

Федеральное агентство по образованию  
Государственное образовательное учреждение  
высшего профессионального образования  
Владимирский государственный университет

В.Н. КРАЕВ

КУРС ЛЕКЦИЙ  
ПО ОБЩЕЙ ТЕОРИИ СТАТИСТИКИ

Владимир 2006

УДК 311 (075.8)  
ББК 60.6я73  
К78

Рецензенты:

Доктор экономических наук, профессор  
кафедры маркетинга и экономики производства  
Владимирского государственного университета  
*Ю.А. Дмитриев*

Кандидат экономических наук, доцент  
зав. кафедрой менеджмента и маркетинга  
Владимирского института бизнеса  
*И.В. Паньшин*

Печатается по решению редакционно-издательского совета  
Владимирского государственного университета

**Краев, В. Н.**

К78 Курс лекций по общей теории статистики / В. Н. Краев ; Владим.  
гос. ун-т. – Владимир : Изд-во Владим. гос. ун-та, 2006. – 96 с.  
ISBN 5-89368-678-0.

Излагаются общие вопросы теории статистики. Рассматриваются методы группировок, относительных и средних величин, показатели вариации, корреляционный и динамический анализ, статистическая взаимосвязь социально-экономических явлений, выборочные наблюдения. В каждой лекции дается решение типовых задач.

Предназначено для студентов экономических специальностей всех форм обучения.  
Ил. 9. Табл. 31. Библиогр.: 11 назв

УДК 311 (075.8)  
ББК 60.6я73

ISBN 5-89368-678-0

© Владимирский государственный  
университет, 2006

## ОГЛАВЛЕНИЕ

ВВЕДЕНИЕ .....	1
Лекция 1. СТАТИСТИЧЕСКОЕ НАБЛЮДЕНИЕ .....	6
1.1. ОСНОВНЫЕ ПОНЯТИЯ И ФОРМЫ СТАТИСТИЧЕСКОГО НАБЛЮДЕНИЯ .....	6
1.2. ВИДЫ И СПОСОБЫ СТАТИСТИЧЕСКОГО НАБЛЮДЕНИЯ .....	7
РЕШЕНИЕ ТИПОВЫХ ЗАДАЧ. ....	10
1.3. ОШИБКИ НАБЛЮДЕНИЯ И ИХ КОНТРОЛЬ .....	11
Лекция 2. СВОДКА И ГРУППИРОВКА СТАТИСТИЧЕСКИХ ДАННЫХ.....	13
2.1. Сводка и группировка. Виды группировок.....	13
2.2. Группировочные признаки.....	15
2.3. Предварительные сведения о рядах распределения .....	16
Лекция 3. АБСОЛЮТНЫЕ И ОТНОСИТЕЛЬНЫЕ ВЕЛИЧИНЫ.....	19
3.1. Абсолютные величины .....	19
3.2. Относительные величины: сущность, виды, единицы ИЗМЕРЕНИЯ .....	20
Лекция 4. СРЕДНИЕ ВЕЛИЧИНЫ.....	24
4.1. Общие положения.....	
4.2. Средняя арифметическая.....	24
4.3. Средняя гармоническая .....	26
4.4. Средняя геометрическая.....	26
4.5. Средняя квадратическая.....	27
4.6. Соотношение между различными видами средних (мажорантность средних) .....	28
4.7. Мода и медиана .....	29
Лекция 5. ПОКАЗАТЕЛИ ВАРИАЦИИ И ДРУГИЕ ХАРАКТЕРИСТИКИ РЯДОВ РАСПРЕДЕЛЕНИЯ.....	35
5.1. Понятие о вариации .....	35
5.2. Показатели вариации и способы их расчета .....	35
5.3. Основные свойства дисперсии .....	37
5.4. Правило сложения дисперсий.....	38

5.5. РЯДЫ РАСПРЕДЕЛЕНИЯ И ИХ МОДЕЛИРОВАНИЕ .....	39
Лекция 6. ВЫБОРОЧНОЕ НАБЛЮДЕНИЕ .....	49
6.1. ОБЩИЕ СВЕДЕНИЯ О ВЫБОРОЧНОМ НАБЛЮДЕНИИ.....	49
6.2. ВИДЫ И СХЕМЫ ОТБОРА .....	50
6.3. ОПРЕДЕЛЕНИЕ СРЕДНИХ И ПРЕДЕЛЬНЫХ ОШИБОК ПРИ РАЗЛИЧНЫХ ВИДАХ ОТБОРА .....	51
6.4. ОПРЕДЕЛЕНИЕ НЕОБХОДИМОЙ ЧИСЛЕННОСТИ ВЫБОРКИ.....	53
Лекция 7. СТАТИСТИЧЕСКОЕ ИЗУЧЕНИЕ СВЯЗИ .....	58
7.1. ВИДЫ И ФОРМЫ ВЗАИМОСВЯЗИ МЕЖДУ ЯВЛЕНИЯМИ .....	58
7.2. ОСНОВНЫЕ ПРИЕМЫ ИЗУЧЕНИЯ ВЗАИМОСВЯЗЕЙ .....	58
7.3. КОРРЕЛЯЦИОННЫЙ АНАЛИЗ .....	59
7.4. МНОЖЕСТВЕННАЯ КОРРЕЛЯЦИЯ.....	62
7.5. ПРОСТЕЙШИЕ МЕТОДЫ ИЗМЕРЕНИЯ ТЕСНОТЫ СВЯЗИ .....	63
Лекция 8. РЯДЫ ДИНАМИКИ .....	70
8.1. ПОНЯТИЕ О ДИНАМИЧЕСКИХ РЯДАХ И ИХ ВИДАХ .....	70
8.2. ОСНОВНЫЕ ПОКАЗАТЕЛИ, ИСПОЛЬЗУЕМЫЕ ПРИ АНАЛИЗЕ ДИНАМИЧЕСКИХ РЯДОВ .....	72
8.3. ВАЖНЕЙШИЕ ПРИЕМЫ ОБРАБОТКИ И АНАЛИЗА ДИНАМИЧЕСКИХ РЯДОВ .....	74
8.4. КОРРЕЛЯЦИОННЫЙ АНАЛИЗ И СЕЗОННЫЕ КОЛЕБАНИЯ В РЯДАХ ДИНАМИКИ.....	75
КОНТРОЛЬНЫЕ ВОПРОСЫ И ТЕМЫ ДЛЯ ПОДГОТОВКИ К ЗАЧЕТУ И ЭКЗАМЕНУ .....	79
БИБЛИОГРАФИЧЕСКИЙ СПИСОК.....	95

## ВВЕДЕНИЕ

Во всем мире возрастает интерес к статистике. В нашей стране это внимание тем более обострено в связи с осуществлением экономических реформ, затрагивающих интересы всех людей.

Одно из неперемennых условий правильного восприятия и тем более практического использования статистической информации – знание статистической методологии.

Курс «Общая теория статистики» является одной из дисциплин, формирующих специалистов высокой квалификации – менеджеров и финансистов. Он тесно связан с экономическими, финансовыми дисциплинами, менеджментом, маркетингом, которые обеспечивают современную фундаментальную подготовку специалистов.

После изучения курса студенты должны знать:

- содержание основных этапов статистического исследования;
- порядок сводки и группировки статистических данных; методы сбора и обработки первичной статистической информации для проведения качественного экономического анализа; уметь проверять достоверность первичных данных в формах статистической отчетности;
- основные формулы и зависимости, используемые при анализе статистических данных, уметь анализировать и находить зависимости в изучаемых явлениях;
- методы вычисления основных статистических показателей.

Студенты должны выработать практические навыки для проведения статистического исследования.

В предлагаемом издании содержится системное изложение основных разделов курса (за исключением статистических таблиц и индексов), изучение которых во многом способствует формированию деловых качеств менеджера.

Особенностью курса лекции является то, что вопросы статистической методологии рассматриваются на примерах решения типовых задач.

Издание рассчитано на студентов всех форм обучения технических и экономических специальностей, изучающих данный курс.

## Лекция 1. СТАТИСТИЧЕСКОЕ НАБЛЮДЕНИЕ

### 1.1. Основные понятия и формы статистического наблюдения

Всякое статистическое исследование делится на три последовательных этапа: статистическое наблюдение, сводка и группировка материалов, полученных в результате наблюдения; вычисление обобщающих показателей, всесторонне характеризующих изучаемое явление, и их анализ.

Таким образом, статистическое наблюдение (СН) – исходный этап статистического исследования, который состоит в планомерном, научно-организованном собирании массовых данных о явлениях и процессах общественной жизни. Целью СН является регистрация элементов, составляющих изучаемое массовое явление, по заранее установленным наиболее существенным признакам. Это означает, что СН должно быть организовано как планомерное, массовое и систематическое.

**Планомерность** статистического наблюдения заключается в том, что его подготавливают и осуществляют по заранее подробно разработанному плану и программе, которые охватывают все вопросы методологии, организации и техники сбора информации. В программно-методологическую часть плана статистического наблюдения включаются вопросы определения цели, объекта наблюдения, единицы наблюдения и единицы совокупности.

Объектом статистического наблюдения называется совокупность единиц изучаемого явления, подлежащих статистическому изучению (совокупность предприятий, лиц, магазинов, фермерских хозяйств и др.).

Единицей наблюдения называется первичная ячейка совокупности, которая является источником информации.

Единицей совокупности называется элемент единицы наблюдения, служащий основой счета и обладающий признаками, подлежащими регистрации.

Каждая единица наблюдения и единица совокупности должны характеризоваться совокупностью признаков, которые по форме и содержанию делятся на *количественные*, т.е. выражающиеся числом (возраст, заработная плата и т.д.) и *атрибутивные*, характеризующие качественное состояние данного явления и не имеющие количественного выражения (пол, специальность,

образование и пр.). В свою очередь, и те и другие могут быть *факторными*, т.е. характеризующими условия, определяющие размер того или иного явления, и *результативными*, являющимися результатом действия факторных признаков (например, стаж работы – факторный признак, производительность труда – результативный признак).

**Массовый характер** статистического наблюдения означает, что оно организовано и направлено на охват возможно большего, массового числа случаев проявления данного явления или процесса

**Систематичность** статистического наблюдения определяется тем, что оно производится не стихийно, а регулярно или непрерывно.

Статистическое наблюдение осуществляется в двух формах: путем представления отчетности и проведения специально организованного статистического наблюдения, причем основной формой СН является представление отчетности.

## 1. 2. Виды и способы статистического наблюдения

Статистическое наблюдение различается по видам и способам проведения. Их можно классифицировать следующим способом:

### I. По степени охвата единиц исследуемой совокупности

По этому классификационному признаку СН подразделяется на два способа.

1. **Сплошное наблюдение**, когда охватываются все единицы совокупности (например перепись населения, текущая отчетность предприятия).

2. **Несплошное (частичное) наблюдение** – обследование охватывает определенную часть изучаемой совокупности. Это наблюдение занимает очень важное место в статистике и подразделяется на следующие виды:

а) *метод основного массива* (наблюдение наиболее важных объектов) заключается в том, что обследованию подвергается та часть совокупности, у которой величина изучаемого признака является определяющей во всем объеме. Например, отслеживание объема реализации и динамики цен на колхозных рынках в России проводилось в 264 крупных городах, составляющих менее 5 % всех городов, но в которых проживало более 50 % городского населения.

б) *монографическое наблюдение* состоит в подробном описании небольшого числа или отдельных единиц совокупности (описание работы передового или отстающего предприятия и т.д.);

в) суть *анкетного вида* несплошного наблюдения заключается в рассылке и сборе анкет;

г) *случайный отбор* определенного количества единиц из совокупности (выборочный метод). В свою очередь, он может выполняться следующими способами:

- при *собственно-случайном отборе*; отбор единиц для выборочной совокупности осуществляется случайным образом по жребию;
- сущность *механического отбора* состоит в том, что единицы изучаемой совокупности предварительно располагают в определенном порядке (например по возрастанию или убыванию), а потом механическим путем выбирают необходимое количество единиц (например, берут каждую 4-ю, каждую 10-ю единицу и пр.);
- *типологическому отбору* предшествует распределение единиц изучаемой совокупности на группы по типичному признаку, затем внутри каждой группы проводится тот или иной отбор (собственно-случайный, механический или любой другой);
- *серийный отбор* предполагает образование выборочной совокупности не из отдельных единиц, а из целых серий, причем серии выбираются путем механического или собственно-случайного отбора, а внутри каждой серии проводится сплошное наблюдение.

Все способы случайного отбора могут выполняться по двум схемам:

– повторный отбор (схема возвращенного шара) – после выбора какой-либо единицы она вновь возвращается в исходную совокупность и может быть выбрана снова;

– бесповторный отбор (схема невозвращенного шара) – отобранная единица обратно в совокупность не возвращается и не может быть выбрана снова.

II. В зависимости от времени статистическое наблюдение может быть непрерывным (текущим), периодическим и единовременным.

1. Непрерывным, или текущим, называется такое наблюдение, которое проводится непрерывно, по мере возникновения явлений, например учет выпуска продукции на предприятии.

