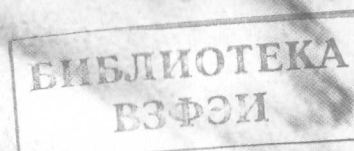


ФЕДЕРАЛЬНОЕ АГЕНТСТВО ПО ОБРАЗОВАНИЮ
ГОСУДАРСТВЕННОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ
ВЫСШЕГО ПРОФЕССИОНАЛЬНОГО ОБРАЗОВАНИЯ
ВСЕРОССИЙСКИЙ ЗАОЧНЫЙ ФИНАНСОВО-ЭКОНОМИЧЕСКИЙ
ИНСТИТУТ



СТАТИСТИКА

*Методические указания
по выполнению лабораторной работы*

Для студентов III курса всех специальностей
(первое и второе высшее образование)



МОСКВА 2005

ФЕДЕРАЛЬНОЕ АГЕНТСТВО ПО ОБРАЗОВАНИЮ

**ГОСУДАРСТВЕННОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ
ВЫСШЕГО ПРОФЕССИОНАЛЬНОГО ОБРАЗОВАНИЯ**

**ВСЕРОССИЙСКИЙ ЗАОЧНЫЙ
ФИНАНСОВО-ЭКОНОМИЧЕСКИЙ ИНСТИТУТ**

**НАУЧНЫЙ АБОНЕМЕНТ
ВЗФЭИ**

СТАТИСТИКА

**Методические указания
по выполнению лабораторной работы № 1
Автоматизированный априорный анализ
статистической совокупности в среде MS Excel**

**Для студентов III курса всех специальностей
(первое и второе высшее образование)**

**Москва
ВУЗОВСКИЙ УЧЕБНИК
2005**

Методические указания по выполнению лабораторной работы № 1
«Автоматизированный априорный анализ статистической совокупности в среде MS Excel» подготовили:

д-р физ. мат. наук, проф. *Г.П. Кожеевникова*,
канд. техн. наук, доцент *А.В. Голикова*,
канд. экон. наук, доцент *А.М. Каманина*,
преподаватель *А.М. Бобров*

Ответственный редактор *Г.П. Кожеевникова*

Методические указания по выполнению лабораторной работы № 1
«Автоматизированный априорный анализ статистической совокупности в среде MS Excel» одобрены на заседании
Научно-методического совета ВЗФЭИ

Проректор, председатель НМС, профессор *Д.М. Дайитбеков*

Статистика. Компьютерные лабораторные работы: Методические указания к лабораторной работе № 1 «Автоматизированный априорный анализ статистической совокупности в среде MS Excel». — М.: Вузовский учебник, 2005. — 72 с.

ББК 65.290-2

Лабораторная работа № 1

Автоматизированный априорный анализ статистической совокупности в среде MS Excel

I. ЦЕЛИ, СОДЕРЖАНИЕ И ОРГАНИЗАЦИЯ ВЫПОЛНЕНИЯ ЛАБОРАТОРНОЙ РАБОТЫ

1. ЦЕЛЬ И ЗАДАЧИ РАБОТЫ

Цель работы — освоение методики априорного статистического анализа структуры наблюдаемой совокупности с применением компьютерных средств экономико-статистических расчетов.

Априорный анализ исходного статистического материала (другие термины — «предварительный», «первичный», «разведочный») является важным этапом статистического исследования совокупности. На этом этапе решаются две основные задачи:

- критическая оценка исходных данных с точки зрения их достоверности и научной обоснованности;
- выявление типических черт и закономерностей, присущих наблюдаемому явлению в целом.

Всесторонняя и качественная оценка исходных данных на этапе априорного анализа — это предпосылка получения на последующих этапах статистического исследования научно обоснованных, логически выверенных статистических характеристик изучаемого явления.

Методы априорного анализа включают:

- обобщение исходных данных на основе построения вариационных рядов распределения по рассматриваемым признакам;
- оценку однородности исследуемой совокупности;
- анализ характера распределения единиц совокупности по изучаемым признакам.

Мощные и гибкие компьютерные средства для решения вышеперечисленных задач предоставляет программная надстройка **Пакет анализа** табличного процессора Microsoft Excel.

Изучение возможностей этого пакета и освоение технологии работы с ним при проведении априорного статистического анализа будет способствовать повышению уровня подготовки студентов в области компьютерного решения экономико-статистических задач.

2. КОМПЬЮТЕРНЫЕ СРЕДСТВА

Операционная среда:	Windows 2000/XP.
Программное средство:	Пакет программ для работы с электронными таблицами Microsoft Excel.
Версии MS Excel:	Excel 97, Excel 2000.
Технологическая среда:	Программная надстройка MS Excel Пакет анализа и библиотека из 78-ми статистических функций, встроенных в Excel.

3. ТРЕБОВАНИЯ К УРОВНЮ ПОДГОТОВЛЕННОСТИ СТУДЕНТА К ЛАБОРАТОРНОЙ РАБОТЕ

Лабораторная работа проводится после изучения студентами следующих тем курса статистики: *статистическое наблюдение, сводка и группировка статистических данных, ряды распределения, средние величины и показатели вариации, выборочный метод.*

Для выполнения работы студент *должен знать*:

- цель и содержание работы, порядок ее выполнения и отчетности (*раздел I «Методических указаний»*);
- статистическую сущность задач и методику априорного анализа статистических данных (*раздел II «Методических указаний»*);
- основные теоретические положения выполняемых заданий (*из раздела III «Методических указаний» пункты «Краткие теоретические сведения к заданиям»*).

Студент должен *обладать навыками* работы в среде Microsoft Excel:

- строить электронные таблицы;
- составлять и копировать расчетные формулы;
- использовать статистические и математические функции инструмента **Мастер функций**;
- строить статистические графики с использованием инструмента **Мастер диаграмм**.

Перед выполнением лабораторной работы студенту следует *ознакомиться с технологией выполнения каждого задания.*

4. СОДЕРЖАНИЕ И СТРУКТУРА ЛАБОРАТОРНОЙ РАБОТЫ

4.1. Постановка задачи

При проведении статистического наблюдения за деятельностью предприятий корпорации получены выборочные данные по 32-м предприятиям, выпускающим однородную продукцию (выборка 10%-ная, механическая), о среднегодовой стоимости основных производственных фондов и о выпуске продукции за год.

В проводимом статистическом исследовании обследованные предприятия выступают как единицы выборочной совокупности, а показатели *Среднегодовая стоимость основных производственных фондов* и *Выпуск продукции* — как изучаемые признаки единиц.

Для проведения автоматизированного статистического анализа совокупности выборочные данные представлены в формате электронных таблиц процессора Excel в диапазоне ячеек B4:C35. Для демонстрационного примера (ДП) выборочные данные приведены в табл. 1-ДП.

Таблица 1-ДП

ИСХОДНЫЕ ДАННЫЕ ДЕМОНСТРАЦИОННОГО ПРИМЕРА

	А	В	С
3	Номер единицы наблюдения	Среднегодовая стоимость основных производственных фондов, млн. руб.	Выпуск продукции, млн. руб.
4	1	94,00	110,00
5	2	107,00	101,00
6	3	134,00	120,00
7	4	157,00	81,00
8	5	163,00	80,00
9	6	167,00	114,00
10	7	173,00	161,00
11	8	173,00	90,00
12	9	177,00	178,00
13	10	179,00	107,00
14	11	200,00	125,00
15	12	201,00	108,00
16	13	205,00	133,00
17	14	208,00	124,00
18	15	212,00	201,00
19	16	213,00	161,00
20	17	214,00	151,00
21	18	216,00	169,00
22	19	218,00	149,00
23	20	230,00	180,00

Окончание табл. 1-ДП

	А	В	С
3	Номер единицы наблюдения	Среднегодовая стоимость основных производственных фондов, млн. руб.	Выпуск продукции, млн. руб.
24	21	234,00	148,00
25	22	237,00	162,00
26	23	241,00	166,00
27	24	248,00	168,00
28	25	45,00	224,00
29	26	276,00	171,00
30	27	290,00	191,00
31	28	298,00	220,00
32	29	167,00	114,00
33	30	205,00	133,00
34	31	330,00	53,00
35	32	260,00	224,00

В процессе исследования совокупности необходимо решить ряд статистических задач для *выборочной и генеральной совокупностей*.

Статистический анализ выборочной совокупности

1. Выявить наличие среди исходных данных резко выделяющихся значений признаков («выбросов» данных) с целью исключения из выборки аномальных единиц наблюдения.

2. Рассчитать обобщающие статистические показатели совокупности по изучаемым признакам: среднюю арифметическую (\bar{x}), моду (Mo), медиану (Me), размах вариации (R), дисперсию (σ_n^2), средние отклонения — линейное (d) и квадратическое (σ_n), коэффициент вариации (V_n), структурный коэффициент асимметрии К. Пирсона (As_n).

3. На основе рассчитанных показателей в предположении, что распределения единиц по обоим признакам близки к нормальному, оценить:

- а) степень колеблемости значений признаков в совокупности;
- б) степень однородности совокупности по изучаемым признакам;
- в) устойчивость индивидуальных значений признаков;
- г) количество попаданий индивидуальных значений признаков в диапазоны $(\bar{x} \pm \sigma)$, $(\bar{x} \pm 2\sigma)$, $(\bar{x} \pm 3\sigma)$.

4. Дать сравнительную характеристику распределений единиц совокупности по двум изучаемым признакам на основе анализа:

- а) вариации признаков;

- б) количественной однородности единиц;
- в) надежности (типичности) средних значений признаков;
- г) симметричности распределений в центральной части ряда.

5. Построить интервальный вариационный ряд и гистограмму распределения единиц совокупности по признаку *Среднегодовая стоимость основных производственных фондов* и установить характер (тип) этого распределения. Рассчитать моду Mo полученного интервального ряда и сравнить ее с показателем Mo несгруппированного ряда данных.

Статистический анализ генеральной совокупности

1. Рассчитать генеральную дисперсию σ_N^2 , генеральное среднее квадратическое отклонение σ_N и ожидаемый размах вариации признаков R_N . Сопоставить значения этих показателей для генеральной и выборочной дисперсий.

2. Для изучаемых признаков рассчитать:

- а) среднюю ошибку выборки;
- б) предельные ошибки выборки для уровней надежности $P = 0,683$, $P = 0,954$, $P = 0,997$ и границы, в которых будут находиться средние значения признака генеральной совокупности при заданных уровнях надежности.

3. Рассчитать коэффициенты асимметрии As и эксцесса Ek . На основе полученных оценок сделать вывод о степени близости распределения единиц генеральной совокупности к нормальному распределению.

4.2. Экономическое содержание задач статистического исследования

1. Типичны ли образующие выборку предприятия по значениям изучаемых экономических показателей?

- Построить точечный график-диаграмму рассеяния значений показателей.
- Выделить область диаграммы, отражающую группирование предприятий с близкими по значению показателями.
- Выявить предприятия с резко выделяющимися характеристиками и исключить их из проводимого статистического исследования вследствие нетипичности (аномальности) этих предприятий для изучаемой совокупности.

Аномальные значения показателей являются предметом отдельного исследования.

2. Каковы наиболее характерные для предприятий значения показателей среднегодовой стоимости основных фондов и выпуска продукции?

- Рассчитать среднюю арифметическую значений каждого из показателей, а также среднее квадратическое отклонение.

- Установить, какие предприятия входят в диапазон $(\bar{x} \pm \sigma)$, включающий предприятия с наиболее характерными значениями показателей.

3. Насколько сильны различия в экономических характеристиках предприятий выборочной совокупности? Можно ли утверждать, что выборка сформирована из предприятий с достаточно близкими значениями по каждому из показателей?

- Рассчитать коэффициенты вариации, используя вычисленные в п. 2 значения \bar{x} , σ .
- Установить (по значению коэффициента вариации), насколько предприятия однородны по изучаемым экономическим характеристикам.
- Определить максимальное расхождение в значениях показателей (размах вариации).

4. Какова структура предприятий выборочной совокупности по среднегодовой стоимости основных фондов? Каков удельный вес предприятий с наибольшими, наименьшими и типичными значениями данного показателя? Какие именно это предприятия?

- Произвести группировку (построить ряд распределения) предприятий по стоимости основных фондов.
- Найти модальный интервал и определить входящие в него предприятия (наиболее типичные).
- Установить, какие предприятия входят в группы с наименьшей и наибольшей стоимостью основных фондов.
- Определить удельный вес предприятий модального интервала и интервалов с наибольшими и наименьшими значениями показателя.

5. Носит ли распределение предприятий по группам закономерный характер и какие предприятия (с более высокой или более низкой стоимостью основных фондов) преобладают в совокупности?

- Построить гистограмму ряда распределения и визуально установить, имеется ли в распределении определенная закономерность.
- В случае, если распределение близко к нормальному, рассчитать показатель асимметрии, по знаку которого определить, доминируют ли в совокупности предприятия с более высокой или с более низкой стоимостью основных фондов.

6. Каковы ожидаемые средние величины среднегодовой стоимости основных фондов и выпуска продукции на предприятиях *корпорации в целом*? Какое максимальное расхождение в значениях показателя можно ожидать?

- Рассчитать предельную ошибку средней для каждого из показателей.

- Определить границы, в которых будут находиться средние значения показателей.
- Определить ожидаемый размах вариации показателей.

4.3. Структура лабораторной работы

Лабораторная работа выполняется по вариантам. Исходные данные своего варианта студент формирует самостоятельно в соответствии с алгоритмом, приведенным в разделе III — **Порядок выполнения лабораторной работы**.

Лабораторная работа состоит из трех этапов — подготовительного, расчетного и аналитического.

На *подготовительном этапе* формируется индивидуальная рабочая среда проведения вычислений по исходным данным варианта:

- осуществляется запуск MS Excel;
- в персональной папке студента подготавливаются два файла: **рабочий файл** с исходными данными варианта и макетами результирующих таблиц и **отчетный файл** с информацией, необходимой для составления отчета;
- согласовываются форматы чисел, используемые в компьютере и в тексте методических указаний к лабораторной работе;
- в случае отсутствия в меню Excel инструмента **Пакет анализа** производится его активизация.

На *расчетном этапе* вычисляются с применением инструментов **Пакет анализа** и **Мастер функций** обобщающие статистические показатели изучаемой совокупности. Этап включает выполнение следующих трех заданий:

Задание 1. Выявление и исключение из выборки аномальных единиц наблюдения.

Задание 2. Оценка описательных статистических параметров совокупности.

Задание 3. Построение и графическое изображение интервального вариационного ряда распределения единиц совокупности по признаку *Среднегодовая стоимость основных производственных фондов*.

Каждое задание имеет следующую структуру:

1. Краткие теоретические сведения.
2. Технология выполнения задания.
3. Алгоритмы выполнения задания.

Краткие теоретические сведения необходимы для понимания студентом статистической сущности задания.

В *технологической части* излагаются особенности применения инструментов **Пакет анализа**, **Мастер функций** и других средств Excel при автоматизированном решении статистических задач, указанных в заданиях.

В *алгоритмической части* представлены алгоритмы действий в среде Excel, выполнение которых реализует технологические процессы решения статистических задач.

На *заключительном, аналитическом, этапе* анализируются полученные обобщающие показатели изучаемой совокупности, интервальный вариационный ряд и его гистограмма, делаются выводы о статистических свойствах совокупности, выполняется их экономическая интерпретация применительно к деятельности изучаемой совокупности предприятий.

В методических указаниях к выполнению заданий используется 3 вида таблиц:

- результативные таблицы рассматриваемых показателей (макеты таблиц приведены в ПРИЛОЖЕНИИ 3);
- результативные таблицы демонстрационного примера «Методических указаний»;
- таблицы собственно «Методических указаний».

Во избежание коллизий при ссылке на различные виды таблиц к номерам таблиц второго и третьего вида добавляются соответственно идентификаторы «ДП» (*демонстрационный пример*) и «М» (*методические указания*).

5. ОТЧЕТНОСТЬ ПО РАБОТЕ

По результатам выполнения лабораторной работы студент представляет отчет. Отчет должен содержать:

- результативные таблицы с рассчитанными обобщающими показателями;
- рисунки статистических графиков;
- заключения о статистических свойствах изучаемой совокупности, сделанные на основе анализа таблиц и графиков;
- экономическую интерпретацию полученных статистических характеристик применительно к изучаемой совокупности предприятий.

Материалы отчета располагаются в следующем порядке:

1. Титульный лист (образец дан в ПРИЛОЖЕНИИ 1, электронная копия — в файле **Формат отчета.doc**).
 2. Постановка задачи статистического исследования, включая исходные данные варианта (п. 4.1, 4.2 раздела I «Методических указаний»; их электронная копия — в файле **Формат отчета.doc**).
 3. Распечатка рабочего файла с результативными таблицами и графиками (при копировании таблиц из Excel в Word следует предварительно снять цветную заливку заголовков таблиц Excel).
 4. Выводы о статистических свойствах изучаемой совокупности, сделанные на заключительном этапе работы на основе анализа полученных обобщающих показателей и статистических графиков.
- Выводы излагаются в текстовой форме в порядке, соответствующем перечню восьми задач п. 4.1 — **Постановка задачи**, и сопровож-

даются ссылками на соответствующие результативные таблицы и графики.

