.53	2.46	2.39	2.35	2.31	2.27	2.22	2.18	2.13
15	2.54	2.48	2.40	2.33	2.29	2.25	2.20	2.16	2.11	2.07
16	2.49	2.42	2.35	2.28	2.24	2.19	2.15	2.11	2.06	2.01
17	2.45	2.38	2.31	2.23	2.19	2.15	2.10	2.06	2.01	1.96
18	2.41	2.34	2.27	2.19	2.15	2.11	2.06	2.02	1.97	1.92
19	2.38	2.31	2.23	2.16	2.11	2.07	2.03	1.98	1.93	1.88
20	2.35	2.28	2.20	2.12	2.08	2.04	1.99	1.95	1.90	1.84
21	2.32	2.25	2.18	2.10	2.05	2.01	1.96	1.92	1.87	1.81
22	2.30	2.23	2.15	2.07	2.03	1.98	1.94	1.89	1.84	1.78
23	2.27	2.20	2.13	2.05.	2.01	1.96	1.91	1.86	1.81	1.76
24	2.25	2.18	2.11	2.03	1.98	1.94	1.89	1.84	1.79	1.73
25	2.24	2.16	2.09	2.01	1.96	1.92	1.87	1.82	1.77	1.71
26	2.22	2.15	2.07	1.99	1.95	1.90	1.85	1.80	1.75	1.69
27	2.20	2.13	2.06	1.97	1.93	1.88	1.84	1.79	1.73	1.67
28	2.19	2.12	2.04	1.96	1.91	1.87	1.82	1.77	1.71	1.65
29	2.18	2.10	2.03	1.94	1.90	1.85	1.81	1.75	1.70	1.64
30	2.16	2.09	2.01	1.93	1.89	1.84	1.79	1.74	1.68	1.62
40	2.08	2.00	1.92	1.84	1.79	1.74	1.69	1.64	1.58	1.51
60	1.99	1.92	1.84	1.75	1.70	1.65	1.59	1.53	1.47	1.39
120	1.91	1.83	1.75	1.66	1.61	1.55	1.50	1.43	1.35	1.25
∞	1.83	1.75	1.67	1.57	1.52	1.46	1.39	1.32	1.22	1.00

[8] Источник: Merrington M. , Thompson C. M. – Tables of Percentage Points of the Inverted Beta (D-Distribution// Biometrika. 1943. 33. P 73-88. Reproduced by permission of the Biometrika Trustees. Приложения 4 - 8 приведены с сохранением знаков источников.

Приложение 8

Значения интеграла вероятностей $F(t) = \frac{1}{\sqrt{2\pi}} \int_{-t}^{+t} e^{-\frac{t^2}{2}} dt$

t	Сотые доли									
	0	1	2	3	4	5	6	7	8	9
0,0	0000	0080	0160	0239	0319	0399	0478	0558	0638	0717
0,1	0797	0876	0955	1034	1114	1192	1271	1350	1428	1507
0,2	1585	1663	1741	1819	1897	1974	2051	2128	2205	2282
0,3	2358	2434	2510	2586	2661	2737	2812	2886	2961	3035
0,4	3108	3182	3255	3328	3401	3473	3545	3616	3688	3759
0,5	3829	3899	3969	4039	4108	4177	4245	4313	4381	4448
0,6	4515	4581	4647	4713	4778	4843	4907	4971	5035	5098
0,7	5161	5223	5285	5346	5407	5467	5527	5587	5646	5705
0,8	5763	5821	5878	5935	5991	6047	6102	6157	6211	6265
0,9	6319	6372	6424	6476	6528	6579	6629	6679	6729	6778
1,0	6827	6875	6923	6970	7017	7063	7109	7154	7199	7243
1,1	7287	7330	7373	7415	7457	7499	7540	7580	7620	7660
1,2	7699	7737	7775	7813	7850	7887	7923	7959	7995	8030
1,3	8064	8098	8132	8165	8198	8230	8262	8293	8324	8355
1,4	8385	8415	8444	8473	8501	8529	8557	8584	8611	8638
1,5	8664	8690	8715	8740	8764	8789	8812	8836	8859	8882
1,6	8904	8926	8948	8969	8990	9011	9031	9051	9070	9089
1,7	9109	9127	9146	9164	9182	9199	9216	9233	9249	9265
1,8	9281	9297	9312	9327	9342	9357	9371	9385	9399	9412
1,9	9425	9439	9451	9464	9476	9488	9500	9512	9523	9534
2,0	9545	9556	9566	9576	9586	9596	9606	9615	9625	9634
2,1	9643	9651	9660	9668	9676	9684	9692	9700	9707	9715
2,2	9722	9729	9736	9743	9749	9755	9762	9768	9774	9780
2,3	9785	9791	9797	9802	9807	9812	9817	9822	9827	9832
2,4	9836	9840	9845	9849	9853	9857	9861	9865	9869	9872
2,5	9876	9879	9883	9886	9889	9892	9895	9898	9901	9904
2,6	9907	9909	9912	9915	9917	9920	9924	9926	9927	9929
2,7	9931	9933	9935	9937	9939	9940	9942	9944	9946	9947
2,8	9949	9950	9952	9953	9955	9956	9958	9959	9960	9961
2,9	9963	9964	9965	9966	9967	9968	9969	9970	9971	9972
3,0	99730	99739	99747	99755	99763	99771	99779	99786	99793	99800
3,1	99807	99813	99819	99825	99831	99837	99842	99847	99853	99858
3,2	99863	99867	99872	99876	99880	99884	99889	99892	99896	99900
3,3	99903	—	—	—	—	—	—	—	—	—
3,4	99933	—	—	—	—	—	—	—	—	—
3,5	99953	—	—	—	—	—	—	—	—	—
4,0	99994	—	—	—	—	—	—	—	—	—
5,0	99999	—	—	—	—	—	—	—	—	—

Куприенко Николай Владимирович
Пономарева Ольга Алексеевна
Тихонов Дмитрий Владимирович

СТАТИСТИКА. МЕТОДЫ АНАЛИЗА РАСПРЕДЕЛЕНИЙ. ВЫБОРОЧНОЕ НАБЛЮДЕНИЕ (С ИСПОЛЬЗОВАНИЕМ ППП STATISTICA)

Учебное пособие

Лицензия ЛР №020593 от 07.08.97
Налоговая льгота – Общероссийский классификатор продукции ОК 005 – 93, т.2; 95 3005 – учебная литература

Подписано в печать . . . Формат 60×84/16.
Печать офсетная. Усл. печ. л. . Уч.-изд. л. Тираж 500.
Заказ .

Отпечатано с готового оригинал-макета, предоставленного авторами,
в типографии Издательство Политехнического университета.
195251, Санкт-Петербург, Политехническая ул., 29