2. Если наблюдение проводится через определенные промежутки времени, то оно называется периодическим (сессия в вузах).

3. Единовременное наблюдение проводится по мере необходимости, например перепись населения.

Установление времени наблюдения – очень важно, и в статистике различают объективное и субъективное время и критический момент.

Объективным временем называется время, к которому относятся данные наблюдения. Оно характеризует тот момент или период времени, на которые по состоянию были собраны и зарегистрированы данные. Например, данные о выпуске продукции, размерах потребления и так можно получить только за определенный период. Сведения же о численности населения, числе больничных коек и тому подобного можно собрать только по состоянию на определенную дату.

Если наблюдение приурочено к определенному моменту, то этот момент называется критическим.

Субъективное время наблюдения – это время производства наблюдения, т.е. период, в течение которого проводится регистрация единиц совокупности. В целях обеспечения точности наблюдения субъективное время должно измеряться возможно более коротким периодом. Когда наблюдение основано на документальных данных, длительность субъективного времени не имеет значения.

III. В зависимости от источников собираемых данных различают:

1. Непосредственное наблюдение, т.е. наблюдение лично регистратором – снятие товарных остатков, изучение и замер норм времени и т.д.

2. Документальное наблюдение, когда используются различного рода документы.

3. Наблюдение базируется на опросе заинтересованных лиц и получение данных в форме ответов.

IV. По способу организации наблюдения различают:

1. Наблюдение, заключающееся в обработке отчетных данных, – отчетность – наиболее распространено в практике работы.

2. Экспедиционный способ – к каждой единице совокупности посылается специальное лицо, которое в соответствующих формулярах фиксирует необходимые сведения.

3. Саморегистрация – заполнение специально выданных бланков.

4. Анкетный способ – рассылка анкет и их обработка.

Наиболее распространенная форма статистического наблюдения – представление отчетности. Статистическая отчетность делится на типовую и специализированную; по периодичности различают недельную, двухнедельную, месячную, квартальную и годовую отчетность; в зависимости от установленного порядка прохождения отчетность может быть централизованной (наиболее распространена) и децентрализованной.

### **Решение типовых задач.**

**Задача 1.** *На заводе ведется учет выполнения норм выработки. Укажите виды наблюдения по времени проведения, охвату единиц совокупности и источнику данных.*

Так как наблюдению подвергаются все без исключения рабочие, то по степени охвата единиц наблюдение сплошное; а также непрерывное, так как проводится по мере возникновения явления; источником данных служат документы.

**Задача 2.** *Проведено обследование части школьных буфетов в городе. Укажите вид наблюдения по степени охвата единиц и способу проведения.*

Так как обследовалась только часть буфетов, то наблюдение несплошное и, по-видимому, был применен экспедиционный способ, т.е. в каждый проверяемый буфет посылался специальный человек, который и фиксировал необходимую информацию.

**Задача 3.** *Первая Всесоюзная перепись населения 1926 г. проводилась по состоянию на 12 ч. ночи с 16 на 17 декабря и продолжалась в городах 7, а в сельской местности – 14 дней. Укажите объективное и субъективное время и критический момент наблюдения.*

Объективное время – 12 ч ночи с 16 на 17 декабря. Этот же момент является критическим. Субъективное время: в городах – 7 дней; в сельской местности – 14 дней.

**Задача 4.** *Срок представления годовых отчетов за 1995 г. – 1 – 25 января 1996 г. Укажите объективное и субъективное время и критический момент наблюдения.*

Объективное время – 1995 г., так как данные наблюдения относятся именно к этому периоду; субъективное время – 1-25 января 1996 г., так как в это время составляется отчет, собираются данные; критический момент отсутствует.

**Задача 5.** *Проводится перепись художественной литературы в библиотеках области. Укажите объект, единицу наблюдения и единицу совокупности.*

Объектом наблюдения являются библиотеки области, единица наблюдения – фонд художественной литературы, так как другие ее виды не учитываются, а единица наблюдения – каждая конкретная книга, поскольку она обладает признаками, подлежащими регистрации.

### **1.3. Ошибки наблюдения и их контроль**

Ошибкой называется расхождение между результатом наблюдения и истинным значением исследуемой величины.

Все ошибки статистического наблюдения можно классифицировать:

1. По характеру ошибок:

а) *случайные* ошибки, которые вызывают самые разные причины. Так как искажение может быть как в ту, так и в другую сторону, то случайные ошибки могут не очень сильно влиять на общий результат;

б) *систематические* ошибки, которые искажают явление только в одну сторону, более опасны и, как правило, вызываются действием систематического фактора (например, тяготение к круглым цифрам вносит большие осложнения при собирании сведений о возрасте населения).

2. По стадии возникновения:

а) *ошибки регистрации;*

б) *ошибки при подготовке данных к обработке;*

в) *ошибки при обработке.*

3. По причинам возникновения:

а) *свойственные только выборочному методу и связанные с неправильным выбором части совокупности ошибки репрезентативности (представительности);*

б) *непреднамеренные* ошибки, совершаются неумышленно, непреднамеренно, т. е. не имеют целью исказить результаты наблюдения;

в) *преднамеренные* ошибки возникают в результате умышленного искажения фактов. Эти ошибки – систематические.

Для исключения ошибок статистические данные проверяют арифметическим и логическим контролем.

Логический контроль состоит в сопоставлении ответов на взаимосвязанные между собой вопросы программы наблюдения.

Счетный и арифметический контроль сводится к проверке общих и групповых цифровых итогов и сопоставлений; его задача – обнаружить и исправить неверные итоги числовых показателей.

### **Решение типовых задач**

*Задача 1. В сводку завода о выпуске продукции за февраль была включена и продукция, выпущенная 1-го и 2-го марта. Укажите характер допущенной ошибки и причины возникновения.*

Ошибка систематическая и преднамеренная, так как были сознательно искажены действительные результаты для получения определенных материальных выгод.

*Задача 2. В переписном листе записано: «лет-, месяцев – 11» вместо «лет – 11, месяцев-». Укажите характер и причину ошибки.*

Ошибка случайная (описка) и непреднамеренная.

*Задача 3. Проведите логический контроль правильности заполнения переписного листа Всесоюзной переписи населения 1970 г. В ответах на какие вопросы допущены ошибки?*

#### Переписной лист № 4

1. Фамилия имя и отчество – Антонова Мария Павловна.
2. Пол – женский.
3. Отношение к главе семьи – внучка.
4. Возраст (число исполнившихся лет) – 4.
5. Состоит ли в браке в настоящее время – да.
6. Образование – не имеет начального.

Логика ответов подсказывает, что ошибка в 5-м вопросе. Должно быть – «нет».

## Лекция 2. СВОДКА И ГРУППИРОВКА СТАТИСТИЧЕСКИХ ДАННЫХ

### 2.1. Сводка и группировка. Виды группировок

В результате статистического наблюдения получают сведения о каждой единице совокупности в отдельности. Чтобы на основе этих данных сделать определенные выводы, необходимо провести сводку полученных материалов, т.е. в узком смысле сводка – это подсчет или подведение итогов.

Однако в более широком смысле под статистической сводкой понимают сложную операцию научной обработки первичных статистических данных, которая включает группировку материала, разработку системы показателей для характеристики типичных групп и подгрупп, подсчет (подведение) итогов по группам и по совокупности в целом и изображение сгруппированных данных в виде статистических таблиц.

Статистическая группировка – расчленение общей совокупности единиц по одному или нескольким существенным признакам на однородные группы, различающиеся между собой в качественном и количественном отношении и позволяющие выделить социально-экономические типы, изучить структуру совокупности или проанализировать связи между отдельными признаками. Группировка проводится только для целей конкретного исследования.

В соответствии с решаемыми задачами все группировки подразделяются на типологические, структурные и аналитические.

1. Расчленение разнородной совокупности на качественные однородные группы, выделение качественно однородных экономических групп или типов общественных явлений осуществляется при помощи *типологических* группировок (табл. 2.1).

Таблица 2.1

*Распределение промышленной продукции, произведенной в различных формах хозяйствования за отчетный период*

Группы предприятий по формам хозяйствования	Объем продукции, млрд руб.	Процент к итогу
Государственные, с традиционными методами управления	405,0	89,20
Арендные	19,0	4,19
Кооперативные	30,0	6,61
<i>Итого:</i>	454,0	100,00

Приведенные данные характеризуют социально-экономический состав промышленной продукции (выделены типы явления), а также свидетельствуют о возникших в последнее время новых формах хозяйствования.

2. Группировка, цель которой – расчленение совокупности единиц на группы, характеризующие структуру по определенным признакам, называется *структурной группировкой*.

Приведенная в табл. 2.2 группировка содержит систему показателей, характеризующих структуру изучаемой совокупности по ряду признаков, а также концентрацию торгово-закупочного процесса, нашедшего свое выражение в укрупнении магазинов по величине товарооборота.

Таблица 2.2

*Группировка торговых предприятий района по объему товарооборота (в процентах к итогу)*

Группы магазинов по объему товарооборота, тыс. р.	Количество магазинов	Розничный товарооборот	Торговая площадь
До 1700	21,87	11,22	18,05
1700 - 2000	28,13	19,04	21,38
2000 - 3000	21,87	20,0	19,08
3000 - 4200	15,63	22,23	19,47
Свыше 4200	12,5	27,51	22,02
Итого	100,00	100,00	100,00

3. Группировка, целью которой является установление взаимосвязей между изучаемыми явлениями, называется *аналитической*. Примером аналитической группировки могут служить данные табл. 2.3.

Таблица 2.3

*Зависимость урожайности зерновых культур от количества внесенных удобрений по группе совхозов*

Количество внесенных удобрений, % от нормы	Количество совхозов	Средняя урожайность по группе, ц / га
До 30	2	20,2
30 – 50	5	21,8
50 – 80	8	25,6
80 – 100	14	30,3
Итого	29	26,8

Данные показывают, что увеличение количества внесенных удобрений (факторный признак) приводит при прочих равных условиях к повышению урожайности (результативный признак).

## 2.2. Группировочные признаки

Признаки в статистике – это свойства, характерные черты или особенности явлений, которые можно выразить рядом статистических величин. Признаки, положенные в основу группировки, называются группировочными.

Группировки по одному признаку называются *простыми*. Когда же для выделения групп берут два признака, и более т.е. группы, образованные по одному признаку, подразделяются на подгруппы по другому, а полученные в результате этого подгруппы подразделяются (каждая в отдельности) еще на подгруппы и так далее, то такие группировки называются *комбинационными*.

Различают четыре вида группировочных признаков: атрибутивные, количественные, признаки пространства и признаки времени.

Атрибутивным называется признак, который характеризует свойство, качество данного явления и не имеет количественного выражения. При группировке по атрибутивным (качественным) признакам статистическая совокупность разбивается на столько групп, сколько разновидностей имеет признак (по полу – на две группы, по национальному составу – на столько групп, сколько имеется национальностей, и т.д.).

Разновидностью атрибутивных признаков являются *альтернативные*, т.е. такие признаки, которыми одни единицы совокупности обладают, а другие – нет. Например, одни работники имеют высшее образование, а другие не имеют, и т.д.

2. Количественным называется признак, характеризующий размеры, величину изучаемой совокупности и дающий возможность расчленить её на группы по величине индивидуальных значений группировочного признака.

3. Признак пространства – это адресный признак (адрес предприятия и пр.). Группировки по признаку пространства применяются для изучения пространственных (географических) закономерностей.

4. Признак времени позволяет установить хронологию событий, и его применяют для изучения изменений явлений во времени.

Признаки бывают *первичные и вторичные*. Первичные характеризуют абсолютные размеры изучаемых явлений (например численность рабочих), вторичные являются производными от первичных и показывают структуру группируемых явлений (фондовооруженность, себестоимость и т.д.).

### 2.3. Предварительные сведения о рядах распределения

Ряд цифровых показателей, представляющий распределение единиц совокупности по одному признаку, разновидности которого расположены в определенной последовательности, называется рядом распределения.

Ряд распределения – это первичная характеристика массовой статистической совокупности, в которой находят количественное выражение закономерности массовых явлений и процессов.

Ряды распределения могут быть построены как по атрибутивному признаку – атрибутивные ряды распределения, так и по количественному – вариационные ряды распределения.

Числовые значения количественного признака в вариационном ряду называются вариантами и располагаются в определенном порядке; численность единиц в каждой группе называется частотой, а если частоты выражены в процентах, то они называются частостями. Рассмотрим табл. 2.4.

Таблица 2.4

*Распределение рабочих завода по уровню месячной заработной платы*

Зарплата, тыс. руб.	Количество рабочих	Проценты к итогу
Признак $x$	Частота $f$	Частость $f$
..... 350 – 450 ↙ варианта .....	..... 89 .....	..... 48,0 .....
В с е г о	180 (объем ряда распределения)	100,0

В данной таблице уровень заработной платы является вариантой, а число рабочих – частотой, показывающей, сколько раз повторяется данная варианта (т.е. сколько рабочих получают заработную плату, лежащую в пределах от 350 до 450 тыс. руб.). Частота, выраженная в процентах (в данном примере - 48,0) - это частость. Сумма всех частот называется объемом ряда распределения, или его численностью (в данном случае было обследовано 180 рабочих).