5. Экономическая интерпретация результатов статистического исследования предприятий.

При этом необходимо последовательно ответить на вопросы, поставленные в п. 4.2 — **Экономическая сущность задач статистического исследования**.

При подготовке отчета студент должен руководствоваться **Рекомендациями по составлению отчета** (раздел IV «Методических указаний»), устанавливающими формат изложения результатов проведенного статистического исследования. Титульный лист и постановка задачи статистического исследования (п. 4.1) заготовлены в файле **Формат отчета.doc** и *копируются в отчетный файл персональной папки студента на подготовительном этапе*.

Подготовка отчета производится вне рамок времени, отведенного на выполнение лабораторной работы. Отчет сдается преподавателю, ведущему лекционный курс.

Студент, не сдавший отчет по лабораторной работе, считается не выполнившим учебный план и к экзамену по статистике не допускается.

II. ТЕОРЕТИЧЕСКИЕ ОСНОВЫ И МЕТОДИКА ПРОВЕДЕНИЯ АПРИОРНОГО СТАТИСТИЧЕСКОГО АНАЛИЗА ДАННЫХ

1. СТАТИСТИЧЕСКАЯ СУЩНОСТЬ ЗАДАЧ АПРИОРНОГО АНАЛИЗА ДАННЫХ

1. **Анализ однородности совокупности.** При проведении статистического исследования основополагающим требованием к изучаемой совокупности является ее *однородность*.

Под однородностью статистической совокупности понимается ее качественная и количественная однородность. *Качественная однородность* означает, что единицы совокупности, характеризуясь множеством различных свойств и особенностей и участвуя в различных по своей природе процессах, в то же время в определенном отношении обладают *общими свойствами, общим законом развития* и в этом отношении являются *однородными, принадлежат одному определенному типу*. Именно в силу их сходства по признакам, характерным для данного типа, они и объединены в единую совокупность. *Количественная однородность* совокупности — это близость числовых значений признаков, определяющих качественное содержание совокупности.

Выбор меры близости зависит от решаемой статистической задачи и характера изучаемого признака.

Таким образом, однородность статистической совокупности означает, что все ее единицы обладают сходством по некоторому кругу признаков, обуславливающих качественную определенность совокупности, а количественные значения этих признаков оказываются близкими друг к другу.

Понятие однородности совокупности является относительным: *одна и та же совокупность может быть однородной по одним признакам и разнородной по другим.*

Однородность совокупности устанавливается в каждом конкретном статистическом исследовании в соответствии с его целями и задачами. *Необходимое условие однородности — однотипность единиц наблюдаемой совокупности* — устанавливается уже на первых шагах статистического исследования. Затем в рамках выполнимости условия однотипности наблюдаемых единиц проверяется их количественная однородность.

Выяснение степени однородности исходного статистического материала — одна из основных задач этапа априорного анализа данных, поскольку неоднородность первичных данных влечет за собой превращение многих статистических показателей в цифры, лишённые реального смысла и значения (типа средней зарплаты по г. Москве с ярко выраженной дифференциацией оплаты труда — от нескольких тысяч рублей до десятков и сотен тысяч долларов в месяц).

В случае выявления неоднородности изучаемой совокупности статистика прибегает к *методу группировки первичных данных* с целью образования однородных групп. Группировка дает возможность выделить из состава всех наблюдаемых случаев единицы разного качества, показать особенности явления при его развитии в различных условиях.

2. Выявление аномальных единиц наблюдения. С задачей анализа однородности совокупности тесно связана *задача выявления в совокупности аномальных наблюдений.*

Любая исследуемая совокупность может содержать единицы наблюдения, значения признаков которых резко выделяются из основной массы значений. Такие нетипичные значения признаков (*выбросы*) могут быть обусловлены воздействием каких-либо сугубо случайных обстоятельств, возникать в результате ошибок наблюдения или же быть объективно присущими наблюдаемому явлению. В любом случае они являются *аномальными для совокупности*, так как нарушают статистическую закономерность изучаемого явления. Следовательно, статистическое

изучение совокупности без предварительного выявления и анализа возможных аномальных наблюдений может не только исказить значения обобщающих показателей (средней, дисперсии, среднего квадратического отклонения и др.), но и привести к серьезным ошибкам в выводах о статистических свойствах совокупности, сделанных на основе полученных оценок показателей.

В случае выявления аномальных наблюдений правильность результатов анализа обеспечивается либо *исключением аномалий* из исходных данных вследствие их нетипичности для изучаемой выборки, либо *корректировкой их влияния* с помощью так называемой «подчистки данных», в основе которой лежат специальные методы робастного статистического оценивания (от англ. robust — устойчивый, крепкий).

В пользу первого варианта (исключения аномалий из выборки) говорит то, что многие выборочные данные существенно зависят от влияния условий наблюдения, т.е. являются неустойчивыми, и, как следствие, результаты анализа теряют статистическую точность. Второй вариант означает, что факт появления резко выделяющихся наблюдений («подозрительных» данных) *объективно имеет место*. Его следует проанализировать и понять причины возникновения аномальных выбросов. Для отдельных статистических задач наблюдаемые выбросы в данных могут быть даже более интересными для анализа, чем типичные данные. Тем более что необоснованное исключение «подозрительных» выбросов может привести к существенному искажению статистических оценок.

3. Построение и анализ вариационных рядов распределения. Статистические свойства, структура (внутреннее строение) и закономерности, присущие наблюдаемой совокупности, исследуются путем анализа *рядов распределения единиц совокупности по изучаемым признакам*. Такие ряды строятся на основе эмпирических данных (фактических данных наблюдения) и называются *эмпирическими распределениями*.

Эмпирические распределения представляются в виде дискретных или интервальных *вариационных рядов*.

В *дискретном вариационном ряду* варианты значений признака располагаются в порядке возрастания их величины, и для каждого варианта указывается его *частота* — число, показывающее, как часто у единиц совокупности встречается данный вариант значения признака.

Интервальный вариационный ряд представляет признак в виде упорядоченного набора интервалов значений признака с указанием для каждого интервала его частоты, фиксирующей

число попаданий значений признака в данный интервал. Количество интервалов и их величина выбираются таким образом, чтобы *наиболее отчетливо могла бы проявиться закономерность вариации признака*.

В силу своего построения интервальный вариационный ряд однозначно определяет разбиение множества единиц совокупности на однородные группы (выделенные в соответствии с выбранным способом группировки значений признака). При этом частоты интервалов дают представление о степени наполненности выделенных групп единицами совокупности, а характер изменения частот — о той или иной закономерности распределения единиц по группам.

Эмпирическое распределение может характеризоваться не только частотами, но и *частотами* вариационного ряда, представляющими собой частоты, выраженные в долях единицы или в процентах к итогу (соответственно сумма частостей равна 1 или 100%).

Распределяя единицы совокупности по группам, интервальный вариационный ряд показывает различия между ними по группировочному признаку, характеризует структуру совокупности по изучаемому признаку, позволяет судить о ее однородности по этому признаку, границах изменения однородности при переходе от одной группы к другой, степени колеблемости (вариации) значений признака, закономерности развития наблюдаемого явления.

Построение и статистическое изучение вариационных рядов распределения выполняется на этапе априорного анализа совокупности. При этом для каждого изучаемого признака строится вариационный ряд распределения единиц совокупности по данному признаку и рассчитываются обобщающие статистические характеристики ряда — средняя \bar{x} , мода Mo , медиана Me , показатели вариации признака R , d , σ , σ^2 , V_o и особенностей формы распределения As , Ek . На их основе оцениваются устойчивость индивидуальных значений признака x_i , надежность их среднего значения \bar{x} , степень вариации признака, устанавливается характер (тип) закономерности изменения частот в распределении и другие статистические свойства распределений.

Для визуального анализа эмпирических распределений используются их графические изображения — *полигоны, гistogramмы, кумуляты, огивы*. Графики дают наглядное и выразительное представление о характере изменения частот (частостей) в вариационном ряду, позволяя тем самым выявить тип закономерности распределения и выбрать для описания закономерности адекватную теоретическую модель.

4. **Определение типа закономерности распределения.** Важной характеристикой вариации признака в совокупности является *определенный порядок изменения частот ряда распределения* в зависимости от изменения величины признака. Например, с возрастанием признака частоты вначале также могут возрасти, а затем, достигнув в середине ряда своей максимальной величины, — уменьшаться по мере дальнейшего роста значений признака. Такой характер изменения частот прослеживается, в частности, в графике функции $f(x)$ на рис. 1.

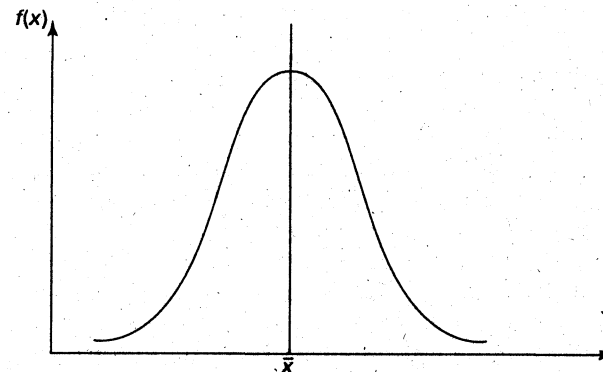


Рис. 1. Кривая нормального распределения

Четко выраженный *порядок изменения частот* в соответствии с изменением величины признака называют *закономерностью распределения*.

В характере (типе) закономерностей распределений отражены определенные условия и причины воздействующего на вариацию признака: сущность явления и те его свойства и условия протекания, которые определяют колеблемость значений признака.

Механизм формирования закономерности описывается в аналитической форме в виде функциональной (однозначной) зависимости $y = f(x)$ значений частот $f(x)$ от вариантов x . *Вид функции $f(x)$ определяет тип закономерности*.

Графическое изображение зависимости $y = f(x)$ (в виде непрерывной линии) называют *кривой распределения*, а форму этой кривой — *формой распределения*.

В статистической практике встречаются различные типы распределений, изображаемые графически в виде одновершинных и многовершинных кривых, имеющих симметричную или асимметричную форму с незначительной или существенной асимметрией, J-образную или U-образную форму и др.

Выявление закономерности распределения, определение *типа* этой *закономерности* и формы кривой распределения — одна из основных задач исследования вариационных рядов на этапе априорного анализа совокупности. Важность этой задачи обусловлена следующими факторами.

Знание типа закономерности распределения (а следовательно, и формы кривой) необходимо прежде всего для выяснения *типичности условий получения первичного статистического материала*. Поскольку всякое социально-экономическое явление в обычных для него условиях дает по определенному признаку *характерную, типичную* для него кривую распределения, то любое искажение формы кривой означает нарушение или изменение нормальных условий возникновения первичного статистического материала. Так, появление многовершинной или существенно асимметричной кривой говорит о разнотипном составе совокупности и о необходимости перегруппировки данных с целью выделения более однородных групп.

Знание типа закономерности распределения важно также для *обеспечения правильности выполнения практических расчетов и прогнозов*, поскольку многие из статистических методов и приемов требуют при их применении подчиненности единиц совокупности определенной закономерности распределения. Так, применение формулы Г. Стерджесса для расчета оптимального числа групп интервального ряда, правила «трех сигм», коэффициента вариации V в качестве индикатора однородности совокупности, метода наименьших квадратов при моделировании корреляционной связи явлений, методов дисперсионного анализа и других правомочно лишь в условиях нормального и близких к нему распределений.

Теоретический аппарат для определения и описания типа закономерности эмпирического распределения предоставляет математическая статистика.

Конкретный вид эмпирического распределения зависит от разных причин — как основных, так и второстепенных. *Основные причины* органически связаны с природой изучаемого явления, они *действуют систематически*, придают явлению регулярность, повторяемость и, следовательно, непосредственно формируют тот или иной тип закономерности распределения. *Второстепенные причины* являются *случайными* для данного явления, действуют эпизодически, хаотично, и их влияние проявляется в виде более или менее существенных отклонений от закономерного распределения. Воздействие на вариацию признака случайных причин затемняет основной, подлинный ха-

актер изменения частот в эмпирическом распределении, поэтому для определения типа закономерности (формы кривой) необходимо «отсеять» мешающие случайные факторы и тем самым *исключить из общей вариации признака ее случайную составляющую*.

Эта задача решается в статистике путем моделирования эмпирических распределений теоретическими (вероятностными) распределениями. *Теоретическое распределение* — это гипотетическое (предполагаемое) распределение вероятностей частот, которое получилось бы при *полном погашении всех случайных причин*, воздействующих на вариацию признака.

Как известно из математической статистики, всякое теоретическое распределение характеризуется тем или иным *законом распределения вероятностей*, устанавливающим функциональную зависимость $y = f(x)$ между значениями признака и их частотами. Графически такая зависимость выражается кривой определенной формы, называемой *формой теоретического распределения*.

Математическая статистика предоставляет для моделирования эмпирических распределений широкий спектр теоретических распределений — нормальное, логарифмически нормальное, биномиальное, Пуассона, Стьюдента, Пирсона, Фишера, Шарлье и др. Каждое распределение имеет свою специфику, свои условия возникновения и область применения. Исходя из этих факторов и выбирают для эмпирического распределения адекватную (т.е. наиболее подходящую) теоретическую модель, выраженную вероятностным законом распределения $y = f(x)$. При этом функция $f(x)$ выступает в качестве *приближенного описания закономерности эмпирического распределения*, определяя тем самым тип закономерности.

Если, например, моделью служит *нормальное распределение*, то эмпирическое распределение приближенно описывается законом нормального распределения случайной величины (законом Гаусса):

$$f(x) = \varphi(x, \bar{x}, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{x-\bar{x}}{\sigma^2}},$$

где $f(x)$ — плотность распределения случайной величины признака x ; \bar{x} и σ^2 — соответственно средняя арифметическая и дисперсия значений признака; π и e — математические константы ($\pi = 3,14$; $e = 2,72$).

Кривая нормального распределения (см. рис. 1) имеет од-
новершинную симметричную колоколообразную форму. Ее ле-

вая и правая ветви равномерно и симметрично убывают, асимптотически приближаясь к оси абсцисс. Высота кривой в точке x_i (частота появления варианта x_i), согласно закону Гаусса, зависит от значения x_i и параметров \bar{x} , σ^2 , причем чем больше отклонение варианта x_i от \bar{x} (величина σ), тем реже он появляется (меньше высота кривой в точке x_i). Максимальной высоте кривой соответствует совпадение центров распределения: $\bar{x} = M_0 = Me$. Форму описанной кривой называют *нормальной формой распределения*.

Таким образом, тип закономерности эмпирического распределения можно установить путем нахождения для него идеализированной теоретической (вероятностной) модели, в которой происходит освобождение данных от влияния случайных факторов и проявляется в максимально обобщенном виде характер вариации, закономерность изменения частот наблюдаемых единиц в зависимости от изменения признака. Кривая теоретической модели называется *теоретической кривой эмпирического распределения*.

Существуют различные способы нахождения теоретической модели эмпирического распределения. Прежде всего это *теоретический анализ изучаемого явления*. Например, известно, что распределение числа требований на обслуживание определенными услугами может быть изучено посредством распределения Пуассона; распределение людей по росту или по размеру обуви будет нормальным и т.д.

В ряде случаев тип закономерности можно установить по форме полигона или гистограмм. Форма полигона дает лишь первоначальное, ориентировочное представление о форме теоретической кривой. Однако с ростом числа наблюдений и одновременным сужением интервалов ряда число сторон полигона увеличивается, его зигзаги постепенно сглаживаются, и он все более и более приближается к некоторой плавной линии, а в пределе (*при бесконечно большом числе наблюдений и бесконечно малой ширине интервалов*) — принимает форму этой линии. *Гипотетическая кривая, выступающая в качестве предела полигона, и является искомой теоретической кривой эмпирического распределения*.

Аналогичным образом форму теоретической кривой можно определить и по виду гистограммы, используя предельный переход для сглаживания зубчатых столбцов гистограммы (рис. 2).

По форме гипотетической кривой устанавливается подходящий для ее описания закон распределения $y = f(x)$, который и определяет тип закономерности эмпирических данных.

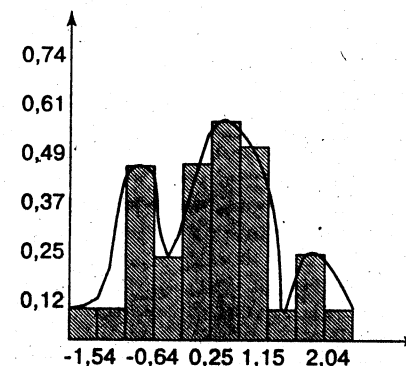


Рис. 2. Гистограмма и теоретическая кривая эмпирического распределения.