Вариация признака может быть дискретной (прерывной) или непрерывной. При дискретной вариации признака значение вариантов отличается друг

от друга на вполне определенную величину, и варианты обычно выражаются целыми числами (например, число членов семьи – 1, 2, 3, 4 и т.д.). Такие ряды называются вариационными дискретными рядами.

При непрерывной вариации признака его величина может принимать любые значения в определенном интервале (интервальные ряды распределения), например рост человека, урожайность, заработная плата и т.д. Построение интервального ряда основано на принципах статистической группировки.

Интервалы могут быть неравными, поэтому сравнивать частоты и судить о степени заполнения разных интервалов нельзя. В целях сравнения заполненности интервалов рассчитывается показатель, называемый *плотностью распределения*, т.е. число единиц совокупности, приходящееся в среднем на одну единицу ширины интервала (ширина интервала – разность между его границами). Если плотность определяется с помощью частот, она называется абсолютной, если с помощью частостей – относительной.

Для облегчения анализа рядов распределения их изображают при помощи графиков.

При дискретной вариации анализа признака графиком вариационного ряда служит *полигон распределения*. Он представляет собой замкнутый многоугольник, абсциссами вершин которого являются значения варьирующего признака, а ординатами – соответствующие им частоты или частости. Для примера рассмотрим построение полигона распределения по следующим данным (табл. 2.5). Полигон распределения приведен на рис. 2.1.

Таблица 2.5

*Распределение квартир жилого дома по числу проживающих в них лиц*

Количество лиц, живущих в квартире $x$	1	2	3	4	5	6	7	Всего 100
Число квартир, $f$	4	6	22	46	18	3	1	
Накопленные частоты	4	10	32	78	96	99	100	

Для иллюстрации рядов распределения часто используется кумулятивная кривая, или кумулята, построенная по накопленным частотам. Накопленная частота показывает, какое число единиц имеет величину варианта не большую данной и находится путем последовательного суммирования частот (стр. 3 в табл. 2.5) Например, из табл. 2.5 видно, что в 10 квартирах проживают семьи, состав которых не превышает двух человек, в 78 квартирах проживают семьи, состав которых не превышает 4 человек и т. д. Кумулята приведена на рис.2.2.

Если при построении кумуляты поменять местами значения признака и частоты (т.е. на оси ординат отложить значения признака, а на оси абсцисс – частоты) то получим кривую, называемую огивой.

Для иллюстрации интервального ряда распределения применяют график, называемый гистограммой распределения. Гистограмма представляет собой ряд смежных столбиков, основание каждого равно величине соответствующего интервала, а высота – частоте (правильнее – пропорциональна частоте). Пример гистограммы приведен на рис.2.3.

Разновидностью кумулятивной кривой является график Лоренца или кривая концентрации. Например, необходимо дать характеристику уровня концентрации производства в промышленности России по числу рабочих (см. табл. 2.6). График приведен на рис. 2.4. Если бы численность рабочих распределялась равномерно по числу предприятий, т.е. процесс концентрации отсутствовал полностью, график представлял бы собой прямую линию. При неравномерном распределении линия концентрации отходит от прямой, причем, чем выше уровень концентрации, тем больше ее кривизна.

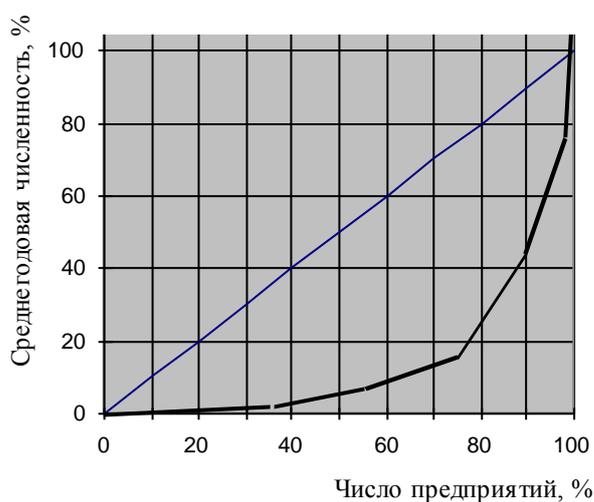


Рис. 2.4. График Лоренца

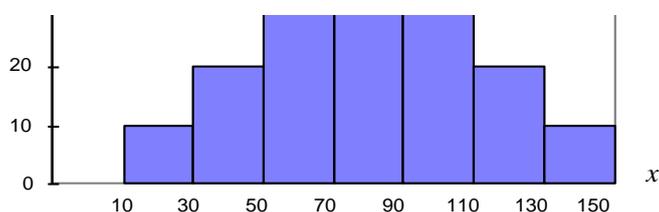


Рис. 2.3. Гистограмма распределения электродвигателей по мощности.

Таблица 2.6

Группировка промышленных предприятий по численности  
рабочих за 19\_\_ г.

Группы предприятий по числу рабочих, чел	Число предприятий		Среднегодовая численность	
	% к итогу	кумулятивные итоги	% к итогу	кумулятивные итоги
до 100	35,00	35,0	3,3	3,3
101-200	19,6	54,6	5,4	8,7
201-500	22,9	77,5	14,2	22,9
501-1000	11,3	88,8	15,3	38,2
1001-3000	8,4	97,2	26,8	65,0
3001-10000	2,5	99,7	23,6	88,6
10001 и более	0,3	100,0	11,4	100,0

### РАЗДЕЛ 3. АБСОЛЮТНЫЕ И ОТНОСИТЕЛЬНЫЕ ВЕЛИЧИНЫ

#### 3.1. Абсолютные величины

Абсолютной величиной называется статистический показатель, выражающий абсолютные размеры изучаемого явления.

По способу выражения размеров изучаемых явлений абсолютные величины подразделяются на индивидуальные, которые отражают абсолютные размеры каждой отдельной единицы совокупности, и суммарные (групповые), которые получаются суммированием индивидуальных абсолютных величин.

Абсолютные величины - всегда числа именованные, имеющие определенную размерность, единицу измерения. В зависимости от различных причин и целей анализа применяют *натуральные, условно-натуральные, трудовые и стоимостные* единицы измерения.

**Натуральные** измерители в большинстве своем соответствуют природным или потребительским свойствам предмета изучения и выражаются в физических мерах веса, длины и т. д. Учет в натуральных единицах измерения называется натуральным учетом. Иногда применяется составная единица измерения (например, тоннокилометры - ткм и т. д.).

Однородная, но неодинаковая продукция измеряется в **условно-натуральных** единицах измерения. Такие единицы получают, приводя различные натуральные единицы к одной, принятой за базу, например, в консервной продукции емкость банки в 354,4 см<sup>3</sup> принята за условную и вся продукция пересчитывается в этих условных банках. Аналогично производится пересчет в условно-натуральные измерители и в других отраслях (топливной, текстильной и пр.).

Показатели, отражающие результаты труда, измеряются в *трудовых* единицах измерения, а наиболее распространенные – в *стоимостных* (или денежных) единицах измерения.

### **3.2. Относительные величины: сущность, виды, единицы измерения**

Статистические показатели, выражающие относительные размеры изучаемого явления, называются относительными величинами.

Относительные величины получают в результате деления одной величины на другую, принимаемую за базу сравнения (или основание относительной величины). При этом следует иметь в виду, что в числителе всегда находится показатель, отражающий то явление, которое изучает, т.е. сравнимый показатель, а в знаменателе - показатель, с которым производится сравнение.

База сравнения выступает в качестве своеобразного измерителя. Если значение базы принимается за единицу (производится простое деление двух величин), то относительная величина выражается в коэффициенте (например, темпы роста), если основание (база) - 100%, то относительная величина выражается в процентах.

Иногда за базу принимают 1000 (при расчете коэффициентов рождаемости или смертности, естественного или механического прироста показате-

ля, в фармацевтике и т. д.). Такие единицы измерения называются *промилле* и обозначаются ‰.

Все относительные величины делятся на относительные величины выполнения плана и планового задания (в настоящее время лучше говорить о выполнении договорных обязательств), динамики, структуры, координации, интенсивности, сравнения, уровня экономического развития.

**Относительная величина выполнения плана (ОВВП)** - соотношение величины показателя, достигнутое за какое-то время или какому-либо моменту, и величины его, установленной по плану за это же время:

$$\text{ОВВП} = \frac{\text{Факт (95)}}{\text{План}}$$

(95)

- факт (95) - фактическое выпол-

нение за 1995 г.;

- план (95) - плановое значение за 1995 г.

**Относительная величина планового задания (ОВПЗ)** - соотношение плановой величины показателя к его фактически достигнутому уровню в предыдущем периоде (или в периоде, принятом за базу):

$$З = \frac{\text{ОВП}}{\text{Факт (94)}}$$

(94)

- факт (94) - фактическое вы-

полнение за 1994 г.

Произведение этих двух величин дает относительную величину динамики (см. ниже):  $\text{ОВД} = \text{ОВВП} \cdot \text{ОВПЗ}$ ;

**Относительная величина динамики (ОВД)** - соотношение величины показателя за данное время и величины его за какое-либо аналогичное предшествующее время, принятое за базу сравнения. Относительная величина динамики характеризует степень, скорость изменения показателя во времени, в частности, темпы роста. Она выражается в виде краткого отношения (коэффициента) или в процентах. Если имеется ряд динамики (ряд во времени) абсолютных величин, то ОВД может исчисляться как отношение величины показателя за каждое последующее время к величине его за непосредственно предшествующее ему время или как отношение к величине его за одно и то же время, принятое за базу сравнения. В первом случае относительная величина динамики называется ОВД с переменной базой или цепная, во втором – с постоянной базой сравнения или базисная.

Если уровни ряда обозначить через  $Y_i$ , то ОВД вида:

$\frac{Y_2}{Y_1}; \frac{Y_3}{Y_2}; \dots; \frac{Y_i}{Y_{i-1}}; \dots; \frac{Y_n}{Y_{n-1}}$  называются цепными, а вида:

$\frac{Y_2}{Y_1}; \frac{Y_3}{Y_1}; \dots; \frac{Y_{i-1}}{Y_1}; \dots; \frac{Y_n}{Y_1}$ ; базисными (более подробно эта тема раскрывается в разделе рядов динамики).

**Относительная величина структуры** характеризует состав изучаемой совокупности и показывает какой удельный вес (какую долю) в общем итоге составляет каждая ее часть. Она получается в результате деления значения каждой части совокупности на их общий итог, принятой за базу.

**Относительная величина координации** характеризует соотношение отдельных частей целого, одна из которых принимается за базу сравнения и является дополнением к относительной величине структуры. Например, число сельских жителей, приходящихся на 100 городских и т. д.

**Относительная величина интенсивности** (степени) показывает степень распространенности данного явления в определенной среде. Обычно это отношение двух равноименных абсолютных величин (плотность населения на квадратный километр; производство сельскохозяйственной продукции на 1 га сельхозугодий и пр.).

**Относительная величина сравнения** характеризуют соотношение одноименных показателей, относящихся к различным объектам или территориям, но за один и тот же период времени.

**Относительная величина уровня экономического развития** – обычно соотношение величины важнейших экономических показателей (страны, региона, отрасли) и численности населения (так называемые показатели на душу населения).

### **Решение типовых задач**

**Задача 1.** Мыловаренный завод произвел за отчетный период следующее количество продукции, т.:

мыло хозяйственное 40%-ное – 25,0; мыло туалетное – 20,0;

мыло хозяйственное 60%-ное – 22,0; порошок стиральный – 55,0.

Определите общий выпуск продукции в пересчете на условное 40%-ное мыло по следующим переводным коэффициентам: мыло 60%-ное и туалетное – 1,75, порошок стиральный – 0,5.

**Решение.** Перемножив фактический выпуск продукции на коэффициенты пересчета в условное 40%-ное мыло, получим:

$$25 \cdot 1.0 + 22.0 \cdot 1.75 + 20.0 \cdot 1.75 + 55.0 \cdot 0.5 = \\ = 126 \text{ тыс. т условного 40\%-ного мыла.}$$

**Задача 2.** По плану предусматривалось собрать по 30 ц пшеницы с гектара, а собрано по 33 ц. Определить выполнение плана по урожайности.

Решение. Поделим фактическую урожайность на плановую и получим

$$\frac{33 \cdot 100}{30} = 110\%.$$

План по урожайности выполнен на 110%, или перевыполнен на 10%.

**Задача 3.** Планом предприятия предусматривалось повысить производительность труда на 5% и снизить затраты на 2%. Фактически производительность труда возросла на 6%, затраты были снижены на 4%. Определите выполнение плана по росту производительности труда и снижению затрат.

Решение. Степень выполнения плана по повышению производительности труда равна отношению достигнутого уровня - 106% (100% + 6%) к показателю плана - 105% (100% + 5%) в процентах

$$\frac{106 \cdot 100}{105} = 101\%.$$

Аналогично, фактическое снижение затрат - 96% (100% - 4%), а плановое - 98% (100% - 2%) их отношение дает 98%, т.е. план по снижению затрат перевыполнен на 2%.

**Задача 4.** Плановое задание по выпуску продукции на 1995 г. составило 104%, а выполнено на 105%. Определить относительную величину динамики.

Решение. Здесь 104% - это относительная величина планового задания, а 105% - это относительная величина выполнения плана, а так как ОВД = ОВВП·ОВПЗ, то  $\frac{104 \cdot 105}{100} = 109,2\%$ , т.е. выпуск продукции в 1995 г. составил по сравнению с 1994 годом 109,2% или увеличился на 9,2%.

**Задача 5.** По данным переписи населения 1970 г. в СССР проживало в возрасте 100 лет и старше 19304 чел., из них мужчин – 4252 чел., женщин – 15052. Определите структуру долгожителей по полу и относительную величину координации, приняв число мужчин за базу, равную 100.

Решение. Для вычисления относительных величин структуры нужно разделить значение каждой части на общий итог, принимаемый за целое (100%). Количество мужчин и женщин в возрасте 100 лет и старше надо поделить на общее их число или исчислить удельный вес мужчин, а затем отнять его от 100%:

$$\frac{4250 \cdot 100}{19304} = 22\% \text{ мужчин,}$$

женщин – 78% (100%-22%).

Относительные величины координации характеризуют соотношение отдельных частей совокупности. В задаче требуется вычислить соотношение между численностью мужчин и женщин в возрасте 100 лет и старше. Для этого, приняв число мужчин за базу сравнения, равную 100, поделим численность женщин на численность мужчин и получим

$$\frac{15052 \cdot 100}{4252} = 354.$$

Это значит, что на каждые 100 мужчин в возрасте 100 лет и старше приходится 354 женщины в том же возрасте.

## РАЗДЕЛ 4. СРЕДНИЕ ВЕЛИЧИНЫ

Средней величиной называется показатель, который дает обобщенную характеристику варьирующего признака единиц однородной совокупности.

Средняя отражает то общее, что скрывается в каждой единице совокупности, улавливает общие черты, общую тенденцию, закономерность, присущую данному распределению, она является равнодействующей, потому что в ней находит свое отражение, суммируется и синтезируется влияние всей совокупности факторов, под воздействием которых формируется ряд распределения. Средняя дает также характеристику центра распределения.

### 4.1. Средняя арифметическая

Обозначив индивидуальные значения признака через

$$x_1, x_2, x_3, \dots, x_n,$$

их количество - через  $n$ , можно записать:

$$\bar{x}_{\text{ар}} = \frac{\sum_{i=1}^n x_i}{n}.$$

Исчисленная таким образом средняя называется средней арифметической простой, т. е. она равна частному от деления суммы индивидуальных значений признака на их количество.

Средняя арифметическая простая применяется в тех случаях, когда каждое индивидуальное значение признака встречается один, или одинаковое количество раз, т. е. когда средняя рассчитывается по не сгруппированным данным.

В том случае, когда мы имеем дискретный ряд распределения, т. е. когда значение признака встречается несколько раз, применяют среднюю арифметическую взвешенную, рассчитываемую по формуле:

$$\bar{x}_{ар} = \frac{\sum xf}{\sum f}.$$

Таблица 4.1

Заработная плата рабочих цеха за месяц, тыс. руб.

з/ п (X)	10	20	30	40	50	60	70
ко л. (f)	2	5	20	30	15	10	5

По данным таблицы средняя заработная плата рассчитывается:

$(100 \cdot 2 + 200 \cdot 5 + 300 \cdot 20 + 400 \cdot 30 + 500 \cdot 15 + 600 \cdot 10 + 700 \cdot 5) / 87 = 36200 / 87 = 416$  тыс.руб.

Таким образом, средняя арифметическая взвешенная равна сумме произведений индивидуальных значений признака  $x$  на их частоты или веса  $f$ , поделенной на сумму частот  $\sum f$ .

Довольно часто в статистике приходится вычислять среднюю арифметическую в интервальном ряду, среднюю из групповых средних и среднюю из относительных величин.

При вычислении средней интервального ряда необходимо найти середину каждого интервала, и, взяв ее за значение признака использовать формулу средней арифметической взвешенной.

При вычислении средней из групповых (или частных) средних (например, при вычислении средней заработной платы по отрасли, когда имеется средняя заработная плата по предприятиям этой отрасли), в качестве индивидуальных значений признака берется соответствующая групповая средняя и рассчитывается по формуле средней арифметической взвешенной.

При вычислении средней из относительных величин в качестве весов (или частот) берут основание относительной величины, а в качестве значения признака - соответствующую этому основанию относительную величину.

### Свойства средней арифметической

1. Средняя арифметическая суммы варьирующих величин равна сумме средних арифметических этих величин.
2. Алгебраическая сумма отклонений индивидуальных значений признака от средней равно нулю.
3. Если все варианты ряда (значения признака) изменить на одно и то же число  $a$  или изменить в  $A$  раз, то и средняя изменится на  $a$  или в  $A$  раз соответственно.
4. Если все частоты ряда изменить в  $A$  раз, то средняя не изменится.

### 4.2. Средняя гармоническая

Во многих статистических исследованиях приходится сталкиваться с таким положением, когда известны значения индивидуального признака –  $x$  и произведения  $x \cdot f$ , т.е. действительные значения весов (частот) неизвестны. В этом случае расчет средней производится с использованием средней гармонической взвешенной, которая определяется по формуле:

$$\bar{x}_{\text{гарм}} = \frac{\sum z}{\sum \frac{z}{x}},$$

где  $Z = x \cdot f$ , т.е. произведению значения признака на частоту.

В тех случаях, когда произведения  $x \cdot f$  одинаковы или равны единице

$$x_1 \cdot f_1 = x_2 \cdot f_2 = \dots \dots = x_n \cdot f_n; \quad x_1 \cdot f_1 = x_2 \cdot f_2 = \dots \dots = x_n \cdot f_n = 1,$$

применяется средняя гармоническая простая, определяемая по формуле:

$$\bar{x}_{\text{гарм}} = \frac{n}{\sum \frac{1}{x}},$$

где  $x$  – отдельные варианты (значения признака);

$n$ - число наблюдений (общее число признаков или вариантов).

Таким образом, средняя гармоническая представляет собой особый вид средней, которая применяется в тех случаях, когда известны варианты  $x$  и произведения вариантов на частоты –  $x \cdot f$ , при отсутствии действительных весов.

### 4.3. Средняя геометрическая

В некоторых случаях приходится вычислять средний коэффициент роста в единицу времени. Коэффициент роста характеризует скорость изменения статистических показателей и представляет собой отношение величины показателя за два периода времени, как правило, за ряд смежных лет (табл. 4.2).

Таблица 4.2

Выпуск продукции предприятием за 1990 - 1993 гг. (у.д.е.)

Годы Показатели	1990	1991	1992	1993
Выпуск продукции, у	20.0	22.0	26.4	50.1
Коэффициент роста выпуска продукции по сравнению с предыдущим годом, к		22.0:20.0=1.1	26.4:22.0=1.2	50.1:26.4=1.9

Средняя, которая отражает средний коэффициент роста показателя за определенный период называется средней геометрической, которая равна корню степени  $m$  из произведений коэффициентов роста ( $m$  - число коэффициентов роста),

$$\bar{x}_{\text{геом}} = \sqrt[m]{\prod k}$$

Средний коэффициент роста (среднюю геометрическую) можно определить и по значениям первого и последнего членов динамического ряда. Если первый уровень ряда обозначить  $y_1$ , а последний -  $y_n$ , то

$$\bar{x}_{\text{геом}} = \sqrt[n-1]{\frac{y_n}{y_1}},$$

где  $n$  - число членов ряда (число лет).

#### 4.4. Средняя квадратическая

В тех случаях, когда осреднению подлежат величины, выраженные в виде квадратных функций, применяется средняя квадратическая. Средние

диаметры колес, труб, стволов, средние стороны квадратов и т.д. определяются при помощи средней квадратической.

Средняя квадратическая простая вычисляется путем извлечения квадратного корня из частного от деления суммы квадратов отдельных значений признака на их число:

$$\bar{x}_{\text{кв}} = \sqrt{\frac{\sum x^2}{n}},$$

#### 4.5. Соотношение между различными видами средних (мажорантность средних)

Все выше рассмотренные виды средних величин можно получить из формулы степенной средней вида

$$\bar{x} = \sqrt[k]{\frac{\sum x^k}{n}}.$$

При различных значениях показателя получаются различные средние, средняя арифметическая, гармоническая, геометрическая и квадратическая.

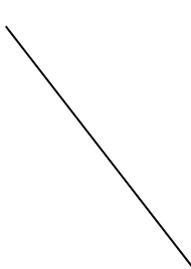
Мажорантность средних состоит в том, что средняя некоторого вида всегда больше средней некоторого другого вида (для признака, не могущего иметь отрицательных значений). В частности, если для одной совокупности вычислить средние гармоническую, геометрическую, арифметическую и квадратическую, то по численному значению они расположатся в возрастающем порядке. Порядок возрастания этих средних определяет показатель степени  $k$  в формуле степенной средней, т.е. чем больше  $k$ , тем больше средняя (табл.4.3).

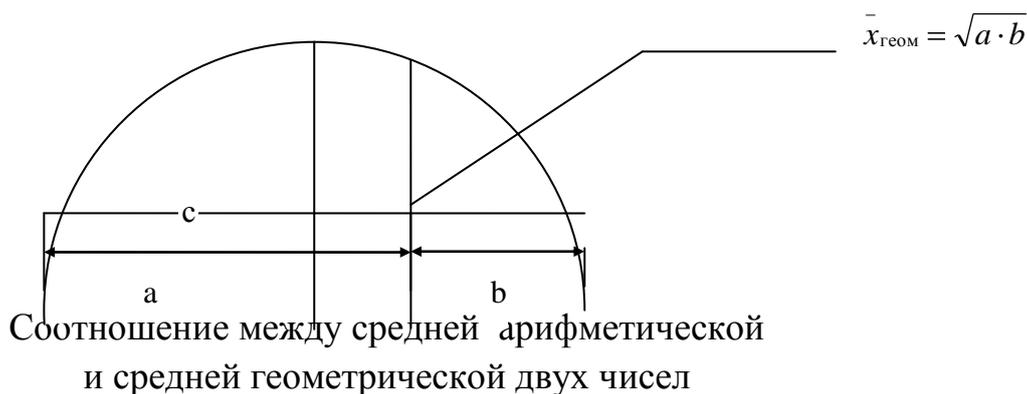
Таблица 4.3

Порядок возрастания средних

$k$	-1	0	1	2
Название средней	гармоническая	геометрическая	арифметическая	квадратическая

Подробно общее условие мажорантности т.е. соотношения между различными видами средних было сформулировано А. Я. Боярским. Проиллюстрируем правило мажорантности геометрически (см. рисунок).

$$\bar{x}_{\text{ар}} = \frac{a+b}{2}$$




Опишем на отрезке  $c = a + b$ , как на диаметре, полукруг и восстановим из общей точки отрезка  $a$  и  $b$  перпендикуляр, длина которого будет равна  $\sqrt{a \cdot b}$ . Восстановим также перпендикуляр из центра полукруга, длина которого будет равна  $\frac{a + b}{2}$ .

Из рисунка видно, что средняя геометрическая меньше средней арифметической. При этом, чем меньше отличаются отрезки  $a$  и  $b$  друг от друга, тем меньше разница между средними.

#### 4.6. Мода и медиана

Кроме средних в статистике для описательной характеристики величины варьирующего признака пользуются показателями *моды* и *медианы*.

Мода - это наиболее часто встречающийся вариант ряда. Мода применяется, например при определении размера одежды, обуви, пользующейся наибольшим спросом у покупателей, наиболее распространенной цены на тот или иной товар и пр.

Модой в дискретном ряду называется варианта (значение признака), имеющая наибольшую частоту (повторяющаяся самое большое количество раз), например: имеем данные о продаже магазином обуви по размерам (табл.4.4):

Таблица 4.4

Размер обуви $x$	3	3	3	3	3	3	3	3	3	3	Итого
Число пар $f$	3	4	<b>5</b>	6	7	8	7	9			500
Накоп-	0	8	<b>87</b>	4	2	6	3				

ленная	1	5	2	2	3	4	
частота	0	8	35	89	71	47	00

В этом примере модой является 35-й размер, так как обуви такого размера продано больше всего - 187 пар.

Если мы имеем интервальный ряд, то для определения моды необходимо сначала определить модальный интервал, причем, если интервалы равны, то модальный интервал определяется по наибольшей частоте, а если неравны, то по наибольшей плотности. При равных интервалах мода внутри модального интервала может определяться по формуле:

$$M_o = x_0 + h \frac{f_m - f_{m-1}}{(f_m - f_{m-1}) + (f_m + f_{m+1})},$$

где  $x_0$  - нижняя граница модального интервала;

$h$  - величина (ширина) интервала;

$f_m$  - частота модального интервала;

$f_{m-1}$  - частота интервала, предшествующего модальному;

$f_{m+1}$  - частота интервала, следующего за модальным.