Однако во многих случаях ни теоретический анализ явления, ни непосредственное рассмотрение эмпирических данных, ни графики распределения не дают ответа на вопрос о типе закономерности распределения. Тогда обычно ведется исследование *на близость эмпирических данных к нормальному распределению*, поскольку при статистическом изучении социально-экономических явлений обычно формируются совокупности однородных единиц, *а для однородных совокупностей характерны распределения, которые по своему типу относятся к нормальным*.

Установлено, что в кривой нормального распределения выражается закономерность, возникающая при взаимодействии множества случайных факторов. Подчиненность явления нормальному закону распределения проявляется тем точнее, чем больше случайных причин влияют на величину признака, причем эти причины действуют независимо друг от друга. Если ни одна из причин по своему воздействию не окажется преобладающей над другими, то эмпирическое распределение очень близко подходит к нормальному (*в чистом виде нормальное распределение в социально-экономических явлениях практически не встречается*).

Предположение о характере того или иного эмпирического распределения — это гипотезы, а не категорические утверждения, поэтому они обычно подвергаются статистической проверке. Для этой цели используется ряд особых критериев, называемых *критериями согласия* (Пирсона, Романовского, Колмогорова и др.). Критерии согласия позволяют отвергнуть или подтвердить правильность выдвинутой гипотезы о характере распределения в эмпирическом ряду и дать ответ, можно ли принять для данного

эмпирического распределения модель, выраженную некоторым теоретическим законом распределения.

2. ОБЩАЯ МЕТОДИКА АПРИОРНОГО АНАЛИЗА СТАТИСТИЧЕСКИХ ДАННЫХ

1. Обобщение исходных данных:
 - построение интервальных вариационных рядов распределения по каждому из рассматриваемых признаков на базе предварительного определения целесообразного количества групп;
 - графическое изображение полученных рядов распределения в виде гистограммы, полигона, кумуляты, огивы.
2. Анализ однородности совокупности:
 - определение степени однородности всей совокупности в целом, а также однородности единиц в группах интервального ряда распределения;
 - выявление и анализ аномальных наблюдений.
3. Оценка характера распределения совокупности эмпирических данных:
 - вычисление и анализ показателей центра распределения и вариации признаков;
 - установление формы эмпирических распределений, полученных в виде вариационных рядов;
 - проверка гипотезы о форме распределения на базе одного из критериев согласия.

В лабораторной работе методика априорного анализа данных реализуется следующим образом.

1. Выявление и исключение из выборочной совокупности аномальных единиц наблюдения — Задание 1.
2. Расчет обобщающих статистических показателей выборочной совокупности непосредственно по первичным данным наблюдения (без построения вариационных рядов распределения) — Задание 2.
3. Оценка обобщающих параметров генеральной совокупности — Задание 2:
 - установление границ, в которых будет находиться генеральная средняя при заданных уровнях надежности;
 - расчет показателей вариации;
 - расчет показателей, характеризующих форму распределения.
4. Построение интервального ряда распределения, гистограммы и кумуляты — Задание 3.
5. Установление близости распределения эмпирических данных к нормальной форме — Заключительный этап.
6. Анализ однородности и других статистических свойств совокупности на основе полученных обобщающих показателей — Заключительный этап.

III. ПОРЯДОК ВЫПОЛНЕНИЯ ЛАБОРАТОРНОЙ РАБОТЫ

1. ПОДГОТОВИТЕЛЬНЫЙ ЭТАП

На данном этапе студент должен проделать следующие обязательные действия, связанные с организацией индивидуальной рабочей среды:

- запустить Excel и подготовить персональную папку с рабочим и отчетным файлами;
- сформировать индивидуальный вариант исходных данных и записать его в отчетный файл;
- проверить наличие в Excel надстройки **Пакет анализа**;
- проверить установку форматов чисел на компьютере.

1. Запуск Excel и подготовка персональной папки студента с рабочим и отчетным файлами

Для выполнения расчетов обобщающих показателей и подготовки отчета по лабораторной работе студент формирует персональную папку с именем **ФИО**, содержащую два файла: расчетный с именем **Лаб1.xls** и отчетный с именем **Отчет1.doc**.

Для создания рабочего файла проделать следующие действия:

1. Загрузить файл с исходными данными и макетами таблиц по следующему алгоритму:

1. На рабочем столе активизировать **Мой компьютер**;

2. В диалоговом окне выбрать файл

Datadisk на «primary» (E:)\Преподаватели\Кафедра статистики\Априорный анализ\stat_lab.xls.

2. Сохранить файл с исходными данными в качестве рабочего файла по алгоритму:

1. **Файл → Сохранить как...**;

2. В диалоговом окне **Сохранение документа** выбрать путь: **Datadisk на «primary» (E:)\Статистика\ Работы студентов\<Специальности>\<Номер группы>\<Ф.И.О.>;**

3. Сохранить файл в указанной папке под именем **Лаб1.xls**.

Для создания отчетного файла проделать следующие действия:

1. Загрузить файл **Формат отчета.doc** из директории

Datadisk на «primary» (E:)\Преподаватели\Кафедра статистики\Априорный анализ.

2. Сохранить файл по алгоритму:

1. **Файл → Сохранить как...**;

2. В диалоговом окне **Сохранение документа** выбрать путь: **Datadisk на «primary» (E:)\Статистика\ Работы студентов\<Специальности>\<Номер группы>\<Ф.И.О.>;**

3. Сохранить файл в указанной папке под именем **Отчет1.doc**.

ных производственных фондов, а по оси Y — соответствующие значения признака *Выпуск продукции*.

Обнаружение резко выделяющихся наблюдений производится визуально, путем выявления точек, отстоящих от основной массы точек на существенном расстоянии (для демонстрационного примера см. рис. 3).

Каждый «выброс» из основной массы точек означает аномальность единицы наблюдения либо по признаку X , либо по признаку Y . В обоих случаях такие единицы наблюдения (предприятия) подлежат удалению из первичных данных.

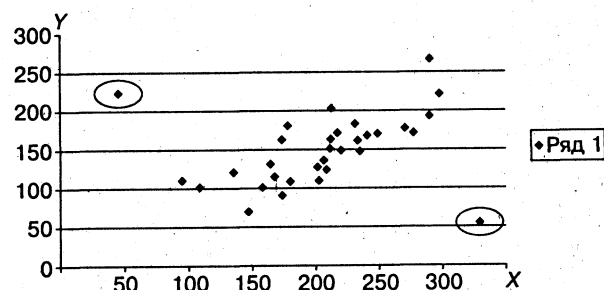


Рис. 3. Аномальные значения признаков на диаграмме рассеяния

Технология выполнения Задания 1

1. Построение диаграммы рассеяния изучаемых признаков

1. Построение диаграммы рассеяния в среде Excel осуществляется с помощью инструмента построения графиков **Мастер диаграмм**.

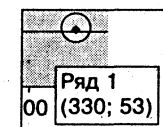
При построении точечного графика в режиме **Мастер диаграмм** данные первого выделенного столбца (*Стоимость основных производственных фондов*) автоматически сопоставляются оси X , данные второго выделенного столбца (*Выпуск продукции*) — оси Y .

Полученный график можно произвольно перемещать по полю рабочего листа, изменять его размеры. Для перемещения графика в удобное для просмотра место следует осуществить так называемый «захват мышью», т.е. установить курсор на произвольное место белой области графика, нажать левую кнопку мыши и, удерживая ее, переместить график в требуемое место, затем отпустить клавишу. Для изменения размеров графика производятся аналогичные действия, однако местом «захвата мышью» должен быть один из углов графика.

2. Для поиска аномальных наблюдений на построенной диаграмме рассеяния визуально находятся аномальные точки. При

подведении к ним курсора появляется надпись, содержащая значения признаков этого наблюдения в формате (X ; Y).

Для демонстрационного примера такая надпись выглядит следующим образом:



Обозначение (330; 53) означает, что выбранная аномальная точка соответствует наблюдению (предприятию), которое имеет среднегодовую стоимость основных фондов, равную 330 млн. руб., и выпуск продукции, равный 53 млн. руб.

3. Единица наблюдения, соответствующая выявленной аномальной точке, отыскивается в исходных данных табл. 1 визуально либо с помощью поисковых средств Excel.

4. Для фиксации выявленных аномальных единиц наблюдения в рабочем файле персональной папки студента выделена таблица (табл. 2), которая располагается в диапазоне ячеек A37–C41. Формат табл. 2 совпадает с форматом исходной табл. 1. **Перед исключением аномальных единиц из первичных данных информацию о них следует скопировать в табл. 2.**

Для демонстрационного примера табл. 2 имеет следующий вид.

Таблица 2-ДП

АНОМАЛЬНЫЕ ЕДИНИЦЫ НАБЛЮДЕНИЯ

Номер предприятия	Среднегодовая стоимость основных производственных фондов, млн. руб.	Выпуск продукции, млн. руб.
12	50,00	150,00
31	30,00	53,00

5. Для удаления из исходных данных табл. 1 строк с аномальными данными необходимо выделить мышью соответствующую **адресную строку вместе с ее номером!**

Для демонстрационного примера это адресная строка с номером 34, содержащая значения 31, 330 и 53:

	A	B	C	D	E
32	29	167	114		
33	30	205	133		
34	31	330	53		
35					

Алгоритм выполнения Задания 1

Этап 1. Построение диаграммы рассеяния изучаемых признаков

1. Выделить мышью оба столбца исходных данных в диапазоне B4:C35.

2. Вставка → Диаграмма → Точечная → Готово.

В результате выполнения этих действий на рабочем листе Excel появится диаграмма рассеяния.

Этап 2. Визуальный анализ диаграммы рассеяния, выявление и фиксация аномальных значений признаков, их удаление из первичных данных

1. Найти на графике точку, соответствующую аномальному наблюдению. Если таких точек нет, то перейти к действию 7, если есть — к действиям 2 — 6.

2. Подвести курсор к точке на диаграмме рассеяния, соответствующей аномальному наблюдению. После непродолжительного времени возле точки автоматически появится надпись, содержащая значения признаков этого наблюдения в формате (X, Y).

3. В исходных данных визуально (либо с помощью поисковых средств Excel) найти в табл. 1 строку, соответствующую выявленной аномальной единице наблюдения (предприятию). Скопировать эту строку в табл. 2.

4. Выделить мышью всю адресную строку с данными, подлежащими удалению.

5. Правка → Удалить.

6. Выполнять действия 1–5 до полного удаления всех аномальных наблюдений.

7. Переместить диаграмму рассеяния в область ячеек, начиная с ячейки F4.

Задание 2

Оценка описательных статистических параметров совокупности

Обобщающие статистические показатели совокупности исчисляются на основе анализа вариационных рядов распределения (см. п. 3 раздела II — «Теоретические основы лабораторной работы»). Однако пакет Excel позволяет рассчитать многие из этих показателей непосредственно по первичным данным наблюдения, используя инструмент **Описательная статистика** надстройки **Пакет анализа**, а также статистические функции инструмента **Мастер функций**.

Выполнение Задания 2 заключается в автоматизированном решении двух статистических задач:

1. Расчет описательных показателей выборочной и генеральной совокупностей по несгруппированным выборочным данным с использованием инструментов **Описательная статистика** и **Мастер функций**.

2. Оценка средней и предельной ошибок выборки для средней величины признака, а также границ, в которых эта средняя будет находиться в генеральной совокупности при заданных уровнях надежности.

Краткие теоретические сведения

1. Показатели описательной статистики

Описательная (дескриптивная) статистика является инструментом статистического описания данных, представляющих всю наблюдаемую совокупность в целом. Цель описательной статистики — получение сводных (обобщающих) показателей, характеризующих **исходную совокупность данных как генеральную** (а не как выборку из некоторой другой совокупности большего объема).

Для численной оценки обобщающих показателей совокупности используются так называемые **описательные статистики**, представляющие собой **однозначные функции** на множестве наблюдаемых данных, определяющие значения оцениваемых обобщающих показателей совокупности.

Описательные статистики рассчитываются **по несгруппированным данным** и реализуют **точные функциональные зависимости** значений показателей от исходных данных (в отличие от приближенных статистических оценок, выводимых с заданным уровнем надежности).

Показатели, вычисляемые с помощью описательных статистик (так называемые **описательные параметры**), можно разбить на 3 группы — показатели **положения вариантов значений признака, вариации признака и особенностей формы его распределения**.

1.1. Показатели положения описывают положение в первичном ряду данных тех или иных вариантов значений признака, характеризующих ряд. К ним относятся:

- максимальное x_{\max} и минимальное x_{\min} значения признака;
- средняя арифметическая величина \bar{x} (выступающая в качестве статистической оценки математического ожидания $M[\bar{x}]$ средней величины признака);
- мода Mo — наиболее часто встречающийся вариант значений признака или тот вариант, который соответствует максимальной ординате эмпирической кривой распределения;
- медиана Me — срединное значение ранжированного ряда вариантов значений признака;
- нижний и верхний квартили Q_1 и Q_3 , ограничивающие центральную зону ранжированного ряда, в которую попадают 50% вариантов значений признака: 25% вариантов значений, меньших срединного значения Me , и 25% вариантов значений, больших Me .

Среди показателей этой группы наиболее часто используются **показатели центра распределения** — \bar{x} , Mo и Me . При этом

\bar{x} рассчитывается для *первичного* ряда наблюдаемых данных, Mo и Me — для *ранжированного (упорядоченного)* ряда.

Для \bar{x} и Me характерны свойства:

$$\sum_{i=1}^n (x_i - \bar{x}) = 0, \quad \sum_{i=1}^n (x_i - Me) = \min. \quad (1)$$

В зависимости от целей статистического исследования в качестве центра распределения выбирается один из показателей \bar{x} , Mo или Me . В случае *однородной совокупности* (с нормальным или близким к нему распределением единиц) в качестве центра чаще всего используется средняя величина \bar{x} , характеризующая типичный уровень значений признака.

Для *неоднородной совокупности* (не поддающейся нормальному закону распределения) роль центра распределения обычно выполняет медиана Me .

1.2. Показатели вариации (колеблемости) признака описывают степень рассеяния вариантов значений признака относительно своего центра \bar{x} (или Me). Различают показатели размера и интенсивности вариации. К *показателям размера вариации* относятся:

- **размах вариации** $R = x_{\max} - x_{\min}$, устанавливающий предельное значение амплитуды колебаний признака;
- **межквартильный размах** $Q_3 - Q_1$, определяющий максимальную амплитуду колебаний в центральной зоне ряда (ограниченной квартилями Q_1 и Q_3);
- **среднее линейное отклонение** \bar{d} , вычисляемое как среднее арифметическое из абсолютных отклонений $|x_i - \bar{x}|$:

$$\bar{d} = \frac{\sum_{i=1}^n |x_i - \bar{x}|}{n}; \quad (2)$$

- **дисперсия** σ^2 (или D), рассчитываемая как среднее арифметическое из квадратов отклонений $(x_i - \bar{x})^2$:

$$\sigma^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}; \quad (3)$$

- **среднее квадратическое (стандартное) отклонение** σ , вычисляемое как корень квадратный из дисперсии σ^2 :

$$\sigma = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}}. \quad (4)$$

Интенсивность вариации признака измеряется относительными показателями

$$V_{\sigma} = \frac{\sigma}{\bar{x}}, \quad V_d = \frac{\bar{d}}{\bar{x}}, \quad V_R = \frac{R}{\bar{x}}, \quad V_{Me} = \frac{Me}{\bar{x}}.$$

Показатели R , \bar{d} и σ являются величинами именованными и выражаются в тех же единицах, что и изучаемый признак. Дисперсия σ^2 считается безразмерной величиной. Относительные показатели интенсивности вариации, как правило, измеряются в процентах.

В статистической практике для оценки вариации наиболее широко применяются показатели размера вариации σ^2 , σ и показатель интенсивности вариации V_{σ} .

Показатели σ^2 , σ , основанные на учете отклонений $(x_i - \bar{x})$ индивидуальных значений признака x_i от средней арифметической \bar{x} , являются *обобщающими характеристиками различия в значениях признака*.

Дисперсия σ^2 оценивает средний квадрат отклонений $(x_i - \bar{x})$. Величина σ очень чутко реагирует на вариацию признака (за счет возведения отклонений в квадрат) и органически вписывается в аппарат математической статистики (дисперсионный, корреляционный анализ и др.). На расчете дисперсии основаны многие статистические показатели.

Среднее квадратическое отклонение σ показывает, на сколько в среднем отклоняются индивидуальные значения признака x_i от их средней величины \bar{x} . Размерность отклонения σ совпадает с размерностью самого признака, поэтому этот показатель экономически хорошо интерпретируется. Отклонения, выраженные в σ , принято считать *стандартными*.