Мода является наиболее распространенной и в этом смысле типичной величиной в распределении. Но мода и средняя величина по разному характеризуют совокупность. Мода определяет непосредственно размер признака, свойственный хотя и значительной части, но все же не всей совокупности. Поэтому мода по своему обобщающему значению уступает средней, которая характеризует совокупность в целом, так как складывается под воздействием всех без исключения элементов совокупности.

При наличии одной моды распределение называют унимодальным, при двух модах - бимодальным, при трех и более модах - мультимодальным.

**Медианой** называется варианта, которая приходится на середину ряда, расположенного в порядке возрастания или убывания численных значений признака. Медиана делит ряд на две равные части.

Если в совокупности нечетное число единиц, т. е.  $2m+1$ , то значение признака у  $(m+1)$ -ой единицы будет медианным. Если в совокупности четное число, т.е.  $2 \cdot m$  единиц, то медиана равна средней арифметической из двух срединных значений вариантов.

Для определения медианы в дискретном ряду при наличии частот сначала вычисляют полусумму частот, а затем определяют, какое значение признака приходится на нее. В примере с размером обуви (см. табл.4.4) медианой

является 35-й размер, так как именно он приходится на полусумму частот ( $500:2=250$ ). Это значит, что 35-й размер делит ряд на две равные части.

При вычислении медианы для интервального вариационного ряда вначале определяют медианный интервал, т.е. первый интервал, накопленная частота которого принимает полусумму частот, а затем приближенное значение медианы по формуле:

$$Me = x_0 + h \cdot \frac{\sum f_m - S_{m-1}}{f_m},$$

где  $f_m$  - сумма частот;

$x_0$  - нижняя граница медианного интервала;

$h$  - ширина интервала;

$S_{m-1}$  - сумма накопленных частот интервалов, предшествующих медианному;

$f_m$  - частота медианного интервала.

Главное свойство медианы состоит в том, что сумма абсолютных отклонений вариантов от медианы меньше, чем от любой другой величины (в том числе и от средней арифметической):

$$\sum |x - Me| = \min.$$

Медиану, являющуюся описательной характеристикой вариационного ряда, иногда называют непараметрической средней. Медиана меньше, чем средняя арифметическая, зависит от формы распределения признака. Она не зависит ни от амплитуды колебаний ряда, ни от распределения частот в пределах двух равных частей ряда. Вот почему в медиане не находят отражения важные свойства совокупности и она используется обычно для решения некоторых частных задач, связанных с определением оптимума, совпадающего с вариантом, приходящейся на середину ряда.

Мода и медиана являются описательными характеристиками совокупностей с количественно варьирующими признаками и не могут заменить среднюю обобщающую величину. Величина моды и медианы, как правило, отличается от величины средней, совпадая с ней только в случае симметрии вариационного ряда.

### **Решение типовых задач**

**Задача 1.** По следующим данным определите средний стаж рабочего (табл.4.5):

Таблица 4.5

Общий стаж ра- боты, лет	о 5	-10	0-15	5-20	0-25	25 и более	того
Число рабочих	0	0	0	2	8	10	00

Решение. Признаком в данной задаче является общий стаж рабочего, а частотами соответственно количество рабочих, имеющих тот или иной стаж. Ряд распределения – интервальный, причем первый и последний интервал – открытые.

Если интервалы открыты, то по правилам принимаем величину первого интервала равной второму, а последнего предпоследнему. Так как имеются и значения признака и частоты, то средний стаж находим по формуле средней арифметической взвешенной. А так как ряд интервальный, то в качестве значения признака в каждой группе берём середины интервала

$$\bar{x}_{ар} = \frac{\sum xf}{\sum f} = \frac{2,5 \cdot 70 + 7,5 \cdot 60 + 12,5 \cdot 20 + 17,5 \cdot 22 + 22,5 \cdot 18 + 27,5 \cdot 10}{200} = 9,7.$$

**Задача 2.** Все частоты уменьшились в два раза, а все варианты увеличились на две единицы. Что произойдет со средней?

Решение. Согласно свойствам средней арифметической, если все частоты ряда уменьшить или увеличить в одинаковое количество раз, то средняя не изменится, т.е. с точки зрения частот – средняя не изменится. Если все варианты увеличить или уменьшить на одно и то же число, то и средняя изменится на это же число.

В нашем случае средняя увеличится на две единицы.

**Задача 3.** Двое рабочих в течение 8-часового рабочего дня изготавливают одни и те же детали. Первый из них тратит на изготовление детали 30 мин., второй – 40 мин. Вычислите среднюю затрату времени на изготовление одной детали.

Решение. В этой задаче явно даны только значения признака – затраты времени, а частоты, которыми является количество изготовленных каждым рабочим деталей, в явном виде не присутствуют. Однако произведения значений признака на частоты дает количество проработанного времени – 8 час. Так как произведения признака на частоту равны, то средняя определяется по формуле средней гармонической простой:

$$\bar{x}_{\text{гарм}} = \frac{n}{\sum \frac{1}{x}} = \frac{2}{\frac{1}{30} + \frac{1}{40}} = 34 \text{ мин.}$$

**Задача 4.** Автомобиль проехал 1000 км, из них 480 км он прошел со скоростью 60 км/час, 320 – со скоростью 80 км/час и 200 км – со скоростью 50 км/час. Определите среднюю скорость, с которой совершался рейс.

Решение. В этой задаче опять известны только значения признака, а значения частот (время) не даны, однако имеются данные о пройденном расстоянии, которое является произведением признака на частоту. В этом случае средняя рассчитывается по формуле средней гармонической взвешенной:

$$\bar{x}_{\text{гарм}} = \frac{\sum z}{\sum \frac{z}{x}} = \frac{\sum xf}{\sum \frac{xf}{x}} = \frac{480 + 320 + 200}{\frac{480}{60} + \frac{320}{80} + \frac{200}{50}} = \frac{1000}{16} = 62,5 \text{ км/ч.}$$

**Задача 5.** Определите среднегодовой темп роста выпуска продукции на заводе, если в 1990 г. было произведено продукции на 21,15 у.д.е., а в 1995 г. было запланировано произвести продукции на 35 у.д.е.

Решение. Для определения средних темпов роста применяется средняя геометрическая. Когда имеются данные о первом периоде (в нашем случае – выпуск продукции в 1990 г. на сумму 21,15 у.д.е.) и в последнем периоде (в задаче — выпуск продукции по плану в 1995 г. на сумму 35 у.д.е.), среднегодовой темп роста определяется по формуле:

$$\bar{K} = \sqrt[n-1]{\frac{y_n}{y_1}} = \sqrt[n-1]{\frac{y_{95}}{y_{90}}} = \sqrt[6-1]{\frac{35,00}{21,15}} = \sqrt[5]{1,655} = 1,106.$$

**Задача 6.** Определить моду и медиану по следующим данным (табл 4.6):

Таблица 4.6

Распределение студентов заочного отделения по возрасту

Возрастные группы	Число студентов	Накопленные частоты
до 20 лет	346	346
20-25	872	1218
25-30	1054	2272
30-35	781	3053

35-40	212	3265
40-45	121	3386
45 лет и выше	76	3462
Итого:	3462	

Решение. Для определения моды определяем модальный интервал. Им является интервал 25-30 лет, так как его частота наибольшая (1054), тогда

$$M_o = 25 + 5 \frac{1054 - 872}{(1054 - 872) + (1054 + 781)} = 27 \text{ лет.}$$

Для определения медианы тоже необходимо определить медианный интервал. Медианным интервалом является интервал 25-30, так как он является первым интервалом, накопленная частота которого превышает полусумму частот ( $3462 : 2 = 1731$ ). Тогда медиана определится как:

$$M_e = 25 + 5 \frac{\frac{3462}{2} - 1218}{1054} = 27,4 \text{ года.}$$

## РАЗДЕЛ 5. ПОКАЗАТЕЛИ ВАРИАЦИИ И ДРУГИЕ ХАРАКТЕРИСТИКИ РЯДОВ РАСПРЕДЕЛЕНИЯ

### 5.1. Понятие о вариации

При изучении процессов и явлений общественной жизни статистика встречается с разнообразной вариацией признаков, характеризующей отдельные единицы совокупности. Величина признака изменяется, колеблется под влиянием различных причин и условий. Чем разнообразнее условия, влияющие на размер данного признака, тем больше его вариация.

Таблица 5.1

Дневная выработка рабочих двух бригад

Но мер бри- гады	Число рабочих	Дневная выработка деталей отдельных рабо- чих, шт.	Средняя дневная выработ- ка
1	6	75 90 78 82 93 86	84
2	6	65 122 84 70 105 58	84

Одна и та же средняя может характеризовать совокупность, в которой размеры вариации признака существенно отличаются друг от друга (табл.5.1). Как видно, средняя дневная выработка в обеих бригадах одинакова, хотя в первой бригаде средняя значительно меньше отличается от индивидуальных значений признака, чем во второй.

Следовательно, для всесторонней характеристики рядов распределения необходимы показатели, определяющие меру, степень колеблемости отдельных значений признака от средней, т.е. степень вариации, а также форму (тип) распределения, характеризующую ее закономерности.

### 5.2. Показатели вариации и способы их расчета

а) **Размах вариации** является наиболее простой мерой колеблемости значений признака и представляет собой разность между максимальным и минимальным значением признака:  $R = X_{\max} - X_{\min}$ .

Размах вариации имеет недостаток, проявляющийся в том, что при вычислении  $R$  используется только крайние значения ряда распределения, и он не всегда правильно характеризует колеблемость признака.

В связи с тем, что каждое индивидуальное значение признака отклоняется от средней на определенную величину, мерой вариации может служить средняя из отклонений каждой отдельной варианты от их средней.

Таковыми показателями являются среднее линейное отклонение, дисперсия, среднее квадратическое отклонение.

б) **Среднее линейное отклонение** представляет собой среднюю из абсолютных значений отклонений индивидуальных значений признаков от их средней:

$$d = \frac{\sum |x - \bar{x}|}{n}; \quad d = \frac{\sum |x - \bar{x}| \cdot f}{\sum f}.$$

Недостаток среднего линейного отклонения в том, что оно берется без учета знака. Поэтому в статистике чаще используют дисперсию и среднее квадратическое отклонение.

в) **Дисперсия и среднее квадратическое отклонение.** Дисперсией называется средний квадрат отклонений индивидуальных значений признака от их средней величины

$$\sigma^2 = \frac{\sum (x - \bar{x})^2}{n}; \quad \sigma^2 = \frac{\sum (x - \bar{x})^2 \cdot f}{\sum f}.$$

Дисперсия имеет очень большое значение в анализе. Однако ее применение как меры вариации в ряде случаев бывает не совсем удобным, потому что размерность дисперсии равна квадрату размерности изучаемого признака. Поэтому вычисляют среднее квадратическое отклонение, равное корню квадратному из суммы квадратов отклонений индивидуальных значений признака от их средней, т.е. из дисперсии

$$\sigma = \sqrt{\sigma^2}.$$

Сопоставление линейных или средних квадратических отклонений по нескольким совокупностям дает возможность определять степень их однородности в отношении того или иного признака. Чем меньше:  $R$ ,  $d$ ,  $\sigma^2$ ,  $\sigma$ , тем совокупность более однородна, тем более типичной будет средняя величина.

г) **Коэффициент вариации.** Так как дисперсия и среднее квадратическое отклонение характеризуют абсолютный размер отклонений, то представляет интерес сопоставить среднее квадратическое отклонение с его средней величиной. Такой показатель называется коэффициентом вариации

$$V = \frac{\sigma}{\bar{x}} \cdot 100\%.$$

Если взять отношение среднего линейного отклонения к средней арифметической, то получим *линейный* коэффициент вариации

$$V_{\text{л}} = \frac{d}{\bar{x}} \cdot 100\%.$$

Отношение размаха вариации к средней арифметической называется коэффициентом *осцилляции*:

$$K_{\text{o}} = \frac{R}{\bar{x}} \cdot 100\%.$$

### Взаимосвязь показателей вариации

В нормальном ряду распределения между показателями вариации имеются следующие примерные соотношения:

$$\sigma \cong \frac{R}{6}; \quad R \cong \bar{x} \pm 3 \cdot \sigma; \quad \sigma \cong 1,25 \cdot d.$$

### **5.3. Основные свойства дисперсии**

1) Если все значения признака уменьшить или увеличить на какое-то постоянное число  $a$ , то дисперсия не изменится.

2) Если все значения признака уменьшить или увеличить в  $K$  раз, то дисперсия изменится в  $K^2$  раз.

3) Сумма квадратов отклонений индивидуальных значений признака  $x$  от их средней  $\bar{x}$  меньше суммы квадратов отклонений индивидуальных значений признака от любого числа  $a$ , при условии, что  $a \neq \bar{x}$

$$\sum (x - \bar{x})^2 < \sum (x - a)^2.$$

4) Дисперсия признака равна разности между средним квадратом значения признака и квадратом их средней:

$$\sigma^2 = \overline{x^2} - (\bar{x})^2.$$

### Дисперсия альтернативного признака

В ряде случаев возникает необходимость измерить вариацию альтернативного признака, то есть такого, который может принимать только два значения. Обозначив отсутствие интересующего нас признака через 0, его наличие через 1, долю единиц, обладающих данным признаком – через  $p$ , не обладающих — через  $q$ , дисперсию этого признака можно определить как

$$\sigma^2 = p \cdot q$$

Например, если 64% работников предприятия имеют высшее образование  $p$ , то дисперсия будет равна:

$$\sigma^2 = 0,64 \cdot (-0,64) = 0,23.$$

#### 5.4. Правило сложения дисперсий

На вариацию признака влияют различные причины и факторы, которые делятся на случайные и систематические. Поэтому и вариация может быть случайной, вызванной действием случайных причин и систематической, обусловленной воздействием постоянных причин и факторов. В связи с этим возникает необходимость в определении случайной систематической составляющей и её роли в общей вариации. Общую дисперсию мы уже рассматривали. Она характеризует общую вариацию признака под влиянием всех условий, всех причин, вызывающих эту вариацию и исчисляется по формуле:

$$\sigma^2 = \frac{\sum (x - \bar{x})^2}{n}; \quad \sigma^2 = \frac{\sum (x - \bar{x})^2 \cdot f}{\sum f}.$$

Для определения влияния постоянного фактора на величину вариации пользуются аналитической группировкой. Вариация, обусловленная фактором, положенным в основание группировки, называется межгрупповой вариацией. Размеры ее определяются при помощи дисперсии групповых средних или межгрупповой дисперсии, которая характеризует колеблемость групповых или частных средних  $\bar{x}$  около их общей средней:

$$\delta^2 = \frac{\sum (x_i - \bar{x})^2}{n}; \quad \delta^2 = \frac{\sum (x_i - \bar{x})^2 \cdot f}{\sum f},$$

где  $\bar{x}_i$  - средняя по каждой отдельной группе;  $\bar{x}$  - средняя по всей совокупности;  $n$  - число единиц совокупности;  $f$  - частоты или веса.

Таким образом, межгрупповая дисперсия (дисперсия групповых средних) равна средней арифметической из квадратов отклонений частных средних от общей средней. Она характеризует систематическую вариацию, которая возникает под влиянием фактора, признака, положенного в основание группировки.

Для определения влияния случайных факторов и их роли в общей вариации определяют дисперсию в пределах каждой группы, т.е. внутригрупповую дисперсию, а затем и среднюю из внутригрупповых дисперсий:

$$\sigma_i^2 = \frac{\sum (x - x_i)^2}{n}; \quad \sigma_i^2 = \frac{\sum (x - x_i)^2 \cdot f}{\sum f}.$$

где  $x$  - индивидуальные значения признака;  $\bar{x}_i$  - групповые или частные средние:

$$\bar{\sigma}^2 = \frac{\sum \sigma_i^2}{n}; \quad \bar{\sigma}^2 = \frac{\sum \sigma_i^2 \cdot f}{\sum f}$$

В математической статистике доказано, что общая дисперсия признака равна сумме межгрупповой дисперсии и средней из внутригрупповых дисперсий

$$\sigma^2 = \bar{\sigma}^2 + \delta^2$$

Это правило называется **правилом сложения дисперсий**.

## 5.5. Ряды распределения и их моделирование

### а) Моменты распределения

Одной из важных задач анализов рядов распределения является выявление закономерности распределения, определение ее характера и количественного выражения. Эта задача решается при помощи показателей, характеризующих форму, тип распределения.

Кроме рассмотренных выше важной характеристикой рядов распределения являются моменты распределения.

Моментом распределения ( $M_k$ ) называется средняя арифметическая из отклонений значений признака  $x$  от некоторой постоянной величины  $a$  в степени  $k$ :

$$M_k = \frac{\sum (x - a)^k \cdot f}{\sum f}$$

Величина  $k$  определяет порядок момента. В зависимости от величины  $a$  различают начальные, центральные и условные моменты (табл. 5.2)

Таблица 5.2

## Значение моментов распределения

Ве- личина	На- звание мо- мента	Обо- значение мо- мента	Величина момента		
			$\kappa=0$	$\kappa=1$	$\kappa=2$
0	на- чальный	$M$	1	$\bar{x}$	$\bar{x}^2$
$\bar{x}$	цен- тральный	$m$	1	0	$\sigma$
$a$	ус- ловный	$M$	$M_k = \frac{\sum (x - a)^k \cdot f}{\sum f}$		

Условные моменты используются для определения дисперсий высоких степеней. Практически используются моменты первых четырех порядков. Если в качестве весов взять не частоты, а вероятности, то получим теоретические моменты распределения.

**б) Кривые распределения**

Для обобщенной характеристики особенностей формы распределения применяются кривые распределения, которые выражают закономерность распределения единиц совокупности по величине варьирующего признака.

Эмпирическая кривая - это фактическая кривая, полученная по данным наблюдения, в которой отражаются как общие, так и случайные условия, определяющие распределение (кривая 1, рис.5.3)

Теоретическая кривая распределения - это кривая, выражающая функциональную связь между изменением варьирующего признака и изменением частот и характеризующая определенный тип распределения (кривая 2, рис.5.3).

По форме кривые распределения бывают симметричными и асимметричными. В зависимости от того, какая ветвь кривой вытянута, различают *правостороннюю асимметрию* и *левостороннюю асимметрию* (рис.5.1). Кривые распределения также могут иметь различную островершинность (рис.5.2).

Для характеристики степени асимметрии кривой используют *коэффициент асимметрии*, который представляет собой отношение центрального момента третьего порядка к среднему квадратическому отклонению в кубе

$$A = \frac{m_3}{\sigma^3}.$$

Если  $A > 0$ , то асимметрия правосторонняя, а если  $A < 0$ , то асимметрия левосторонняя, в симметричном распределении –  $A=0$ . Кроме этого коэффициент для характеристики асимметрии применяют и соотношение между  $\bar{x}$  и модой или медианой по отношению к среднеквадратическому отклонению

$$A_{Mo} = \frac{\bar{x} - Mo}{\sigma}; \quad A_{Me} = \frac{\bar{x} - Me}{\sigma}.$$

Он менее точен по сравнению с коэффициентом асимметрии и применяется реже.

Для характеристики островершинности кривой распределения применяют *коэффициент эксцесса*, который равен отношению центрального момента четвертого порядка к дисперсии в квадрате

$$E = \frac{m_4}{\sigma^4}.$$

В нормальном распределении  $E=3$ , поэтому, если  $E>3$ , то эксцесс выше нормального (островершинная кривая),  $E<3$ , эксцесс ниже нормального (плосковершинная кривая).

### в) Моделирование рядов распределения

Все рассмотренные выше показатели характеризуют отдельные свойства совокупности. Общую характеристику ряда распределения можно представить аналитически, в виде функции, характеризующей зависимость между изменениями признака и частотами. Если имеется эмпирический ряд распределения, то необходимо найти функцию распределения, т. е. подобрать такую теоретическую кривую, которая наиболее полно бы раскрывала закономерность распределения. Нахождение функции кривой распределения называется **моделированием**.

Для аппроксимации (выравнивания) эмпирических кривых распределения в статистике часто пользуются нормальным распределением, функция которого

$$F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt,$$

где  $F(x)$  - интегральная функция распределения;  $t$  - нормированное от-

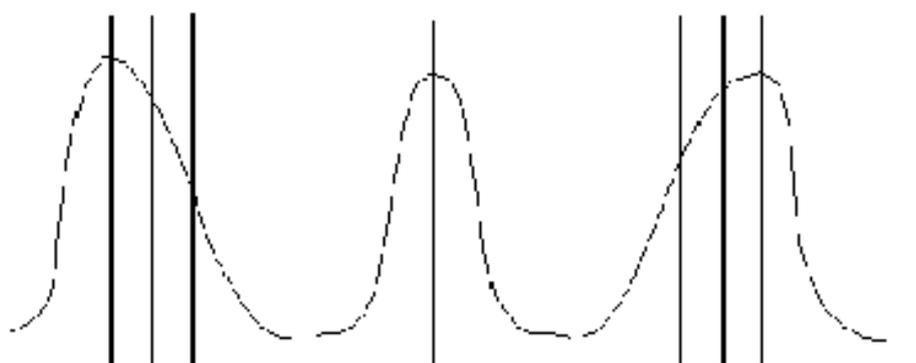
клонение;  $e$  - основание натуральных логарифмов.

Теоретическое распределение вероятностей и частот дает представление о форме, типе распределения, о закономерности, свойственной изучаемому явлению.

$$\bar{x} > Me > Mo$$

$$\bar{x} = Me = Mo$$

$$\bar{x} < Me < Mo$$



Правосторонняя  
Асимметрия

Симметричная

Левосторонняя  
кривая

асиммет-

рия

Рис.5.1. Типы кривых распределения (по асимметрии)



эксцесс выше  
нормального

нормальный  
эксцесс

эксцесс ниже  
нормального

Рис. 5.2. Типы кривых распределения (по островершинности)

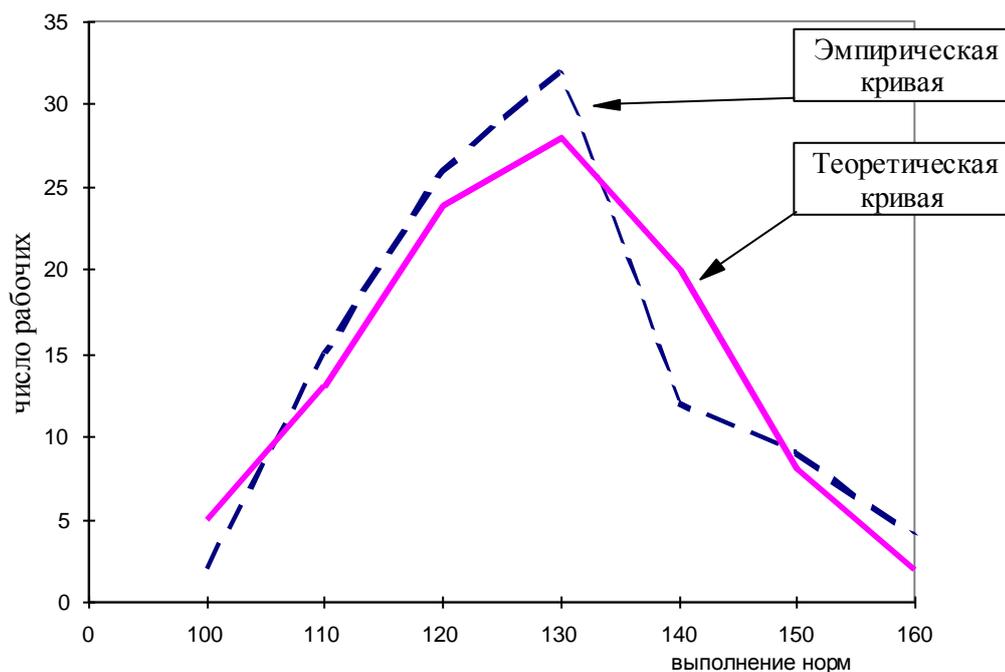


Рис. 5. 3. Эмпирическое и теоретическое распределение рабочих по степени выполнения

Эмпирическое и теоретическое распределение рабочих по степени выполнения норм приведено в табл.5.3 и на рис.5.3.

Таблица 5.3

Распределение рабочих по степени выполнения норм выработки

Группы рабочих по степени выполнения норм,%	Число рабочих (эмпирические частоты) (f)	Теоретические частоты (f')
до 100	2	5
100-110	15	13
110-120	26	24
120-130	32	28
130-140	12	20
140-150	9	8
150 и выше	4	2
Итого:	100	100

г) критерии согласия

Для оценки близости эмпирического и теоретического распределения используются специальные показатели, которые называются критериями согласия. Критерии согласия как правило тем или иным способом оценивают степень расхождения между эмпирическими и теоретическими частотами.

Наиболее часто используется критерий, который называется *хи-квадрат Пирсона* и рассчитывается

$$\chi^2 = \sum \left[ \frac{f - f'}{f'} \right]^2,$$

где  $f'$ -теоретические частоты,  $f$ -эмпирические частоты.

Чем меньше значение  $\chi^2$ , тем лучше теоретическое распределение отражает реальное положение в совокупности и наоборот. Если  $\chi^2=0$ , то расхождение между  $f$  и  $f'$  отсутствует.

Для проверки согласия с помощью  $\chi^2$  рассчитанное значение  $\chi_{\text{ф}}^2$  сравнивают с табличным и при условии  $\chi_{\text{ф}}^2 < \chi_{\text{т}}^2$  можно с определенной вероятностью сказать, что расхождение теоретического и эмпирического распределения случайно, и наоборот.

Также применяется и критерий согласия Колмогорова  $\lambda$ , который равен

$$\lambda = D: \sqrt{n},$$

где  $D$  - максимальная разность накопленных частот (без учета знака),  $n$  - объем совокупности.