Интенсивность вариации обычно измеряют **коэффициентом вариации** V_{σ} , который выражается в процентах и вычисляется по формуле

$$V_{\sigma} = \frac{\sigma}{\bar{x}} \cdot 100. \quad (5)$$

Величина V_{σ} оценивает интенсивность колебаний вариантов относительно их средней величины. Принята следующая *оценочная шкала колеблемости* признака:

$$\begin{array}{ll} 0\% < V_{\sigma} \leq 40\% & \text{— колеблемость незначительная;} \\ 40\% < V_{\sigma} \leq 60\% & \text{— колеблемость средняя (умеренная);} \\ V_{\sigma} > 60\% & \text{— колеблемость значительная.} \end{array} \quad (6)$$

Для нормальных и близких к нормальному распределений показатель V_o служит индикатором однородности совокупности: принято считать, что при выполнении неравенства

$$V_o \leq 33\% \quad (7)$$

совокупность является количественно однородной по данному признаку.

Коэффициент вариации V_o часто используется для сравнения колеблемости признаков в различных рядах распределения, когда сравнивается вариация разных признаков в одной и той же совокупности или же вариация одного и того же признака в различных совокупностях, имеющих разные средние \bar{x} .

1.3. Показатели особенностей формы распределения. Для определения типа закономерности эмпирического распределения оно приближенно описывается подходящим теоретическим (вероятностным) распределением, форму кривой которого называют **формой распределения** (см. п. 4 раздела II — «Теоретические основы лабораторной работы»). В тех случаях, когда форма распределения анализируется на ее близость к нормальной форме, расхождение между ними оценивается показателями асимметрии и эксцесса.

Показатели асимметрии оценивают смещение ряда распределения влево или вправо по отношению к оси симметрии нормального распределения.

В симметричном распределении максимальная ордината прямой располагается точно в середине кривой (рис. 4), а соответствующие ей характеристики центра распределения совпадают:

$$\bar{x} = Mo = Me. \quad (8)$$

В случае асимметричного распределения вершина кривой находится не в середине, а сдвинута либо влево, либо вправо (рис. 4).

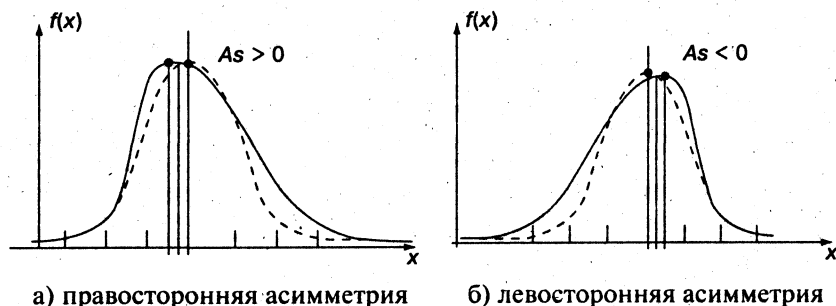


Рис. 4. Кривые асимметричных распределений (пунктиром обозначена нормальная кривая)

Если вершина сдвинута влево, то **правая часть кривой оказывается длиннее левой** (рис. 4, а), т.е. имеет место **правосторонняя асимметрия**, характеризующаяся неравенством

$$\bar{x} > Me > Mo, \quad (9)$$

что означает **преимущественное появление в распределении более высоких значений признака**.

Если же вершина кривой сдвинута вправо и **левая часть оказывается длиннее правой**, то **асимметрия левосторонняя** (рис. 4, б), для которой справедливо неравенство

$$\bar{x} < Me < Mo, \quad (10)$$

означающее, что в **распределении чаще встречаются более низкие значения признака**.

Чем больше величина расхождения между \bar{x} , Me , Mo , тем более асимметричен ряд. Разности $|\bar{x} - Me|$ и $|\bar{x} - Mo|$ являются простейшими показателями асимметрии в рядах распределения.

В нормальном и близких к нему распределениях основная масса единиц (почти 70%) располагается в центральной зоне ряда, в диапазоне $(\bar{x} \pm \sigma)$. Для оценки асимметричности распределения в этом центральном диапазоне служит коэффициент К. Пирсона:

$$As_n = \frac{\bar{x} - Mo}{\sigma}. \quad (11)$$

При правосторонней асимметрии $As_n > 0$, при левосторонней $As_n < 0$. Если $As_n = 0$, вариационный ряд симметричен.

Наиболее точным показателем асимметрии распределения является **коэффициент асимметрии As** , вычисляемый по формуле

$$As = \frac{\sum_{i=1}^n (x_i - \bar{x})^3}{\sigma^3 n}, \quad (12)$$

где n — число единиц совокупности. Как и в случае коэффициента Пирсона, при $As > 0$ имеет место правосторонняя асимметрия при $As < 0$ — левосторонняя. В симметричных распределениях $As = 0$.

Чем больше величина $|As|$, тем более асимметрично распределение. Установлена следующая **оценочная шкала асимметричности**:

$$\begin{array}{ll} |As| \leq 0,25 & \text{— асимметрия незначительная;} \\ 0,25 < |As| \leq 0,5 & \text{— асимметрия заметная (умеренная);} \\ |As| > 0,5 & \text{— асимметрия существенная.} \end{array} \quad (13)$$

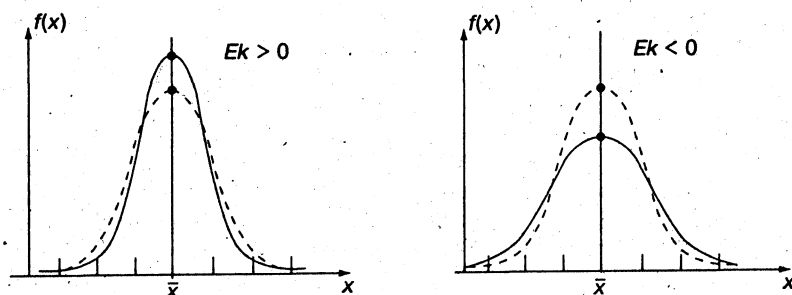
Поскольку коэффициенты As_n и As являются относительными безразмерными величинами, они часто применяются для

сравнительного анализа асимметричности различных рядов распределения.

Показатель эксцесса характеризует *крутизну* кривой распределения — ее *заостренность* или *пологость* по сравнению с нормальной кривой (рис. 5).

Для оценки расхождений в степени крутизны кривых (при одинаковой силе вариации) применяется **коэффициент эксцесса** E_k :

$$E_k = \frac{\sum_{i=1}^n (x_i - \bar{x})^4}{\sigma^4 n} - 3. \quad (14)$$



а) островершинное распределение б) плосковершинное распределение

Рис. 5. Кривые распределения с ненулевым эксцессом (пунктиром обозначена нормальная кривая)

Как правило, **коэффициент эксцесса вычисляется только для симметричных или близких к ним распределений**. Это объясняется тем, что за базу сравнения принята кривая нормального распределения, являющаяся симметричной. Относительно вершины нормальной кривой и определяется выпад вверх или вниз вершины теоретической кривой эмпирического распределения. При этом:

- если $E_k > 0$, то вершина кривой распределения располагается выше вершины нормальной кривой, а форма кривой является более островершинной, чем нормальная (рис. 5а). Это говорит о скоплении значений признака в центральной зоне ряда распределения, т.е. о **преимущественном появлении в данных значений, близких к средним**;
- если $E_k < 0$, то вершина кривой распределения лежит ниже вершины нормальной кривой, а форма кривой более пологая по сравнению с нормальной (рис. 5, б). Это означает, что значения признака не концентрируются в центральной части ряда, а **достаточно равномерно рассеяны по всему диапазону от x_{\max} до x_{\min}** .

Для нормального распределения $E_k = 0$, поэтому **чем больше абсолютная величина $|E_k|$, тем существеннее распределение отличается от нормального**. В частности, большая отрицательная величина E_k означает преобладание у признака крайних значений, причем одновременно и более низких, и более высоких. При этом в центральной части распределения может образоваться «впадина», превращающая распределение в двухвершинное (U-образной формы), что является **индикатором неоднородности совокупности**.

2. Оценка ошибок выборки

Применение выборочного метода наблюдения связано с измерением степени достоверности статистических характеристик генеральной совокупности, полученных по результатам выборочного наблюдения. Достоверность генеральных параметров зависит от **репрезентативности выборки**, т.е. от того, насколько полно и адекватно представлены в выборке статистические свойства генеральной совокупности.

Как правило, статистические характеристики выборочной и генеральной совокупностей не совпадают, а отклоняются на некоторую величину ε , которую называют **ошибкой выборки (ошибкой репрезентативности)**. Ошибка выборки — это разность между значением показателя, который был получен по выборке, и генеральным значением этого показателя. Например, разность

$$\varepsilon_x = |\bar{x} - \bar{X}|$$

определяет ошибку репрезентативности для средней величины признака.

Значения признаков единиц выборочной совокупности являются случайными величинами, следовательно, **ошибки выборки** также **случайны** и могут принимать для разных выборок одной и той же генеральной совокупности разные значения. Ввиду этого принято вычислять среднюю и предельную ошибки выборки.

Для среднего значения признака **средняя ошибка выборки** $\mu_{\bar{x}}$ (ее называют также **стандартной ошибкой**) выражает среднее квадратическое отклонение σ выборочной средней \bar{x} от математического ожидания $M[\bar{x}]$ генеральной средней \bar{X} . Величина ошибки $\mu_{\bar{x}}$ зависит от объема выборки n и от величины вариации признака σ : **чем больше n и меньше σ , тем меньше ошибка $\mu_{\bar{x}}$** .

Предельная ошибка выборки $\Delta_{\bar{x}}$ определяет границы, в пределах которых лежит генеральная средняя \bar{X} . Эти границы задают так называемый **доверительный интервал генеральной средней** \bar{x} — случайную область значений, которая с вероятностью P , близкой к 1, **гарантированно содержит** значение генеральной

средней. Эту вероятность называют *доверительной вероятностью* или *уровнем надежности*.

Наиболее часто используются уровни надежности $P = 0,954$; $P = 0,997$; $P = 0,683$.

В математической статистике доказано, что предельная ошибка выборки $\Delta_{\bar{x}}$ кратна средней ошибке $\mu_{\bar{x}}$ с коэффициентом кратности t , зависящим от значения доверительной вероятности P :

$$\Delta_{\bar{x}} = t \cdot \mu_{\bar{x}}.$$

Величина коэффициента t (называемого также *коэффициентом доверия*) является нормированным отклонением, которое вычисляется по формуле

$$t = \frac{|\bar{x} - \tilde{x}|}{\sigma}$$

и выражается не в натуральных единицах, а в сигмах: 1σ , 2σ , 3σ и т.д.

Значения t подсчитаны для различных уровней надежности P и протабулированы (хранятся в таблицах интегральной функции Лапласа). Для вышеприведенных уровней надежности P коэффициенты доверия t задаются следующим образом:

P	0,683	0,954	0,997
t	1	2	3

Например, если $t = 2$, то с вероятностью $P = 0,954$ можно утверждать, что расхождение между выборочной и генеральной средними $|\tilde{x} - \bar{x}|$ не превысит двукратной величины средней ошибки выборки:

$$\varepsilon_x = |\tilde{x} - \bar{x}| \leq 2\mu_{\bar{x}}.$$

Таким образом, предельная ошибка выборки Δ позволяет определить *предельные значения показателей генеральной совокупности и их доверительные интервалы*. Для генеральной средней предельные значения и доверительные интервалы определяются выражениями:

$$\begin{aligned} \bar{x} &= \tilde{x} \pm \Delta_{\bar{x}}, \\ \tilde{x} - \Delta_{\bar{x}} &\leq \bar{x} \leq \tilde{x} + \Delta_{\bar{x}}. \end{aligned} \quad (15)$$

Что касается величины *дисперсии генеральной совокупности* σ_N^2 , то она может быть оценена непосредственно по выборочной дисперсии σ_n^2 .

В математической статистике доказано, что при малом числе наблюдений (особенно при $n \leq 40-50$) для вычисления генеральной дисперсии σ_N^2 по выборочной дисперсии σ_n^2 следует использовать формулу

$$\sigma_N^2 = \frac{n}{n-1} \cdot \sigma_n^2. \quad (16)$$

При достаточно больших n значение поправочного коэффициента $\frac{n}{n-1}$ близко к 1 (при $n = 100$ его значение равно 1,01, а при $n = 500$ — 1,002 и т.д.). Поэтому при достаточно больших n можно приближенно считать, что обе дисперсии совпадают:

$$\sigma_N^2 \approx \sigma_n^2.$$

Технология выполнения Задания 2

1. Особенности реализации средств описательной статистики в надстройке ПАКЕТ АНАЛИЗА

1. В Пакете анализа инструмент *Описательная статистика* используется для генерации одномерного статистического отчета, который включает ряд *показателей положения, вариации и формы распределения признаков* выборочной и генеральной совокупностей, а также *среднюю и предельную ошибки выборки для средней* (рис. 6).

Столбец 1		Столбец 2	
Среднее		Среднее	
Стандартная ошибка		Стандартная ошибка	
Медиана		Медиана	
Мода		Мода	
Стандартное отклонение		Стандартное отклонение	
Дисперсия выборки		Дисперсия выборки	
Эксцесс		Эксцесс	
Асимметричность		Асимметричность	
Интервал		Интервал	
Минимум		Минимум	
Максимум		Максимум	
Сумма		Сумма	
Счет		Счет	
Уровень надежности (95,4%)		Уровень надежности (95,4%)	

Рис. 6. Макет результирующей таблицы инструмента ОПИСАТЕЛЬНАЯ СТАТИСТИКА

2. Между терминологией инструмента *Описательная статистика* и терминами, принятыми в отечественной статистике, имеется ряд расхождений. Согласование терминологии приводится в табл. 2-М.

Таблица 2-М

**СТАТИСТИЧЕСКАЯ ИНТЕРПРЕТАЦИЯ ПАРАМЕТРОВ
ОПИСАТЕЛЬНОЙ СТАТИСТИКИ**

Параметр инструмента ОПИСАТЕЛЬНАЯ СТАТИСТИКА	Статистический показатель	Обозначение
Среднее	Средняя арифметическая величина признака в выборке, вычисленная по несгруппированным данным	\bar{x}
Стандартная ошибка	Средняя ошибка выборки — среднее квадратическое отклонение выборочной средней \bar{x} от математического ожидания генеральной средней \bar{x}	$\mu_{\bar{x}}$
Медиана	Значение признака, приходящееся на середину ранжированного ряда выборочных данных	Me
Мода	Значение признака, повторяющееся в выборке с наибольшей частотой	Mo
Стандартное отклонение	Генеральное среднее квадратическое отклонение, оцененное по выборке	σ_N
Дисперсия выборки	Генеральная дисперсия, оцененная по выборке	σ_n^2
Экссесс	Коэффициент эксцесса, оценивающий по выборке значение эксцесса в генеральной совокупности	Ek_N
Асимметричность	Коэффициент асимметрии, оценивающий по выборке величину асимметрии в генеральной совокупности	As_N
Интервал	Размах вариации в выборке	R
Минимум	Минимальное значение признака в выборке	x_{\min}
Максимум	Максимальное значение признака в выборке	x_{\max}
Сумма	Суммарное значение элементов выборки	$\sum x_i$
Счет	Объем выборки	n
Уровень надежности (95,0%)	Предельная ошибка выборки, оцененная с заданным уровнем надежности	$\Delta_{\bar{x}}$

3. Вычисленные значения всех вышеперечисленных показателей представляются в *единой результирующей таблице* на рабочем листе Excel. При этом показатели могут рассчитываться сразу для нескольких рядов данных в соответствии с заданным входным диапазоном ячеек. Так, для входного диапазона с двумя рядами данных результирующая таблица будет состоять из двух столбцов значений описательных параметров (рис. 6). Именно такой формат имеет табл. 3, зарезервированная в рабочем файле

персональной папки студента для показателей, рассчитываемых в режиме **Описательная статистика**.

ВНИМАНИЕ!!! Расчет параметров в режиме **Описательная статистика** имеет ряд важных особенностей.

1. В качестве значений параметров
Стандартное отклонение,
Дисперсия выборки,
Экссесс,
Асимметричность.

Excel генерирует оценки соответствующих параметров для **генеральной совокупности**, а не для выборки.

2. Для применения **Описательной статистики** предварительное ранжирование исходных данных не требуется: при вычислении показателей ранжирование выполняется автоматически.

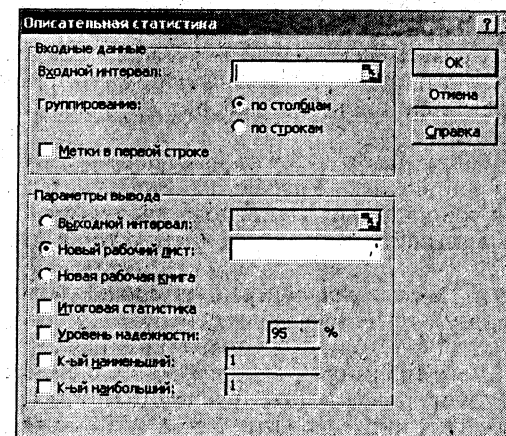
3. Появление в ячейке **Мода** индикатора ошибки **#Н/Д** указывает на то, что в анализируемых данных нет одинаковых значений признака. В этом случае в качестве моды **Мо** выбирается то значение признака, которое соответствует максимальной ординате теоретической кривой распределения (см. рис. 4).