Вычислив фактический критерий  $\lambda$ , по специальной таблице находят вероятность достижения критерием этого значения. Если вероятность значительна, то расхождение между эмпирическим и теоретическим распределением можно считать случайным.

В явлениях общественной жизни асимметричные распределения встречаются гораздо чаще, чем симметричные. Некоторые асимметричные распределения могут быть приведены к симметричному путем преобразования признака  $X$ , например логарифмированием. В этом случае распределение называется логарифмически нормальным. Такое преобразование производится, как правило, для сильно асимметричных распределений.

### Решение типовых задач

**Задача 1.** Проверка пряжи на крепость в лаборатории дала следующие результаты (табл. 5.4):

Таблица 5.4

Кре- пость	1	1	1	18	1	2	2
пря- жи, Г (x)	50	60	70	0	90	00	10
Ко- личество		1	2	31	2	4	6
проб, (f)	2	3	4		0		

Вычислить все показатели вариации.

Решение. В задаче значения признака имеют различную численность, поэтому значения,  $\bar{x}$ ,  $d$ ,  $\sigma^2$ ,  $\sigma$ ,  $V$  должны вычисляться как средние взвешенные величины. Для вычисления показателей вариации проводим дополнительные расчеты (табл. 5.5)

Таблица 5.5

x	f	xf	(x - $\bar{x}$ )	(x - $\bar{x}$ )f	(x - $\bar{x}$ ) <sup>2</sup>	(x - $\bar{x}$ ) <sup>2</sup> f
15	2	300	-	-58	841	1682
0			29			
16	13	208	-	-247	361	4693
0		0	19			
17	24	408	-	-216	81	1944
0		0	9			
18	31	558	1	31	1	31
0		0				
19	20	380	1	220	121	2420
0		0	1			
20	4	800	2	84	441	1764
0			1			
21	6	126	3	186	961	5766
0		0	1			
Ит	10	179		+52		18300
ого	0	00		1		
				-521		

Средняя крепость пряжи определяется по формуле

$$\bar{x} = \frac{\sum xf}{\sum f} = \frac{17900}{100} = 179 \text{ г.}$$

Размах вариации

$$R = x_{\max} - x_{\min} = 210 - 150 = 60 \text{ г.}$$

Среднее линейное отклонение вычисляется по формуле

$$d = \frac{\sum |x - \bar{x}| f}{\sum f} = \frac{521 + 521}{100} = 10,42 \text{ г.}$$

Дисперсия, взвешенная по частоте вариантов, равна

$$\sigma^2 = \frac{\sum (x - \bar{x})^2 f}{\sum f} = \frac{18300}{100} = 183 \text{ г}^2.$$

Среднее квадратичное отклонение равно:

$$\sigma = \sqrt{\sigma^2} = \sqrt{183} = 13,53 \text{ г.}$$

коэффициент вариации составляет

$$V = \frac{\sigma}{\bar{x}} \cdot 100\% = \frac{13,53}{179} \cdot 100\% = 7,6\%.$$

**Задача 2.** На предприятии 64% работников имеют среднее и высшее образование. Определить дисперсию доли рабочих, имеющих среднее и высшее образование.

Решение. Дисперсия альтернативного признака определяется как  $\sigma^2 = pq$ , где  $p$  – доля единиц совокупности, обладающих данным признаком, а  $q = 1 - p$  – доля единиц совокупности, не обладающих данным признаком, следовательно:  $\sigma^2 = 0,64 \cdot 0,36 = 0,2304$ .

**Задача 3.** Имеются следующие данные (табл.5.6):

Таблица 5.6

Заработная плата рабочих по профессиям (1997 год), тыс. руб.

п/п	Группы рабочих по профессиям	Число рабочих	З/п одного рабочего в месяц
	Токари	2	140, 160
	Слесари	3	120, 200, 160
	Кузнецы	5	140, 180, 200, 220, 260

Вычислить:

среднюю месячную заработную плату для каждой группы и для всех рабочих; групповые дисперсии; среднюю из внутригрупповых дисперсий; межгрупповую дисперсию; общую дисперсию; проверить правило сложения дисперсий и сделать выводы.

**Решение.** В задаче требуется определить, как влияют на изменение заработной платы рабочих постоянные причины, положенные в основание группировки (в данном случае – различия в профессии), и случайные причины. Для этого необходимо определить систематическую и случайную вариации и их роль в общей вариации. Для решения задачи и вычисления заданных показателей, построим вспомогательную таблицу (табл.5.7).

Таблица 5.7

№ п/п	$x$	$x - \bar{x}_i$	$(x - \bar{x}_i)^2$	$x - \bar{x}$	$(x - \bar{x})^2$
<b>Токари</b>					
1	140	-10	100	-38	144
2	160	10	100	-18	324
Итого:	300	0	200		
<b>Слесари</b>					
1	120	-40	1600	-58	336
2	200	40	1600	22	484
3	160	0	0	-18	324
Итого:	480		3200		
<b>Кузнецы</b>					
1	140	-60	3600	-38	144
2	180	-20	400	2	4
3	200	0	0	22	484
4	220	20	400	42	176
5	260	60	3600	82	672
Итого:	1000	0	8000		

<b>Все- го:</b>	1780	1140	163
		0	60

Среднюю заработную плату по каждой группе и для всех рабочих определяем по формуле средней арифметической простой:

$$\bar{x}_1 = \frac{300}{2} = 150 \text{ тыс.руб.} \quad \bar{x}_2 = \frac{480}{3} = 160 \text{ тыс.руб.} \quad \bar{x}_3 = \frac{1000}{5} = 200 \text{ тыс.руб.} \quad \bar{x} = \frac{1780}{10} = 178 \text{ тыс.руб.}$$

**Внутригрупповые дисперсии** вычисляем по формуле:

$$\sigma_i^2 = \frac{\sum (x_i - \bar{x})^2}{n}; \quad \sigma_1^2 = \frac{200}{2} = 100 \text{ тыс.руб}^2; \quad \sigma_2^2 = \frac{3200}{3} = 1067 \text{ тыс.руб}^2; \quad \sigma_3^2 = \frac{8000}{5} = 1600 \text{ тыс.руб}^2$$

**Средняя из внутригрупповых дисперсий** равна:

$$\bar{\sigma}^2 = \frac{\sum \sigma_i^2 \cdot f_i}{\sum f_i} = \frac{100 \cdot 2 + 1067 \cdot 3 + 1600 \cdot 5}{2 + 3 + 5} = 1140 \text{ тыс.руб}^2.$$

**Межгрупповая дисперсия** определяется как:

$$\delta^2 = \frac{\sum (x_i - \bar{x})^2 \cdot f_i}{\sum f_i}$$

$$\delta^2 = \frac{(50 - 178)^2 \cdot 2 + (60 - 178)^2 \cdot 3 + (200 - 178)^2 \cdot 5}{2 + 3 + 5} = \frac{4960}{10} = 496 \text{ тыс.руб}^2.$$

**Общая дисперсия** равна:

$$\sigma^2 = \frac{\sum (x_i - \bar{x})^2}{n} = \frac{16360}{10} = 1636 \text{ тыс.руб}^2.$$

По правилу сложения дисперсий общая дисперсия равна сумме средней внутригрупповых и межгрупповой дисперсии:

$$\sigma^2 = \bar{\sigma}^2 + \delta^2 = 1140 + 496 = 1636.$$

**Задача 4.** Средняя успеваемость студентов экономического факультета  $\bar{x}=4$  балла,  $M_o=4,8$  балла,  $M_e=4,2$  балла. Указать, симметричным или асимметричным является распределение студентов по успеваемости. Если асимметричным, то с какой асимметрией (левосторонней или правосторонней).

**Решение.** В симметричном распределении  $\bar{x}=M_o=M_e$ . В нашей задаче такого равенства нет. Следовательно, распределение асимметрично, а так как

$\bar{x} < Me < Mo$ , то разности между  $\bar{x} - Me$  и  $\bar{x} - Mo$  отрицательные и асимметрия левосторонняя.

**Задача 5.** Урожайность сахарной свеклы в сельхозпредприятиях области характеризуется следующими показателями:  $\bar{x} = 300$  ц/га,  $\sigma = 120$  ц/га,  $Mo = 240$  ц/га, а урожайность зерновых  $\bar{x} = 30$  ц/га;  $\sigma = 10$  ц/га,  $Mo = 27$  ц/га. Вычислить показатели асимметрии и указать где она больше. Каково направление асимметрии?

Решение.

$$P_{A_1} = \frac{\bar{x} - Mo}{\sigma} = \frac{300 - 240}{120} = 0,5; \quad P_{A_2} = \frac{30 - 27}{10} = 0,3$$

Асимметрия правосторонняя, так как  $P_A > 0$ , и больше в первом случае.

**Задача 6.** Центральный момент третьего порядка  $m_3 = 34,56$  при  $\sigma = 12$ . Вычислить коэффициент асимметрии и сделать выводы; центральный момент четвертого порядка  $m_4 = 64,32$  при  $\sigma = 2$ . Вычислить коэффициент эксцесса и сделать выводы.

Решение.

$$A = \frac{m_3}{\sigma^3} = \frac{34,56}{12^3} = 0,02$$

Асимметрия правосторонняя, так как  $A > 0$ , и незначительная

$$E = \frac{m_4}{\sigma^4} = \frac{64,32}{2^4} = 4,02$$

Эксцесс выше нормального, потому что  $E > 3$ .

## РАЗДЕЛ 6. ВЫБОРОЧНОЕ НАБЛЮДЕНИЕ

### 6.1. Общие сведения о выборочном наблюдении

В связи с тем, что статистика имеет дело с массовыми совокупностями статистические исследования весьма трудоемки. Поэтому давно возникла мысль о замене сплошного наблюдения выборочным.

Выборочное наблюдение – это наиболее совершенный способ несплошного наблюдения, при котором обследуется не вся совокупность, а лишь ее часть, отобранная по определенным правилам выборки и обеспечивающая получение данных, характеризующих всю совокупность в целом.

При проведении выборочного наблюдения нельзя получить абсолютно точные данные. Как при сплошном, так как при выборочном наблюдении не-

избежны ошибки, которые делятся на ошибки регистрации и ошибки репрезентативности. В свою очередь, ошибки репрезентативности бывают *случайные* и *систематические*.

Важнейшим условием применения выборочного метода является правильный отбор единиц совокупности, а именно:

а) строго объективный отбор единиц совокупности, при котором каждая из них получала бы абсолютно одинаковую возможность попасть в выборку;

б) достаточное количество отобранных единиц совокупности. При соблюдении этих условий выборка будет репрезентативной или представительной.

Вся совокупность единиц, из которой производится отбор, называется **генеральной совокупностью** и обозначается буквой  $N$ . Часть генеральной совокупности, попавшая в выборку, называется **выборочной совокупностью** и обозначается  $n$ .

Обобщающие показатели генеральной совокупности - средняя, дисперсия и доля - называются **генеральными** и соответственно обозначаются  $\bar{x}$ ,  $\sigma$ ,  $p$ , где  $p$  - доля или отношение числа единиц совокупности  $M$ , обладающих данным признаком ко всей численности генеральной совокупности, т.е.  $p = \frac{M}{N}$ . Те же обобщающие характеристики в выборочной совокупности обозначаются соответственно  $\tilde{x}$ ,  $\tilde{\sigma}^2$ ,  $\omega$ .

Теоретической основой выборочного метода является теорема П.Л.Чебышева, которая формулируется следующим образом: с вероятностью, сколь угодно близкой к единице (достоверности), можно утверждать, что при достаточно большом объеме выборки и ограниченной дисперсии генеральной совокупности разность между выборочной средней и генеральной средней будет сколь угодно мала:

$$|\bar{x} - \tilde{x}| < \varepsilon .$$

При практическом использовании теоремы Чебышева генеральную дисперсию  $\sigma^2$ , которая неизвестна, заменяют выборочной дисперсией.

## 6.2. Виды и схемы отбора

Формирование выборочной совокупности из генеральной может осуществляться по-разному. Различают следующие виды отбора: собственно-случайный; механический; типический; серийный; комбинированный.

1. **Собственно-случайный отбор.** Он ориентирован на выборку единиц из генеральной совокупности без всякого расчленения ее на части или

группы. Либо применяется жеребьевка, либо используются таблицы случайных чисел.

2. **Механический отбор.** Он состоит в том, что отбор единиц в выборочную совокупность производится из генеральной совокупности, разбитой на равные интервалы (группы), причем все единицы генеральной совокупности должны располагаться в определенном порядке. Размер интервала или группы равен обратной величине доли выборки (или количеству отбираемых единиц). Из каждой группы (интервала) берется только одна единица. Так, при 2%-ной выборке отбирается каждая 50-ая единица (1:0,02; или формируется 50 групп или интервалов), при 20%-ной выборке - каждая 5-я единица (1:0,2) и т. д.