4. Индикатор ошибки **#ДЕЛ/0!** в ячейке **Экссесс** и/или **Асимметричность** означает, что в результирующей таблице стандартное отклонение является нулевым или же заданный входной диапазон данных содержит менее четырех элементов данных.

**2. Задание управляющей информации
в диалоговом окне инструмента
ОПИСАТЕЛЬНАЯ СТАТИСТИКА**

Запуск инструмента **Описательная статистика** осуществляется следующей последовательностью действий:

Сервис → Анализ данных → Описательная статистика → ОК.



**Рис. 7. Диалоговое окно инструмента
ОПИСАТЕЛЬНАЯ СТАТИСТИКА**

В появившемся диалоговом окне инструмента (рис. 7) задаются следующие параметры:

1. Поле **Входной интервал** — вводится ссылка на диапазон ячеек, содержащих значения анализируемого признака. В качестве входного интервала может быть указан диапазон, который содержит ряды значений сразу нескольких анализируемых признаков. В таком случае показатели **Описательной статистики** будут рассчитаны для каждого ряда и представлены в единой таблице в виде отдельных столбцов (рис. 6).

2. Переключатель **Группирование: по столбцам/строкам** — устанавливается в положение **по столбцам** или **по строкам** в зависимости от того, в каком направлении располагаются анализируемые данные во входном диапазоне — вертикальном (по столбцам) или горизонтальном (по строкам).

3. Флажок **Метки в первой строке** — устанавливается в активное состояние, если первая строка во входном диапазоне содержит заголовки. Если заголовки отсутствуют, поле не активизируется. В этом случае будут автоматически созданы стандартные названия для данных выходного диапазона.

4. Поле **Выходной интервал** — вводится ссылка на ячейку заголовка первого столбца выходной результативной таблицы. Размер выходного диапазона ячеек определяется автоматически. В случае возможного наложения выходного диапазона на другие данные на экране появится соответствующее сообщение.

5. Переключатели **Новый рабочий лист** и **Новая рабочая книга** — устанавливаются в активное положение при необходимости открытия соответственно нового листа или новой книги. В новом листе результаты анализа располагаются начиная с ячейки A1, в новой книге — на первом листе начиная с ячейки A1.

6. Флажок **Итоговая статистика** — устанавливается в активное состояние, если для данных входного диапазона необходимо произвести расчет основных показателей, перечисленных в макете результативной таблицы на рис. 6.

7. Флажок **Уровень надежности** — устанавливается в активное состояние, если в результативную таблицу необходимо включить строку для оценки предельной ошибки выборки (Δ_x) с заданной доверительной вероятностью.

Значение уровня надежности *выражается в процентах* и задается в поле напротив флажка **Уровень надежности**. Уровень надежности **95,0%** (что равносильно доверительной вероятности $P = 0,95$ или же уровню значимости $\alpha = 0,05$) фиксируется в поле автоматически.

8. Флажки **К-тый наименьший** и **К-тый наибольший** — активизируются, если в результативную таблицу необходимо включить строку соответственно для k -того наименьшего (начиная с минимума x_{\min}) и k -того наибольшего (начиная с максимума x_{\max}) значений элементов в выборке. В этом случае в поле, расположенном напротив каждого флажка, вводится число k . При $k = 1$ выходные строки будут содержать соответственно x_{\min} и x_{\max} .

3. Расчет предельной ошибки выборки

Расчет *предельной ошибки выборки* Δ_x осуществляется в режиме **Описательная статистика**. Значение Δ_x для уровня надежности **95,0%** автоматически выводится в результативной таблице. Для расчета Δ_x при уровнях надежности **99,7** и **68,3%** необходимо *дважды* обратиться к инструменту **Описательная статистика** и в диалоговом окне активизировать поля **Входной интервал**, **Группирование**, **Выходной интервал**, **Уровень надежности**, внося в них соответствующие значения. Поскольку флажок **Итоговая статистика** при этом *не активизируется*, дублирования вычислений описательных показателей не происходит.

Для предельных ошибок выборки при уровнях надежности **68,3** и **99,7%** в рабочем файле персональной папки студента зарезервированы **таблицы 4а и 4б** соответственно, имеющие следующий формат:

Таблица 4

ПРЕДЕЛЬНАЯ ОШИБКА ВЫБОРКИ

Столбец 1		Столбец 2	
Уровень надежности (...%)		Уровень надежности (...%)	

4. Использование статистических функций инструмента МАСТЕР ФУНКЦИЙ

Помимо показателей, определяемых в режиме **Описательная статистика**, для анализа статистических свойств изучаемой совокупности необходимы дополнительно следующие выборочные показатели (которые либо вообще не вычисляются инструментом **Описательная статистика**, либо вычисляются для генеральной совокупности):

- выборочное среднее квадратическое отклонение σ_n ;
- выборочная дисперсия σ_n^2 ;
- выборочное среднее линейное отклонение \bar{d} ;
- коэффициент вариации признака в выборке V_o ;
- коэффициент асимметрии Пирсона As_n .

Для этих пяти показателей в рабочем файле персональной папки студента зарезервирована табл. 5 следующего формата.

Таблица 5

ВЫБОРОЧНЫЕ ПОКАЗАТЕЛИ СОВОКУПНОСТИ

Первый признак	Второй признак
Стандартное отклонение σ_n	Стандартное отклонение σ_n
Дисперсия σ_n^2	Дисперсия σ_n^2
Среднее линейное отклонение \bar{d}	Среднее линейное отклонение \bar{d}
Коэффициент вариации V_σ	Коэффициент вариации V_σ
Коэффициент асимметрии As_n	Коэффициент асимметрии As_n

Для вычисления показателей σ_n , σ_n^2 и \bar{d} используются соответствующие статистические функции инструмента **Мастер функций**:

1. **СТАНДОТКЛОНП (Диапазон ячеек)** — оценивает среднее квадратическое отклонение σ в предположении, что исходные данные представляют всю совокупность (не являются выборкой из генеральной совокупности).

2. **ДИСПР (Диапазон ячеек)** — оценивает дисперсию σ^2 в предположении, что исходные данные представляют всю совокупность.

3. **СРОТКЛ (Диапазон ячеек)** — рассчитывает для выборочных данных среднее линейное отклонение \bar{d} .

Технология расчета показателей в режиме МАСТЕР ФУНКЦИЙ:

1. Вызвать **Мастер функций**, нажав кнопку f_x на панели инструментов.

2. В **Мастере функций** среди предлагаемых категорий функций выбрать категорию **Статистические**, а затем — необходимую функцию (**СТАНДОТКЛОНП**, **ДИСПР** или **СРОТКЛ**).

3. В появившемся диалоговом окне в поле **Число 1** указать диапазон ячеек, содержащих значения анализируемого признака.

Для расчета коэффициента вариации V_σ по формуле (5) необходимо использовать среднее арифметическое значение \bar{x} из резульативной таблицы **Описательной статистики** (табл. 3) и выборочное среднее квадратическое отклонение σ_n из таблицы выборочных показателей (табл. 5).

Для расчета коэффициента асимметрии Пирсона As_n по формуле (11) необходимо использовать среднее арифметическое значение \bar{x} и моду Mo из резульативной таблицы **Описательной статистики** (табл. 3), а также выборочное среднее квадратическое отклонение σ_n из таблицы выборочных показателей (табл. 5).

5. Расположение данных на рабочем листе Excel

Исходные данные и их описательные статистические характеристики располагаются в таблицах рабочего файла персональной папки студента в соответствии с табл. 3-М. При этом учитывается следующая особенность автоматического построения выходных таблиц инструментом **Описательная статистика**.

Инструменты **Пакета анализа** строят выходные таблицы со **стандартными заголовками столбцов** (например, таблица описательных статистик на рис. 6 имеет стандартные заголовки *Столбец 1*, *Столбец 2*). При этом размещение выходных таблиц на рабочем листе осуществляется **начиная с ячейки заголовка первого столбца** (в примере это ячейка с заголовком *Столбец 1*). Именно адрес этой ячейки и должен задаваться в поле **Выходной интервал** диалогового окна при формировании выходных таблиц.

Таблица 3-М

РАСПОЛОЖЕНИЕ ДАННЫХ НА РАБОЧЕМ ЛИСТЕ EXCEL

Исходные данные и их статистические характеристики	Признак Среднегодовая стоимость основных производственных фондов, млн. руб.	Признак Выпуск продукции, млн. руб.
Исходные данные после удаления «аномальных» значений — Таблица 1	B4:B33	C4:C33
Описательные статистики — Таблица 3 с ячейкой заголовка первого столбца	A46	C46
Предельная ошибка выборки при $P = 0,683$ — Таблица 4а с ячейкой заголовка первого столбца	A67	C67
Предельная ошибка выборки при $P = 0,997$ — Таблица 4б с ячейкой заголовка первого столбца	A75	C75
Среднее квадратическое отклонение выборки σ_n в Таблице 5	B83	D83
Дисперсия выборки σ_n^2 в Таблице 5	B84	D84
Среднее линейное отклонение \bar{d} в Таблице 5	B85	D85
Коэффициент вариации V_σ в Таблице 5	B86	D86
Коэффициент асимметрии As_n в Таблице 5	B87	D87

В соответствии с приведенным расположением данных в рабочем файле персональной папки студента заготовлены макеты исходной и резульативных таблиц с соответствующими названиями (**ПРИЛОЖЕНИЕ 3**).

Алгоритмы выполнения Задания 2

Выполнение задания включает три этапа:

1. Расчет описательных параметров выборочной и генеральной совокупностей с использованием инструмента **Описательная статистика**.

2. Оценка предельных ошибок выборки для различных уровней надежности в режиме **Описательная статистика**.

3. Расчет описательных параметров выборочной совокупности с использованием инструмента **Мастер функций**.

Этап 1. Расчет описательных параметров выборочной и генеральной совокупностей с использованием инструмента **ОПИСАТЕЛЬНАЯ СТАТИСТИКА**

Алгоритм 1.1. Расчет описательных статистик

1. Сервис → Анализ данных → Описательная статистика → ОК;
2. **Входной интервал** ← диапазон ячеек таблицы, выделенный согласно табл. 3-М для значений признаков *Стоимость основных фондов* и *Выпуск продукции*;
3. **Группирование** → по столбцам;
4. **Итоговая статистика** — Активизировать;
5. **Уровень надежности** — Активизировать;
6. **Уровень надежности** ← 95,4;
7. **Выходной интервал** ← адрес ячейки заголовка первого столбца табл. 3 (см. табл. 3-М);
8. **ОК**;
9. При появлении окна с сообщением «*Выходной интервал накладывается на имеющиеся данные*» → **ОК**.

В результате указанных действий Excel осуществляет вывод таблицы описательных статистик в заданный диапазон рабочего файла (для демонстрационного примера это табл. 3-ДП). Интерпретация терминов таблицы в принятых статистических терминах приведена выше в табл. 2-М.

	A	B	C	D
43	Таблица 3-ДП			
44	Описательные статистики			
45	Среднегодовая стоимость основных производственных фондов, млн. руб.		Выпуск продукции, млн. руб	
46	Столбец 1		Столбец 2	
47				
48	Среднее	203,2333333	Среднее	144,6666667
49	Стандартная ошибка	8,804737927	Стандартная ошибка	7,071772174
50	Медиана	206,5	Медиана	148,5
51	Мода	167	Мода	114
52	Стандартное отклонение	48,22553575	Стандартное отклонение	38,73369141

53	Дисперсия выборки	2325,702299	Дисперсия выборки	1500,298851
54	Эксцесс	0,18281271	Эксцесс	-0,602481285
55	Асимметричность	-0,185105228	Асимметричность	0,218561586
56	Интервал	204	Интервал	144
57	Минимум	94	Минимум	80
58	Максимум	298	Максимум	224
59	Сумма	6097	Сумма	4340
60	Счет	30	Счет	30
61	Уровень надежности (95,4%)	18,356223	Уровень надежности (95,4%)	14,74331526

Этап 2. Оценка предельных ошибок выборки для различных уровней надежности в режиме **ОПИСАТЕЛЬНАЯ СТАТИСТИКА**

Алгоритм 2.1. Расчет предельной ошибки выборки при $P = 0,683$

1. Сервис → Анализ данных → Описательная статистика → ОК;
2. **Входной интервал** ← диапазон ячеек таблицы, выделенный согласно табл. 3-М для значений признаков *Стоимость основных фондов* и *Выпуск продукции*;
3. **Итоговая статистика** — Снять флажок;
4. **Уровень надежности** — Активизировать;
5. **Уровень надежности** ← 68,3;
6. **Выходной интервал** ← адрес ячейки, выделенной согласно табл. 3-М для предельной ошибки выборки при $P = 0,683$;
7. **ОК**;
8. При появлении окна с сообщением «*Выходной интервал накладывается на имеющиеся данные*» → **ОК**.

Алгоритм 2.2. Расчет предельной ошибки выборки при $P = 0,997$

1. Сервис → Анализ данных → Описательная статистика → ОК;
2. **Входной интервал** ← диапазон ячеек таблицы, выделенный согласно табл. 3-М для значений признаков *Стоимость основных фондов* и *Выпуск продукции* после удаления аномальных значений;
3. **Итоговая статистика** — Снять флажок;
4. **Уровень надежности** — Активизировать;
5. **Уровень надежности** ← 99,7;
6. **Выходной интервал** ← адрес ячейки, выделенной согласно табл. 3-М для предельной ошибки выборки при $P = 0,997$;
7. **ОК**;
8. При появлении окна с сообщением «*Выходной интервал накладывается на имеющиеся данные*» → **ОК**.

В результате работы алгоритмов 2.1 и 2.2 Excel выводит в соответствующие ячейки табл. 4 рабочего файла значения предельных ошибок выборки при $P = 0,683$ и $P = 0,997$ (для демонстрационного примера табл. 4а-ДП и табл. 4б-ДП).

	A	B	C	D
	Таблица 4а-ДП			
64	Предельные ошибки выборки			
65				
66	Среднегодовая стоимость основных производственных фондов, млн. руб.		Выпуск продукции, млн. руб	
67	Столбец 1		Столбец 2	
68				
69	Уровень надежности (68,3%)	8,965032289	Уровень надежности (68,3%)	7,200517087
70				
71				
72				Таблица 4б-ДП
73	Предельные ошибки выборки			
74	Среднегодовая стоимость основных производственных фондов, млн. руб.		Выпуск продукции, млн. руб	
75	Столбец 1		Столбец 2	
76				
77	Уровень надежности (99,7%)	28,52208432	Уровень надежности (99,7%)	22,90831186

Этап 3. Расчет описательных параметров выборочной совокупности с использованием инструмента МАСТЕР ФУНКЦИЙ

Алгоритм 3.1. Расчет выборочного стандартного отклонения σ_n для признака Среднегодовая стоимость основных производственных фондов

1. Установить курсор в ячейку, выделенную согласно табл. 3-М для среднего квадратического отклонения первого признака;
2. Вставка → Функция;
3. Статистические → СТАНДОТКЛОНП → ОК;
4. Число 1 ← диапазон ячеек табл. 1, содержащих значения первого признака.

Алгоритм 3.2. Расчет выборочного стандартного отклонения σ_n для признака Выпуск продукции

1. Установить курсор в ячейку, выделенную согласно табл. 3-М для среднего квадратического отклонения второго признака;
2. Вставка → Функция;
3. Статистические → СТАНДОТКЛОНП → ОК;
4. Число 1 ← диапазон ячеек табл. 1, содержащих значения второго признака.

Алгоритм 3.3. Расчет выборочной дисперсии σ_n^2 для признака Среднегодовая стоимость основных производственных фондов

1. Установить курсор в ячейку, выделенную согласно табл. 3-М для выборочной дисперсии первого признака;
2. Вставка → Функция;
3. Статистические → ДИСПР → ОК;
4. Число 1 ← диапазон ячеек табл. 1, содержащий значения первого признака.

Алгоритм 3.4. Расчет выборочной дисперсии σ_n^2 по признаку Выпуск продукции

1. Установить курсор в ячейку, выделенную согласно табл. 3-М для выборочной дисперсии второго признака;
2. Вставка → Функция;
3. Статистические → ДИСПР → ОК;
4. Число 1 ← диапазон ячеек табл. 1, содержащих значения второго признака.

Алгоритм 3.5. Расчет выборочного среднего линейного отклонения \bar{d} по признаку Среднегодовая стоимость основных производственных фондов

1. Установить курсор в ячейку, выделенную согласно табл. 3-М для среднего линейного отклонения первого признака;
2. Вставка → Функция;
3. Статистические → СРОТКЛ → ОК;
4. Число 1 ← диапазон ячеек табл. 1, содержащих значения первого признака.

Алгоритм 3.6. Расчет выборочного среднего линейного отклонения \bar{d} по признаку Выпуск продукции

1. Установить курсор в ячейку, выделенную согласно табл. 3-М для среднего линейного отклонения второго признака;
2. Вставка → Функция;
3. Статистические → СРОТКЛ → ОК;
4. Число 1 ← диапазон ячеек табл. 1, содержащих значения второго признака.

Алгоритм 3.7. Расчет коэффициента вариации V_σ по признаку Среднегодовая стоимость основных производственных фондов

1. Установить курсор в ячейку, выделенную согласно табл. 3-М для коэффициента вариации первого признака;
2. В активизированную ячейку ввести формулу = B83/B48 · 100.

Алгоритм 3.8. Расчет коэффициента вариации V_σ по признаку Выпуск продукции

1. Установить курсор в ячейку, выделенную согласно табл. 3-М для коэффициента вариации второго признака;
2. В активизированную ячейку ввести формулу = D83/D48 · 100.

Алгоритм 3.9. Расчет выборочного коэффициента асимметрии Пирсона As_n по признаку Среднегодовая стоимость основных производственных фондов

1. Установить курсор в ячейку, выделенную согласно табл. 3-М для коэффициента асимметрии первого признака;
2. В активизированную ячейку ввести формулу = (B48 – B51)/B83.

Алгоритм 3.10. Расчет выборочного коэффициента асимметрии Пирсона As_n по признаку Выпуск продукции

1. Установить курсор в ячейку, выделенную согласно табл.3 для коэффициента асимметрии второго признака;
 2. В активизированную ячейку ввести формулу = (D48 – D51)/D83.
- В результате работы алгоритмов 3.1 – 3.10 Excel осуществляет вывод значений выборочных параметров σ_n , σ_n^2 , \bar{d} , V_n и As_n в соответствующие ячейки рабочего листа (для демонстрационного примера табл. 5-ДП).

	A	B	C	D
80	Таблица 5-ДП			
81	Выборочные показатели вариации и асимметрии			
82	Среднегодовая стоимость основных производственных фондов, млн. руб.		Выпуск продукции, млн. руб	
83	Стандартное отклонение	47,41496482	Стандартное отклонение	38,08265864
84	Дисперсия	2248,178889	Дисперсия	1450,288889
85	Среднее линейное отклонение	36,66888889	Среднее линейное отклонение	32,35555556
86	Коэффициент вариации, %	23,33030908	Коэффициент вариации, %	26,32441841
87	Коэффициент асимметрии	0,764175055	Коэффициент асимметрии	0,805265907

Задание 3

Построение и графическое изображение интервального вариационного ряда распределения единиц совокупности по признаку Среднегодовая стоимость основных производственных фондов

Для того чтобы выявить структуру совокупности и тип закономерности распределения ее единиц по варьирующему признаку, строят и анализируют интервальный вариационный ряд

распределения и его гистограмму (см. п. 3 раздела II — Теоретические основы лабораторной работы).

Выполнение Задания 3 заключается в решении двух статистических задач:

1. Построение интервального ряда распределения единиц выборочной совокупности по признаку Среднегодовая стоимость основных производственных фондов.
2. Построение гистограммы и кумуляты сформированного интервального ряда.

Краткие теоретические сведения

Структурная группировка — это разделение совокупности единиц по одному или нескольким существенным признакам на однородные группы, различающиеся между собой в качественном и количественном отношении и позволяющие изучить структуру (внутреннее строение) совокупности. Простейшим видом структурной группировки являются **ряды распределения** — группировки, в которых для характеристики групп применяется лишь один признак — **численность группы**.

Обычно при построении вариационных рядов распределения ставится цель **количественной характеристики вариации**. При этом важное значение имеет правильный выбор **количества интервалов (групп) k и величины интервалов h** . Эти две величины должны выбираться таким образом, чтобы обеспечить выделение групп по принципу количественного сходства и различия единиц совокупности. Сходство единиц — это их количественная однородность в определенных пределах (внутри групп), различие — существенное расхождение значений признака в разных группах. Иными словами, величины k , h должны обеспечивать такое формирование интервалов, при котором **переход через границы каждого интервала означал бы переход от одной количественной особенности единиц к другой**, т.е. появление у единиц некоторого нового качества.

Правильный выбор границ интервалов важен также в связи с тем, что изменения в расстановке границ могут существенно повлиять на форму кривой распределения и тем самым привести к неправильным выводам о типе закономерности распределения.

В случае однородных совокупностей, помимо надлежащего выбора величин k и h , обычно выдвигаются еще два требования к строению интервального ряда. Первое требование — **равенство интервалов**. Оно обусловлено тем, что в однокачественных

совокупностях вариация признака проявляется в сравнительно узких пределах, а распределение его значений носит равномерный характер. Второе требование — *отсутствие групп с открытыми интервалами*, когда установлена лишь одна из границ крайних групп (это требование способствует точности статистических расчетов).

В условиях компьютерной обработки статистических данных обычно используют те или иные стандартные процедуры группировки по количественным признакам. Один из вариантов такого стандарта дает формула Г. Стерджесса для определения величины интервала:

$$k = 1 + 3,322 \lg n, \quad (17)$$

где n — число единиц совокупности. Величина интервала h определяется по формуле

$$h = \frac{x_{\max} - x_{\min}}{k}, \quad (18)$$

где x_{\max} и x_{\min} — соответственно максимальное и минимальное значения изучаемого признака. При дробном значении k берется ближайшее к нему целое.

Построение ряда распределения завершается подсчетом численности единиц в каждой группе — *частоты группы*. Иногда распределение характеризуют с помощью *накопленных частот* или же используют *частости* и *накопленные частости* (см. п. 3 раздела II — Теоретические основы лабораторной работы). Частости обычно применяют для небольших по объему совокупностей. Кроме того, они позволяют сравнивать распределения по одному и тому же признаку в разных по численности совокупностях.

Ниже приведено табличное представление интервального ряда распределения для демонстрационного примера.

ИНТЕРВАЛЬНЫЙ РЯД РАСПРЕДЕЛЕНИЯ ПРЕДПРИЯТИЙ ПО СРЕДНЕГОДОВОЙ СТОИМОСТИ ОСНОВНЫХ ПРОИЗВОДСТВЕННЫХ ФОНДОВ

Группы предприятий по среднегодовой стоимости основных фондов	Число предприятий в группе	Накопленная частость группы, %
94–134,8	3	10,00
134,8–175,6	6	30,00
175,6–216,4	11	66,67
216,4–257,2	6	86,67
257,2–298	4	100,00
Итого	30	

Относительно обобщающих показателей, рассчитываемых для интервального ряда данных, необходимо отметить, что они могут не совпадать с аналогичными показателями для несгруппированных данных; поскольку в расчетные формулы вместо фактических значений признака подставляются центральные значения интервалов. Вычисления будут точными лишь в тех случаях, когда групповые средние арифметические совпадают с центральными значениями интервалов, а это зависит от характера распределения значений признака в интервалах (группах).

Для наглядного представления интервальных рядов распределения используют их графическое изображение в виде *гистограммы* и *кумуляты*.

Гистограмма — столбиковая диаграмма, для построения которой на оси абсцисс откладывают отрезки, равные величине интервалов вариационного ряда. На отрезках строят прямоугольники, высота которых в принятом масштабе по оси ординат соответствует частотам (или частостям). Чем большее количество интервалов имеет ряд, тем более зазубренную будет гистограмма, чем меньшее — тем более «гладкой» она выйдет, однако при этом могут скрадываться характерные черты распределения и пропасть некоторые его подробности.

Для демонстрационного примера гистограмма и кумюлята приведенного выше интервального ряда распределения даны на рис. 8.

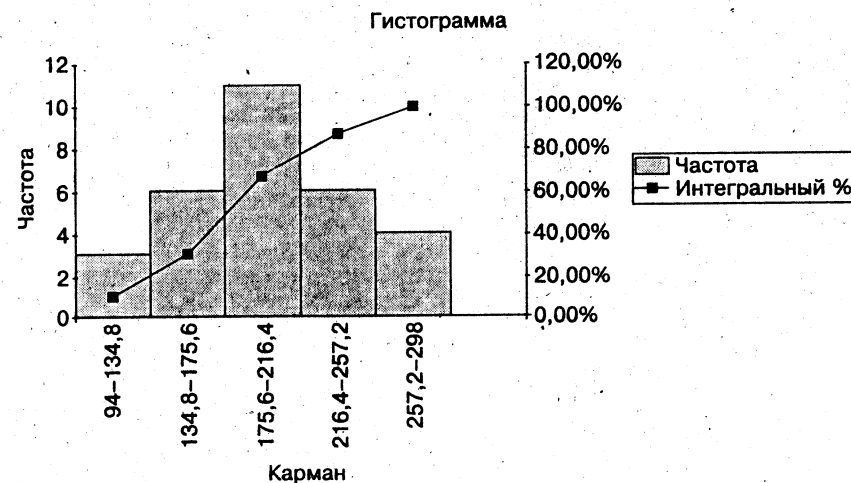


Рис. 8. Гистограмма и кумюлята интервального ряда распределения для демонстрационного примера

Форма гистограммы и ее характеристики несут наглядную информацию о распределении единиц совокупности по изучаемому признаку. Многие выводы и методы статистики базируются на использовании значений описательных параметров распределения, особенно параметров, характеризующих центр распределения (\bar{x} , Mo , Me) и рассеяние значений признака относительно центра (σ , As , Ek). Эти параметры можно не только количественно оценить по гистограмме, но и отобразить их визуально, «увидеть» их. По форме гистограммы можно установить и характер закономерности распределения, т.е. тип функции, описывающей распределение. Так, сопоставляя форму гистограммы на рис. 9 с наложенной на нее кривой нормального распределения, легко видеть существенное расхождение между эмпирическим и нормальным распределением. Форма гистограммы на рис. 8, напротив, дает основание предполагать, что распределение эмпирических данных близко к нормальному (наблюдается незначительная асимметрия).

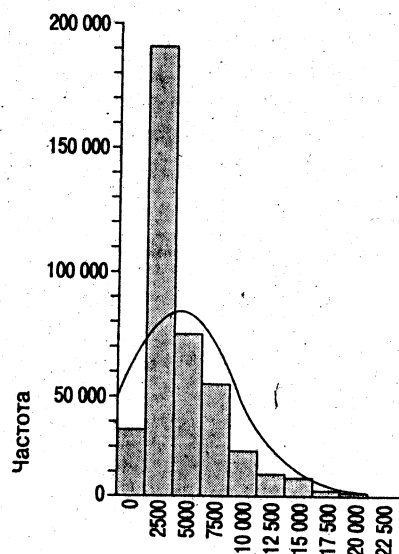


Рис. 9. Гистограмма распределения средней заработной платы по предприятиям региона, руб.

При построении графика гистограммы следует иметь в виду, что неудачный выбор масштаба графика приводит к тому, что гистограмма формируется либо в виде островершинной геометрической фигуры, либо плосковершинной, излишне растянутой по оси абсцисс. В обоих случаях в результате анализа фор-

мы гистограммы могут быть сделаны неправильные выводы о форме кривой распределения. Предотвратить эти недостатки позволяет *правило «золотого сечения»*, согласно которому *основание геометрической фигуры должно относиться к ее высоте как 1:0,62*. Для фигуры гистограммы на рис. 8 это правило соблюдается.

Технология выполнения Задания 3

Построение интервальных вариационных рядов распределения с использованием средств MS Excel можно осуществить различными способами — с помощью инструмента **Сводные таблицы** или же применяя инструмент **Гистограмма** надстройки **Пакет анализа**. В настоящем задании разбиение единиц совокупности на интервалы (группы) будет осуществляться с помощью средств инструмента **Гистограмма**.

1. Особенности построения интервальных вариационных рядов распределения средствами инструмента ГИСТОГРАММА

1. В надстройке Excel **Пакет анализа** инструмент **Гистограмма** используется для генерации интервального вариационного ряда с *равными по величине интервалами*, а также для построения гистограммы и кумуляты сформированного ряда распределения.

Инструмент **Гистограмма** производит следующие действия:

- рассчитывает число интервалов по формуле Г. Стерджесса (17);
- определяет величину интервала h по формуле

$$h = \frac{x_{\max} - x_{\min}}{\sqrt{k} - 1}, \quad (19)$$

несколько отличной от формулы (18);

- определяет нижние границы интервалов;
- формирует интервальный вариационный ряд в соответствии с величинами $\{k\}$, h ;
- рассчитывает частоты и накопленные частоты интервалов, определяя число попаданий данных в сформированные интервалы;
- строит столбиковую диаграмму частот (которая может быть преобразована в гистограмму) и кумуляту накопленных частостей для полученного ряда распределения;
- генерирует для вариационного ряда выходную таблицу в формате (20):

	Карман	Частота	Интегральный %
Еще	0		100%

2. Между терминологией генерируемой в режиме **Гистограмма** выходной таблицы и терминами, принятыми для вариационных рядов, имеются расхождения. Согласование терминологии приводится в табл. 4-М.

Таблица 4-М

СТАТИСТИЧЕСКАЯ ИНТЕРПРЕТАЦИЯ ТЕРМИНОЛОГИИ ИНСТРУМЕНТА ГИСТОГРАММА

Термин инструмента ГИСТОГРАММА	Термин, принятый в статистике
Карманы	Интервалы вариационного ряда
Интервал карманов	Диапазон ячеек, содержащих в возрастающем порядке верхние границы интервалов
Интегральный процент	Накопленная частость, выраженная в процентах

3. Инструмент **Гистограмма** имеет два режима работы:

- режим *автоматического формирования интервалов* вариационного ряда, имеющих равную величину h ;
- режим *формирования интервалов ряда в соответствии с границами, заданными пользователем*. Если при этом заданные интервалы будут не равны между собой, то в сгенерированной столбиковой диаграмме частоты попадания в интервал не будут связаны с размером интервала, что не позволит правильно оценить характер распределения единиц изучаемой совокупности.

2. Задание управляющей информации в диалоговом окне инструмента ГИСТОГРАММА

Запуск инструмента **Гистограмма** осуществляется следующей последовательностью действий:

Сервис → Анализ данных → Гистограмма → ОК.

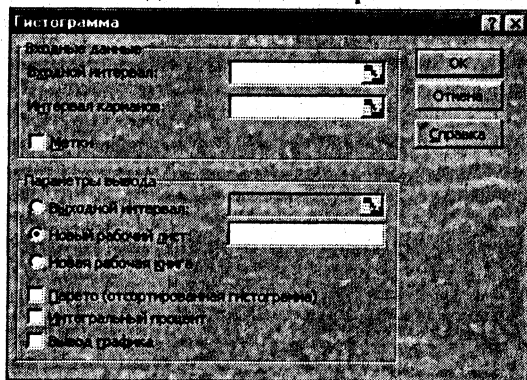


Рис. 10. Диалоговое окно инструмента Гистограмма

В появившемся диалоговом окне инструмента **Гистограмма** (рис. 10) задаются следующие параметры.

1. Поле **Входной интервал** — вводится ссылка на диапазон ячеек, содержащих значения анализируемого признака.

2. **Интервал карманов** (необязательный параметр) — вводится ссылка на диапазон ячеек, в которых задаются верхние границы интервалов. Если такой диапазон не указан, Excel осуществляет расчет нижних границ интервалов автоматически.

3. Флажок **Метки** не активизируется.

4. Поле **Выходной интервал** — вводится ссылка на ячейку заголовка первого столбца формируемой таблицы интервального вариационного ряда.

5. Переключатель **Новый рабочий лист/Новая рабочая книга** — открывает Новый рабочий лист/Новую рабочую книгу.

6. Флажок **Парето** (отсортированная гистограмма) — устанавливается в активное состояние при необходимости представить данные в порядке убывания частоты. Если флажок снят, то данные в выходном диапазоне будут приведены в порядке следования интервалов.

7. Флажок **Интегральный процент** — устанавливается в активное состояние, если необходимо рассчитать накопленные частоты (выраженные в процентах) и построить график кумуляты.

8. Флажок **Вывод графика** — устанавливается в активное состояние при необходимости автоматического построения столбиковой диаграммы.

3. Этапы построения интервального ряда распределения с использованием двух режимов инструмента ГИСТОГРАММА

Для построения интервального вариационного ряда необходимо в диалоговом окне инструмента **Гистограмма** задать *верхние границы* интервалов. Поскольку они неизвестны, для их нахождения можно воспользоваться *режимом автоматического расчета границ интервалов* (п. 1.3). Однако в этом режиме в карманах генерируемой таблицы выдаются *нижние границы* интервалов. Для получения на их основе верхних границ необходимо выполнить два действия:

1) исключить из карманов сгенерированной таблицы нижнюю границу первого интервала x_{\min} ;

2) добавить карман с верхней границей x_{\max} последнего интервала (значение x_{\max} имеется в табл. 3 — Описательные статистики).

Для демонстрационного примера такое преобразование приведено на рис. 11.

Таблица 6		Преобразуется в	Таблица 6	
90	Карман		90	Карман
91	94,0		91	
92	134,8		92	134,8
93	175,6		93	175,6
94	216,4		94	216,4
95	257,2		95	257,2
96	Еще		96	298

а) первичная

б) итоговая

Рис. 11. Схема перехода от нижних границ интервалов к верхним

Построение интервального ряда с использованием описанного приема перехода от нижних границ к верхним осуществляется в два этапа.

Этап 1. Применяется режим автоматического расчета границ интервалов. В этом режиме рассчитываются нижние границы интервалов и формируется *Таблица 6*, рассматриваемая в качестве *промежуточной* (ее формат приведен на рис. 11, а). Помимо нижних границ в табл. 6 автоматически генерируется дополнительный карман с именем «Еще», который зарезервирован для последующего задания верхней границы последнего интервала (значения x_{\max}).

В сгенерированной табл. 6 выполняется переход от нижних границ к верхним (с исключением x_{\min} и добавлением x_{\max}). Преобразование выполняется вручную в соответствии со схемой на рис. 11. В результате промежуточная таблица переходит от своего *первичного* вида (с нижними границами) к *итоговому* (с верхними границами), в котором диапазон карманов смещен на одну ячейку вниз.

Поскольку на первом этапе кроме границ интервалов не рассчитываются никакие другие характеристики ряда и должна быть сгенерирована только табл. 6, то на этом этапе в диалоговом окне *Гистограмма* необходимо задать только два параметра, обеспечивающих генерацию табл. 6, — *Входной интервал*, *Выходной интервал* и не активизировать остальные параметры окна.

Этап 2. Применяется режим формирования ряда по заданным верхним границам интервалов (п. 1.3). При этом в соответствии с карманами итоговой промежуточной табл. 6 генерируется в формате (20) *выходная Таблица 7* интервального вариационного ряда и строится столбчатая диаграмма с нанесенной на ней кумулятой.

Для выполнения этого этапа необходимо вновь обратиться к инструменту *Гистограмма*, задав в диалоговом окне следующие параметры: *Входной интервал*, *Интервал карманов*, *Выходной интервал*, *Интегральный процент*, *Вывод графика*.

Выходная таблица содержит:

- границы интервалов — в графе с именем «Карман»;
- частоты интервалов — во второй графе;
- накопленные частоты — в графе с именем «Интегральный %».

Для демонстрационного примера выходная таблица имеет следующий вид:

	А	В	С
99	Таблица 7-ДП		
100	Интервальный ряд распределения предприятий по стоимости основных производственных фондов		
101	Карман	Частота	Интегральный %
102	134,8	3	10,00
103	175,6	6	30,00
104	216,4	11	66,67
104	257,2	6	86,67
106	298	4	100,00
107	Еще	0	100

4. Приведение выходной таблицы и диаграммы к виду, принятому в статистике

1. Сгенерированную *Гистограммой* выходную таблицу (табл. 7) следует привести к виду, принятому в статистике, учитывая согласование терминологии, данное в табл. 4-М. Для демонстрационного примера Excel-формат результативной таблицы выглядит следующим образом:

	А	В	С
99	Таблица ДП		
100	Интервальный ряд распределения предприятий по стоимости основных производственных фондов		
101	Группы предприятий по среднегодовой стоимости основных фондов	Число предприятий в группе	Накопленная частотность группы, %
102	94 — 134,8	3	10,005
103	134,8 — 175,6	6	30,00
104	175,6 — 216,4	11	66,67
104	216,4 — 257,2	6	86,67
106	257,2 — 298	4	100,00
107			
108	Итого	30	

Для перехода от табл. 7 к результативной таблице необходимо провести ручные следующие преобразования:

- заменить названия столбцов;
- удалить строку «Еще»;
- границы интервалов привести к формату «нижняя граница — верхняя граница» (для первого интервала нижняя граница — это x_{\min} из табл. 3 — *Описательные статистики*);
- добавить и заполнить итоговую строку.

2. В качестве выходного графика интервального ряда инструмент *Гистограмма* строит *столбчатую диаграмму* с нанесенной на ее поле кумулятой. Для преобразования *столбчатой диаграммы* в *гистограмму* необходимо уменьшить ширину зазора между столбцами до 0, воспользовавшись соответствующим средством инструмента *Мастер диаграмм*.

Полученную гистограмму следует переместить в более удобное для анализа место на рабочем листе Excel, расположив ее вслед за результативной таблицей.

Масштаб графика гистограммы должен отвечать правилу «золотого сечения», для чего *ширина и высота гистограммы устанавливается в пропорции 1:0,62*.

Для демонстрационного примера гистограмма приведена на рис. 8.

5. Расположение данных на рабочем листе Excel

Исходные данные и рассчитываемые характеристики ряда распределения располагаются в четырех таблицах рабочего файла персональной папки студента в соответствии с табл. 5-М.

Таблица 5-М

РАСПОЛОЖЕНИЕ ДАННЫХ НА РАБОЧЕМ ЛИСТЕ EXCEL

Исходные данные и их статистические характеристики	Адресация
Значения первого признака — в столбце первого признака из Таблицы 1	B4:B33
Промежуточная таблица с нижними границами интервалов — первичная Таблица 6. Указывается адрес заголовка первого столбца	A90
Промежуточная таблица с верхними границами интервалов — итоговая Таблица 6	A92:A96
Выходная таблица — Таблица 7. Указывается адрес заголовка первого столбца	A101
Гистограмма и кумулята вариационного ряда — с верхней левой ячейкой	A112

Алгоритмы выполнения Задания 3

Выполнение задания осуществляется в три этапа:

1. Построение промежуточной таблицы.
2. Генерация выходной таблицы и графиков.
3. Приведение выходной таблицы и диаграммы к виду, принятому в статистике.

Этап 1. Построение промежуточной таблицы

Алгоритм 1.1. Расчет нижних границ интервалов

1. Сервис → Анализ данных → Гистограмма → ОК;
2. Вводной интервал ← диапазон ячеек, выделенный согласно табл. 5-М для столбца значений первого признака;
ВНИМАНИЕ !!! Здесь возможен ошибочный захват мышью столбца второго признака. Необходимо проконтролировать правильность задания входных данных!

3. Интервал карманов оставить незаполненным;
4. Выходной интервал ← адрес заголовка первого столбца первичной промежуточной табл. 6 (см. табл. 5-М).
5. ОК.

Алгоритм 1.2. Переход от нижних границ к верхним

1. Выделить курсором верхнюю левую ячейку табл. 6 и нажать клавишу [Delete];
2. Ввести в ячейку с именем «Еще» значение x_{\max} первого признака из табл. 3 — Описательные статистики.

Этап 2. Генерация выходной таблицы и графиков

Алгоритм 2.1. Построение выходной таблицы, столбиковой диаграммы и кумуляты

1. Сервис → Анализ данных → Гистограмма → ОК;
2. Вводной интервал ← диапазон ячеек, выделенный согласно табл. 5-М для столбца значений первого признака;
ВНИМАНИЕ !!! Здесь возможен ошибочный захват мышью столбца второго признака. Необходимо проконтролировать правильность задания входных данных!
3. Интервал карманов ← диапазон карманов итоговой промежуточной табл. 6 с верхними границами (см. табл. 5-М);
4. Выходной интервал ← адрес заголовка первого столбца выходной табл. 7 (см. табл. 5-М);
5. Интегральный процент — Активизировать;
6. Вывод графика — Активизировать;
7. ОК;
8. При появлении сообщения о наложении данных — ОК.

Этап 3. Приведение выходной таблицы и диаграммы к виду, принятому в статистике

Алгоритм 3.1. Преобразование выходной таблицы в результативную

1. Заменить названия столбцов выходной табл. 7 в соответствии с табл. 6-М;

Таблица 6-М

Название столбца в выходной таблице	Название столбца в результативной таблице
Карман	Группы предприятий по стоимости основных фондов
Частота	Число предприятий в группе
Интегральный %	Накопленная частота группы

2. Строки первого столбца привести к виду «нижняя граница интервала — верхняя граница интервала», учитывая совпадение верхних границ предыдущего интервала с нижней границей последующего интервала;
3. Строку с именем «Еще» выделить мышью и очистить, нажав клавишу [Delete];
4. Добавить и заполнить строку с именем «Итого».

Алгоритм 3.2. Преобразование столбиковой диаграммы в гистограмму

1. Осуществив «захват мышью», переместить график, расположив его вслед за табл. 7 согласно адресации, указанной в табл. 5-М;
2. Исключить зазоры, выполнив следующие действия:
 - 2.1. Нажать правую кнопку мыши на одном из столбиков диаграммы;
 - 2.2. Формат рядов данных → Параметры;
 - 2.3. Ширина зазора ← 0;
 - 2.4. ОК;
3. Используя «захват мышью» за угол поля графика, установить соотношение ширины и высоты фигуры гистограммы в пропорции 1:0,62.

ВНИМАНИЕ!!! Здесь возможна ошибочная установка указанной пропорции для размеров поля графика, а не для самой геометрической фигуры гистограммы. Необходимо проконтролировать правильность установки пропорции ширины и высоты фигуры гистограммы!

3. ЗАКЛЮЧИТЕЛЬНЫЙ ЭТАП

1. Анализ обобщающих показателей описательной статистики

На основании рассчитанных значений показателей описательной статистики можно не только получить информацию о средних величинах, степени вариации и особенностях формы распределения единиц совокупности, но и сделать заключения о других статистических характеристиках и свойствах совокупности, о внутренней связи между единицами совокупности.

1.1. **Степень колеблемости признака** определяется по значению коэффициента вариации V_o , исходя из оценочной шкалы (6). Показатель V_o используется и для сравнительной оценки вариации в различных рядах распределений.

1.2. **Однородность совокупности** для нормального и близких к нормальному распределений устанавливается по условию (7). Чем однороднее изучаемая совокупность, тем надежнее полученная средняя \bar{x} .

1.3. Для оценки **надежности (типичности) средней величины** \bar{x} можно воспользоваться значением показателя вариации V_o . Если его значение невелико, то индивидуальные значения признака x_i мало отличаются друг от друга, единицы наблюдения количественно однородны и, следовательно, средняя арифметическая величина \bar{x} является надежной характеристикой данной совокупности. Если же оценка V_o достаточно высока (превышает 40%), т.е. наблюдается значительное расхождение между значениями x_i , то средняя \bar{x} будет ненадежной характеристикой совокупности и ее практическое применение становится проблематичным.

1.4. Сопоставление средних отклонений — квадратического σ и линейного \bar{d} позволяет сделать вывод об **устойчивости индивидуальных значений признака**, т.е. об отсутствии среди них «аномальных» вариантов значений.

В условиях симметричного и нормального, а также близких к ним распределений между показателями σ и \bar{d} имеют место равенства

$$\sigma \approx 1,25\bar{d}, \bar{d} \approx 0,8\sigma,$$

поэтому отношение показателей \bar{d} и σ может служить **индикатором устойчивости данных**: если

$$\frac{\bar{d}}{\sigma} > 0,8, \quad (21)$$

то значения признака неустойчивы, в них имеются «аномальные» выбросы. Следовательно, несмотря на визуальное обнаружение и исключение нетипичных единиц наблюдений при выполнении Задания 1, некоторые аномалии в первичных данных продолжают сохраняться. В этом случае их следует выявить (например, путем поиска значений, выходящих за границы $(\bar{x} \pm 2\sigma)$) и рассматривать в качестве возможных «кандидатов» на исключение из выборки.

1.5. По значениям показателей \bar{x} и σ можно определить **границы диапазонов рассеяния значений признака** относительно средней \bar{x} , т.е. установить, какая доля значений признака попадает в тот или иной диапазон отклонений от \bar{x} .

Согласно вероятностной теореме П.Л. Чебышева, следует ожидать, что **независимо от формы распределения** 75% значений признака будут находиться в диапазоне $(\bar{x} \pm 2\sigma)$, а 89% значений — в диапазоне $(\bar{x} \pm 3\sigma)$.

В **нормально распределенных и близких к ним рядах** вероятностные оценки диапазонов рассеяния значений признака таковы:

$$\begin{aligned} 68,3\% \text{ войдет в диапазон } & (\bar{x} \pm \sigma); \\ 95,4\% \text{ попадет в диапазон } & (\bar{x} \pm 2\sigma); \\ 99,7\% \text{ появится в диапазоне } & (\bar{x} \pm 3\sigma). \end{aligned} \quad (22)$$

Соотношение (22) известно как **правило «трех сигм»**.

Для выборочной совокупности значения \bar{x} и σ_n рассчитаны (табл. 3 и 5) и являются точными, поэтому, основываясь на правиле «трех сигм», можно точно оценить границы всех трех диапазонов рассеяния признака и определить, сколько значений x_i попадает в каждый из диапазонов.

В случае генеральной совокупности точно известна только величина σ_n (табл. 3), а для средней \bar{x} рассчитаны лишь предельные ошибки выборки (табл. 3 и 4), поэтому для генеральной совокупности оценки рассеяния значений признака по трем диапазонам являются **прогнозными** и обычно задаются в форме (22) с конкретным числовым значением параметра σ_n .

1.6. Учитывая правило «трех сигм», в статистической практике величину 3σ считают в условиях нормального и близких к нему распределений **максимально допустимой ошибкой наблюдения** и отбрасывают результаты наблюдений, для которых

$$|x_i - \bar{x}| > 3\sigma. \quad (23)$$

1.7. Для нормального распределения справедливо равенство

$$R = 6\sigma. \quad (24)$$

В условиях близости распределения единиц генеральной совокупности к нормальному это соотношение используется для

прогнозной оценки размаха вариации признака в генеральной совокупности.

2. Анализ типа закономерности распределения

При изучении социально-экономических явлений часто возникает эмпирическое распределение, хотя и не отвечающее строго нормальному закону, но имеющее с ним сходство, обусловленное тем, что крайние значения признака (близкие к x_{\max} и x_{\min}) встречаются много реже, чем серединные. (Таков, например, характер распределения признаков в *однородных* совокупностях). Поэтому сопоставление эмпирического распределения с нормальным важно для выяснения степени и характера расхождения между ними.

Строя график распределения, прежде всего пытаются выяснить, насколько сильно нарушено предположение о нормальности. Если эти нарушения невелики, то полученные выводы о статистических свойствах совокупности можно считать достаточно надежными. В противном случае возникает вопрос о целесообразности применения статистических методов, работающих в условиях нормального распределения, и замене их на методы, не чувствительные к распределению данных и устойчивые к различным отклонениям (так называемые робастные методы).

Возможность отнесения кривой распределения эмпирических данных к типу кривых нормального распределения устанавливается путем анализа формы гистограммы ряда распределения с учетом оценок показателей особенностей формы распределения — коэффициентов асимметрии и эксцесса (см. п. 3 раздела **Краткие теоретические сведения к Заданию 2**).

Коэффициенты асимметрии As и As_n характеризуют несимметричность распределения, а коэффициент эксцесса Ek — степень выраженности «хвостов» распределения, т.е. частоту появления значений, удаленных от среднего.

2.1. При анализе формы гистограммы прежде всего следует оценить распределение вариантов значений признака по интервалам (группам). Если на гистограмме четко прослеживаются два-три «горба» частот вариантов (рис. 12), это говорит о том, что значения признака концентрируются сразу в нескольких интервалах, и, следовательно, распределение не является однородным.

Если гистограмма имеет *одновершинную* форму, есть основания предполагать, что выборка является *однородной* по данному признаку.

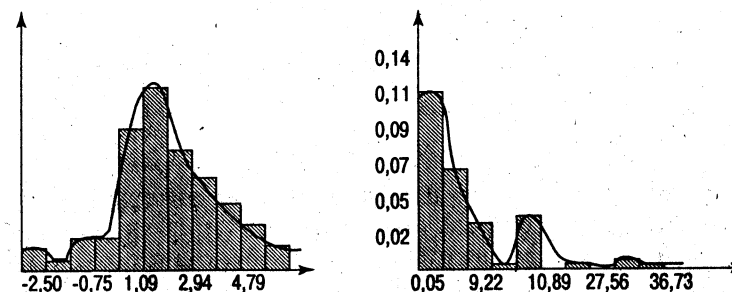


Рис. 12. Примеры гистограмм с длинными «хвостами» и резкой асимметричностью

Вместе с тем, следует иметь в виду, что при незначительном объеме выборки ($n < 50$) слишком углубленный анализ гистограммы может привести к неверным выводам, поскольку слабо выраженные «горбики и ямы» частот могут быть обусловлены не основными факторами, определяющими распределение единиц по группам, а просто *случайными отклонениями* вариантов от \bar{x} .

2.2. Установив по виду диаграммы однородность совокупности, для дальнейшего анализа формы распределения используются описательные параметры выборки — показатели центра распределения (\bar{x} , Mo , Me), вариации (σ), особенностей формы распределения (As_n , As , Ek), позволяющие оценить близость эмпирических данных к нормальной форме распределения.

Нормальное распределение является симметричным, и для него выполняются соотношения:

$$\bar{x} = Mo = Me, As = 0, As_n = 0; R = 6\sigma.$$

Нарушение этих соотношений свидетельствует о наличии асимметрии распределения. Распределение с небольшой или умеренной асимметрией в большинстве случаев по своему типу относится к нормальному.

Нарушение равенства $Ek = 0$ говорит о достаточно частом появлении крайних значений признака.

Таким образом, если гистограмма приблизительно симметрична, ее «хвосты» не очень длинные (не больше 5% вариантов лежат за пределами интервала $[\bar{x} \pm 2\sigma]$), то она *представляет распределение, близкое к нормальному*.

2.3. При резко асимметричной гистограмме (рис. 12) более удобной характеристикой «центра» распределения является медиана Me . Она более устойчива к резким выбросам данных, чем среднее \bar{x} , что позволяет использовать ее при работе с распре-

делениями, имеющими «хвосты». В этом случае для измерения вариации признака применяются коэффициент $V_{Me} = \frac{Me}{\bar{x}}$, учитывающий свойство медианы (1), а также квартильное отклонение $Q = \frac{Q_3 - Q_1}{2}$. Показатель Q рекомендуется и для оценки вариации в интервальных рядах с *открытыми* интервалами, когда показатель σ может быть исчислен лишь приближенно.

2.4. Если распределение единиц выборочной совокупности близко к нормальному, выборка является репрезентативной (значение показателей σ_N^2 и σ_n^2 расходятся незначительно) и при этом коэффициенты As_n , Ek_n указывают на небольшую или умеренную величину асимметрии и эксцесса соответственно, то есть основание полагать, что распределение единиц генеральной совокупности по изучаемому признаку будет близко к нормальному.

IV. РЕКОМЕНДАЦИИ К АНАЛИЗУ СТАТИСТИЧЕСКИХ СВОЙСТВ ИЗУЧАЕМОЙ СОВОКУПНОСТИ И СОСТАВЛЕНИЮ ОТЧЕТА

Ниже изложены рекомендации, которыми следует руководствоваться на заключительном этапе лабораторной работы при анализе полученных обобщающих показателей и подготовке отчета с выводами по результатам работы.

Нумерация пунктов в рекомендациях соответствует нумерации статистических задач из п. 4.1 — **Постановка задачи**.

Анализ выборочной совокупности

Задача 1. Указать количество аномальных единиц наблюдения со ссылкой на табл. 2.

Задача 2. Рассчитанные выборочные показатели представлены в двух таблицах — табл. 3 и 5. На основе этих таблиц необходимо сформировать *единую таблицу* значений выборочных показателей, перечисленных в условии Задачи 2, табл. 8 с заголовком «*Описательные статистики выборочной совокупности*».

Задача 3. Для ответа на вопросы 3а)–3г) следует воспользоваться теоретическими положениями, изложенными в методических указаниях к **Заключительному этапу** в разделе **Анализ обобщающих показателей описательной статистики** (пп. 1.1, 1.2, 1.4, 1.5).

При ответе на вопрос 3в) в случае обнаружения неустойчивых данных возможные аномалии следует выявить и указать в качестве «кандидатов» на исключение из выборки.

При ответе на вопрос 3г) необходимо сформировать таблицу следующего формата (с конкретными числовыми значениями границ диапазонов).

Таблица 9

РАСПРЕДЕЛЕНИЕ ЗНАЧЕНИЙ ПРИЗНАКА ПО ДИАПАЗОНАМ РАССЕЯНИЯ ПРИЗНАКА ОТНОСИТЕЛЬНО \bar{x}

	Границы диапазонов		Количество значений x_i , находящихся в диапазоне	
	Первый признак	Второй признак	Первый признак	Второй признак
$\bar{x} - \sigma_n \leq x_i \leq \bar{x} + \sigma_n$				
$\bar{x} - 2\sigma_n \leq x_i \leq \bar{x} + 2\sigma_n$				
$\bar{x} - 3\sigma_n \leq x_i \leq \bar{x} + 3\sigma_n$				

На основе данных табл. 9 определить процентное соотношение рассеяния значений признака по трем диапазонам и сопоставить его с ожидаемым по правилу «трех сигм».

Задача 4. Для ответа на вопросы 4а) — 4г) необходимо воспользоваться табл. 8 и сравнить величины показателей для двух признаков.

Отвечая на вопрос 4в), следует обратиться к п. 1.3 раздела **Анализ обобщающих показателей описательной статистики** методических указаний к **Заключительному этапу**.

Задача 5. При построении интервального вариационного ряда распределения единиц совокупности по признаку *Среднегодовая стоимость основных производственных фондов* необходимо воспользоваться теоретическими положениями, изложенными в методических указаниях к **Заключительному этапу** в разделе 2 — **Анализ типа закономерности распределения**.

Для полученного интервального ряда значение моды Mo рассчитывается по формуле

$$Mo = x_{Mo} + h \cdot \frac{(f_{Mo} - f_{Mo-1})}{(f_{Mo} - f_{Mo-1}) + (f_{Mo} - f_{Mo+1})},$$

где x_{Mo} — нижняя граница модального интервала; h — величина модального интервала; f_{Mo} — частота модального интервала; f_{Mo-1} — частота интервала, предшествующего модальному; f_{Mo+1} — частота интервала, следующего за модальным.

Полученные значения Mo следует сравнить со значениями моды из табл. 3. При наличии расхождений между ними объяснить причины этих расхождений.

Анализ генеральной совокупности

Задача 1. Генеральные показатели σ_N , σ_N^2 , As , Ek рассчитаны с помощью инструмента **Описательная статистика**, и их значения приведены в табл. 3. Необходимо сформировать для них отдельную таблицу (табл. 10) с заголовком «**Описательные статистики генеральной совокупности**».

Установить степень расхождения между σ_n^2 и σ_N^2 можно по формуле (16) раздела **Краткие теоретические сведения** к Заданию 2.

Прогнозные оценки размаха вариации R_N рассчитываются по формуле (24) из п. 1.7 раздела **Анализ обобщающих показателей описательной статистики** методических указаний к **Заключительному этапу**.

Для каждого признака прогнозное значение R_N следует сравнить с R_n .

Задача 2. Средние ошибки выборки рассчитаны и приведены в табл. 3 (параметр *Стандартная ошибка*). Для ответа на вопрос 2а) их следует выбрать из этой таблицы.

Оценки предельных ошибок выборки имеются в табл. 3, табл. 4а, табл. 4б. На основе этих оценок и формулы (15) необходимо сформировать табл. 11 следующего формата.

Таблица 11

ПРЕДЕЛЬНЫЕ ОШИБКИ ВЫБОРКИ И ОЖИДАЕМЫЕ ГРАНИЦЫ ДЛЯ ГЕНЕРАЛЬНЫХ СРЕДНИХ

Доверительная вероятность P	Коэффициент доверия t	Предельные ошибки выборки		Ожидаемые границы для средних \bar{x}	
		для первого признака	для второго признака	для первого признака	для второго признака
0,683	1			$\leq \bar{x} \leq$	$\leq \bar{x} \leq$
0,954	2			$\leq \bar{x} \leq$	$\leq \bar{x} \leq$
0,997	3			$\leq \bar{x} \leq$	$\leq \bar{x} \leq$

Задача 3. Для заключения о степени близости распределения единиц генеральной совокупности к нормальной форме следует обратиться к п. 2.4 раздела **Анализ типа закономерности распределения** методических указаний к **Заключительному этапу**. Значения коэффициентов асимметрии As и эксцесса Ek имеются в табл. 10.

ПРИЛОЖЕНИЕ 1



ВСЕРОССИЙСКИЙ ЗАОЧНЫЙ ФИНАНСОВО-ЭКОНОМИЧЕСКИЙ ИНСТИТУТ

КАФЕДРА СТАТИСТИКИ

ОТЧЕТ

о результатах выполнения компьютерной лабораторной работы № 1

«Автоматизированный априорный анализ статистической совокупности в среде MS Excel»

Вариант № _____

Выполнил: ст. III курса гр. _____

Ф.И.О. _____

Проверил: _____
Должность _____ Ф.И.О. _____

Москва, 2005 г.

ПРИЛОЖЕНИЕ 2

Таблица 1

БАЗОВЫЙ ВАРИАНТ ИСХОДНЫХ ДАННЫХ

	A	B	C
3	Номер предприятия	Среднегодовая стоимость основных производственных фондов, млн. руб.	Среднегодовая стоимость основных производственных фондов, млн. руб.
4	1	98,00	98,00
5	2	117,00	117,00
6	3	121,00	121,00
7	4	128,00	128,00
8	5	80,00	80,00
9	6	135,00	135,00
10	7	139,00	139,00
11	8	102,00	102,00
12	9	127,00	127,00
13	10	148,00	148,00
14	11	163,00	163,00
15	12	50,00	50,00
16	13	122,00	122,00
17	14	135,00	135,00
18	15	156,00	156,00
19	16	180,00	180,00
20	17	132,00	132,00
21	18	147,00	147,00
22	19	115,00	115,00
23	20	149,00	149,00
24	21	167,00	167,00
25	22	112,00	112,00
26	23	87,00	87,00
27	24	152,00	152,00
28	25	135,00	135,00
29	26	125,00	125,00
30	27	95,00	95,00
31	28	131,00	131,00
32	29	153,00	153,00
33	30	145,00	145,00
34	31	180,00	180,00
35	32	104,00	104,00

Примечание. Для формирования исходных данных вариантов лабораторной работы номер варианта задается в ячейке E2. Необходимо в табл. 1 вместо приведенных конкретных числовых значений поставить формулу их расчета как функцию этих значений и номера варианта. Например, в ячейке B4 может быть записана такая формула: $= 98 \cdot (1 + \$E\$2/2)$.

ПРИЛОЖЕНИЕ 3

РАСПОЛОЖЕНИЕ МАКЕТОВ РЕЗУЛЬТАТИВНЫХ ТАБЛИЦ
В РАБОЧЕМ ФАЙЛЕ ПЕРСОНАЛЬНОЙ ПАПКИ СТУДЕНТА

	A	B	C	D
37	Таблица 2			
38	Аномальные единицы наблюдения			
39	Номер предприятия	Среднегодовая стоимость основных производственных фондов, млн. руб.	Выпуск продукции, млн. руб.	
40				
41				
42	Таблица 3			
43	Описательные статистики			
44				
45	По столбцу «Среднегодовая стоимость основных производственных фондов», млн. руб.		По столбцу «Выпуск продукции», млн. руб.	
46	Столбец 1		Столбец 2	
47				
48	Среднее		Среднее	
49	Стандартная ошибка		Стандартная ошибка	
50	Медиана		Медиана	
51	Мода		Мода	
52	Стандартное отклонение		Стандартное отклонение	
53	Дисперсия выборки		Дисперсия выборки	
54	Эксцесс		Эксцесс	
55	Асимметричность		Асимметричность	
56	Интервал		Интервал	
57	Минимум		Минимум	
58	Максимум		Максимум	
59	Сумма		Сумма	
60	Счет		Счет	
61	Уровень надежности (95,4%)		Уровень надежности (95,4%)	

	A	B	C	D
62	Таблица 4а			
63				
64				
65				
	Предельные ошибки выборки			
66	По столбцу «Среднегодовая стоимость основных производственных фондов», млн. руб.		По столбцу «Выпуск продукции», млн. руб	
67	Столбец 1		Столбец 2	
68				
69	Уровень надежности (68,3%)		Уровень надежности (68,3%)	
70	Таблица 4б			
71				
72				
73				
	Предельные ошибки выборки			
74	По столбцу «Среднегодовая стоимость основных производственных фондов», млн. руб.		По столбцу «Выпуск продукции», млн. руб	
75	Столбец 1		Столбец 2	
76				
77	Уровень надежности (99,7%)		Уровень надежности (99,7%)	
78	Таблица 5			
79				
80				
81				
	Выборочные показатели вариации и асимметрии			
82	По столбцу «Среднегодовая стоимость основных производственных фондов», млн. руб.		По столбцу «Выпуск продукции», млн. руб	
83	Стандартное отклонение		Стандартное отклонение	
84	Дисперсия		Дисперсия	
85	Среднее линейное отклонение		Среднее линейное отклонение	
86	Коэффициент вариации, %		Коэффициент вариации, %	
87	Коэффициент асимметрии		Коэффициент асимметрии	

	A	B	C
88	Таблица 6		
89			
90			
91			
92			
93			
94			
95			
96			
97			
98	Таблица 7		
99			
100			
101			
102			
103			
104			
104			
106			
107			

ЛИТЕРАТУРА

1. Громыко Г.Л. Теория статистики: Учебник. — М.: ИНФРА-М, 2005.
2. Ефимова М.Р., Петрова Е.В., Румянцев В.Н. Общая теория статистики: Учебник. — М.: ИНФРА-М, 2004.
3. Козлов А.Ю., Мхитарян В.С., Шишов В.Ф. Статистические функции MS Excel в экономико-статистических расчетах: Учеб. пособие для вузов/Под ред. проф. В.С. Мхитаряна. — М.: ЮНИТИ-ДАНА, 2003.
4. Козлов А.Ю., Шишов В.Ф. Пакет анализа MS Excel в экономико-статистических расчетах: Учеб. пособие для вузов / Под ред. проф. В.С. Мхитаряна. — М.: ЮНИТИ-ДАНА, 2003.
5. Макарова Н.В., Трофимец В.Я. Статистика в Excel: Учеб. пособие. — М.: Финансы и статистика, 2002.
6. Общая теория статистики: Статистическая методология в изучении коммерческой деятельности: Учебник / Под ред. О.Э. Башиной, А.А. Спирина. — М.: Финансы и статистика, 2005.
7. Практикум по статистике: Учеб. пособие для вузов / Под ред. В.М. Симчеры; ВЗФЭИ. — М.: Финстатинформ, 1999.
8. Салин В.Н., Медведев В.А., Кудряшова С.И., Е.П. Шпаковская Макроэкономическая статистика: Учеб. пособие. — М.: Дело, 2000.
9. Статистика: Учебник / Под ред. И.И. Елисеевой. — М.: ТК Велби, Проспект, 2002.
10. Теория статистики: Учебник / Под ред. Р.А. Шмойловой. — М.: Финансы и статистика, 2004.

СОДЕРЖАНИЕ

Лабораторная работа № 1

Автоматизированный априорный анализ
статистической совокупности в среде MS Excel 3

I. ЦЕЛИ, СОДЕРЖАНИЕ И ОРГАНИЗАЦИЯ ВЫПОЛНЕНИЯ ЛАБОРАТОРНОЙ РАБОТЫ 3

1. Цель и задачи работы 3
2. Компьютерные средства 4
3. Требования к уровню подготовленности студента
к лабораторной работе 4
4. Содержание и структура лабораторной работы 4
 - 4.1. Постановка задачи 4
 - 4.2. Экономическое содержание задач статистического
исследования 7
 - 4.3. Структура лабораторной работы 9
5. Отчетность по работе 10

II. ТЕОРЕТИЧЕСКИЕ ОСНОВЫ И МЕТОДИКА ПРОВЕДЕНИЯ АПРИОРНОГО СТАТИСТИЧЕСКОГО АНАЛИЗА ДАННЫХ 11

1. Статистическая сущность задач априорного анализа данных 11
2. Общая методика априорного анализа статистических данных 20

III. ПОРЯДОК ВЫПОЛНЕНИЯ ЛАБОРАТОРНОЙ РАБОТЫ 21

1. Подготовительный этап 21
2. Этап выполнения статистических расчетов 23
3. Заключительный этап 58

IV. РЕКОМЕНДАЦИИ К АНАЛИЗУ СТАТИСТИЧЕСКИХ СВОЙСТВ ИЗУЧАЕМОЙ СОВОКУПНОСТИ И СОСТАВЛЕНИЮ ОТЧЕТА 62

ПРИЛОЖЕНИЕ 1 65

ПРИЛОЖЕНИЕ 2 66

ПРИЛОЖЕНИЕ 3 67

Литература 70

Статистика. Компьютерные лабораторные работы: Методические указания к лабораторной работе № 1 «Автоматизированный априорный анализ статистической совокупности в среде MS Excel». — М.: Вузовский учебник, 2005.

Сдано в набор 12.08.2005. Подписано в печать 25.08.2005.
Формат 60x88/16. Бумага типографская № 2. Гарнитура «*Newton*».
Печать офсетная. Усл. печ. л. 4,9. Уч.-изд. л. 5,4.
Тираж 10 000 экз. Заказ № 1/12-05.

Издательский Дом «Вузовский учебник»
127247, Москва, ул. С. Ковалевской, д. 1, стр. 52

Отпечатано в полном соответствии
с качеством предоставленных диапозитивов в ФГУП ДПК Роспатента
142001, г. Домодедово, Каширское шоссе, 4, корп. 1.