3. **Типический отбор.** При его осуществлении вся генеральная совокупность делится на группы по типическому признаку, а затем в каждой группе проводится тот или иной отбор. Наиболее часто из каждой группы выбирается количество единиц, пропорциональное удельному весу группы в общей совокупности и как правило - механическим отбором. Такой отбор часто называют пропорциональным типическим отбором с механической выборкой.

4. **Серийный отбор с равновеликими сериями** состоит в выборке не отдельных единиц генеральной совокупности, а целых серий (гнезд). Попавшие в выборку серии подвергаются сплошному наблюдению. Сами серии могут формироваться различными методами и способами.

5. **Комбинированный отбор.** Все вышеперечисленные виды отбора комбинируются между собой.

Используя различные виды отбора, можно применить и различные схемы отбора: **бесповторный отбор** (схема невозвращенного шара) - после выбора какой-либо единицы она обратно в генеральную совокупность не попадает и не может быть выбрана снова; **повторный отбор** (схема возвращенного шара) - после выбора какой-либо единицы она вновь возвращается в генеральную совокупность и может быть выбрана снова.

### 6.3. Определение средних и предельных ошибок при различных видах отбора

**Ошибка выборки  $\varepsilon$**  – разность между характеристиками выборочной и генеральной совокупности.

Если  $\Delta$  представляет собой предел, которого не превосходит абсолютная величина  $\varepsilon$  то

$$\bar{x} - \Delta \leq \bar{x} \leq \bar{x} + \Delta;$$

$$\omega - \Delta \leq p \leq \omega + \Delta$$

Ошибка выборки зависит от многих факторов, и, если из одной и той же генеральной совокупности можно сформировать бесконечное множество выборочных совокупностей, каждая из них даст и свою ошибку. Поэтому при выборочном наблюдении говорят о средней из возможных ошибок (средняя или стандартная ошибка выборки), которую обозначают как  $\mu$ .

Величина  $\mu$  прямо пропорциональна корню квадратному из дисперсии и обратно пропорциональна квадратному корню из объема выборочной совокупности;

$$\mu_x = \sqrt{\frac{\sigma^2}{n}},$$

для доли, соответственно

$$\mu_\omega = \sqrt{\frac{p \cdot q}{n}}.$$

Эти формулы справедливы для повторной схемы отбора. При бесповторном отборе вводится поправочный коэффициент, равный

$$\left(1 - \frac{n}{N}\right), \text{ тогда } \mu_x = \sqrt{\frac{\sigma^2}{n} \left(1 - \frac{n}{N}\right)}; \quad \mu_\omega = \sqrt{\frac{p \cdot q}{n} \left(1 - \frac{n}{N}\right)}.$$

В тех случаях, когда выборка небольшая этим множителем можно пренебречь, так как его значение близко к единице (обычно при  $\frac{n}{N} \leq 0,05$ ).

Для решения практических задач важна не средняя ошибка выборки, а пределы, за которые она не выйдет, т.е. говорят о предельной ошибке выборки.

Предельная ошибка выборки связана со средней ошибкой соотношением  $\Delta = t \cdot \mu$ , где  $t$ - коэффициент доверия, или  $t$ - статистика;  $t$  принимает значения 1, 2 или 3 и связан с вероятностью достижения заданного предела. Если  $t=1$  это значит, что вероятность того, что ошибка выборки не превысит заданного значения, равна 0,683 или 68,3%. При  $t=2$  –  $P=0,954$  или 95,4%;  $t=3$  –  $P=0,997$  или 99,7%.

Таким образом, предельная ошибка выборки зависит от трех факторов: дисперсии  $\sigma^2$ , объема выборки  $n$ , и коэффициента доверия  $t$ . Формулы для определения предельных ошибок при различных видах и схемах отбора приведены в табл. 6.1.

Таблица 6.1

Определение предельной ошибки выборки при различных видах отбора

Схема отбора	Предельная ошибка выборки	
	для средней	для доли
Собственно - случайный и механический отбор		
Повторный отбор	$\Delta_x = t \sqrt{\frac{\sigma^2}{n}}$	$\Delta_\omega = t \sqrt{\frac{\omega(1-\omega)}{n}}$
Бесповторный	$\Delta_x = t \sqrt{\frac{\sigma^2}{n} \left(1 - \frac{n}{N}\right)}$	$\Delta_\omega = t \sqrt{\frac{\omega(1-\omega)}{n} \left(1 - \frac{n}{N}\right)}$
Типический отбор		
Повторный отбор	$\Delta_x = t \sqrt{\frac{\sigma^2}{n}}$	$\Delta_\omega = t \sqrt{\frac{\omega(1-\omega)}{n}}$
Бесповторный отбор	$\Delta_x = t \sqrt{\frac{\sigma^2}{n} \left(1 - \frac{n}{N}\right)}$	$\Delta_\omega = t \sqrt{\frac{\omega(1-\omega)}{n} \left(1 - \frac{n}{N}\right)}$
Серийный отбор		
Повторный отбор	$\Delta_x = t \sqrt{\frac{\delta^2}{s}}$	$\Delta_\omega = t \sqrt{\frac{\omega(1-\omega)}{s}}$
Бесповторный отбор	$\Delta_x = t \sqrt{\frac{\delta^2}{s} \left(1 - \frac{s}{S}\right)}$	$\Delta_\omega = t \sqrt{\frac{\omega(1-\omega)}{s} \left(1 - \frac{s}{S}\right)}$

#### 6.4. Определение необходимой численности выборки

При разработке выборочного наблюдения предполагают заранее заданными величину допустимой ошибки выборки и вероятность ответа (и  $t$ ). Неизвестным, следовательно, остается тот минимальный объем выборки, который должен обеспечить требуемую точность. Из формулы  $\Delta = \mu \cdot t$  и формул предельных ошибок выборки устанавливаем необходимую численность выборки. Формулы для определения численности выборки  $n$  зависят от способа отбора (табл.6.2).

Таблица 6.2

Формулы расчета численности выборки при собственно-случайном отборе

Способ отбора	Для средней	Для доли	Для доли, если даже приблизительно она неизвестна
Повторный	$\frac{t^2 \cdot \sigma^2}{\Delta^2}$	$\frac{t^2 \omega \left( \frac{1-\omega}{\omega} \right)}{\Delta^2}$	$\frac{0,25t^2}{\Delta^2}$
Бесповторный	$\frac{t^2 \sigma^2 N}{N\Delta^2 + t^2 \sigma^2}$	$\frac{t^2 \cdot N \cdot \omega \left( \frac{1-\omega}{\omega} \right)}{N \cdot \Delta^2 + t^2 \omega \left( \frac{1-\omega}{\omega} \right)}$	$\frac{0,25t^2 \cdot N}{N \cdot \Delta^2 + 0,25t^2}$

Численность выборки можно выразить через отношение

$$d = \frac{\Delta}{\sigma},$$

т.е. предельную ошибку выборки выражают в единицах среднего квадратичного отклонения. Так, например, формула численности бесповторной собственно-случайной выборки для средней при  $t=3$  примет вид

$$n = \frac{N}{N \cdot \left( \frac{d}{3} \right)^2 + 1}.$$

Для типического и серийного отбора объем выборки определяется по специальным формулам.

### **Решение типовых задач**

**Задача 1.** Методом случайной повторной выборки было взято для проверки на вес 200 шт. деталей. В результате был установлен средний вес детали - 30 г при среднеквадратическом отклонении равном 4 г. С вероятностью 0,954 требуется определить предел, в котором находится средний вес детали в генеральной совокупности.

**Решение.**

Предельная ошибка средней при собственно-случайном отборе (повторная выборка) определится по формуле

$$\Delta_x = t \sqrt{\frac{\sigma^2}{n}}.$$

Нам известно, что  $t=2$  (т.к.  $P=0.954$ );  $\sigma^2 = 4^2 = 16$ ;  $n=200$ , тогда

$$\Delta_x = 2 \sqrt{\frac{16}{200}} = \pm 0,56.$$

Следовательно, с вероятностью 0,954 можно утверждать, что средний вес детали в генеральной совокупности будет находиться в пределах

$$30 - 0,56 < \bar{x} < 30 + 0,56, \quad 29,44 < \bar{x} < 30,56.$$

**Задача 2.** Был проведен учет городского населения города *A* методом случайного бесповторного отбора. Из общей численности населения 500 тыс. человек было отобрано 500 тыс. и установлено, что 15% имеют возраст старше 60 лет. С вероятностью 0,683 определить предел, в котором находится доля жителей города *A* в возрасте старше 60 лет.

Решение.

Предельная ошибка доли при собственно-случайном бесповторном отборе определится как

$$\Delta_{\omega} = t \sqrt{\frac{\omega(1-\omega)}{n} \left(1 - \frac{n}{N}\right)}.$$

Здесь  $\omega = 0,15$ ;  $1-\omega = 0,85$ ;  $\omega(1-\omega) = 0,15 \cdot 0,85$ ;  $n=50$ ;  $N=500$ ;  $t=1$  ( $P=0,683$ ), тогда подставляя эти данные в формулу получим:

$$\Delta_{\omega} = 1 \sqrt{\frac{0,15 \cdot 0,85}{50} \left(1 - \frac{50}{500}\right)} = \pm 0,048.$$

Следовательно, с вероятностью 0,683 можно утверждать, что доля жителей старше 60 лет находится в пределах:

$$0,15 - 0,048 < P < 0,15 + 0,048; \text{ или } 10,2\% < P < 19,8\%.$$

**Задача 3.** Проведена 10%-ная типическая выборка пропорциональна численности отобранных групп (табл. 6.3).

Таблица 6.3

Группировка рабочих разных профессий по степени выполнения норм выработки

Группы рабочих	Число рабочих	Среднее выполнение норм, %	$\sigma_i$
Токари	40	98	2
Слесари	60	108	3
Фрезеровщики	50	104	5

Требуется с вероятностью 0,954 определить пределы, в которых находится средний процент выполнения норм рабочими завода в целом. Выборка бесповторная.

Решение: Вычислим общий средний процент выполнения норм выработки:

$$\tilde{x} = \frac{\sum \tilde{x}_i f_i}{\sum f_i} = \frac{98 \cdot 40 + 108 \cdot 60 + 104 \cdot 50}{40 + 60 + 50} = 104\% .$$

Далее определим среднюю из групповых дисперсий

$$\overline{\sigma^2} = \frac{\sum \sigma_i^2 \cdot f_i}{\sum f_i} = \frac{4 \cdot 40 + 9 \cdot 60 + 25 \cdot 50}{40 + 60 + 50} = \frac{160 + 540 + 1250}{150} = 13,0 .$$

Рассчитаем предельную ошибку выработки для типического отбора

$$\Delta_x = t \sqrt{\frac{\overline{\sigma^2}}{n} \left(1 - \frac{n}{N}\right)} = 2 \sqrt{\frac{13,0}{150} \left(1 - \frac{150}{1500}\right)} = \pm 0,18 .$$

( $N=1500$ , т.к. выборка 10%-ная).

Таким образом, с вероятностью 0,954 можно утверждать, что средний процент выполнения норм рабочими завода в целом находится в пределах

$$\bar{x} = \tilde{x} \pm \Delta_x \text{ или } \bar{x} = 104\% \pm 0,18$$

т.е. он не меньше 103,82% и не больше 104,18%.

**Задача 4.** Для определения средней урожайности сахарной свеклы в области проведена 20%-ная серийная бесповторная выборка, в которую вошло 5 районов из 25. Средняя урожайность по каждому отобранному району составила: 250, 260, 275, 280, 300 ц/га. Определить с вероятностью 0,954 пределы, в которых будет находиться средняя урожайность сахарной свеклы по области.

Решение. Найдем общую среднюю

$$\tilde{x} = \frac{250 \cdot 800 + 260 \cdot 1000 + \dots}{800 + 1000 + \dots} = 280 \text{ ц/га.}$$

Определим межсерийную дисперсию по формуле

$$\delta^2 = \frac{\sum (x_i - \tilde{x})^2 \cdot f_i}{\sum f_i} = \frac{(250 - 280)^2 \cdot 800 + (260 - 280)^2 \cdot 1000 + \dots}{800 + 1000 + 1200 + \dots} = 337 \text{ ц/га.}$$

Рассчитаем предельную ошибку выборки при серийном бесповторном отборе

$$\Delta_x = 2 \sqrt{\frac{337}{5} \left(1 - \frac{5}{25}\right)} = \pm 7,34 \text{ ц/га.}$$

Следовательно, с вероятностью 0,954 можно утверждать, что средняя урожайность сахарной свеклы в области будет находиться в пределах от 272,66 до 287,34 ц/га.

**Задача 5.** Предполагается, что партия деталей содержит 8 % брака. Определить необходимый объем выборки, чтобы с вероятностью 0,954 можно было установить долю брака с погрешностью не более 2%. Исследуемая партия содержит 5000 деталей.

Решение. По условию задачи  $t=2$ , доля бракованных деталей  $\omega=0,08$ ,

$1-\omega=0,92$ . Предельная ошибка доли по условию равна  $\Delta_{\omega}=0,02$ , а  $N=5000$ . Подставляем эти данные в формулу и получим

$$n = \frac{2^2 \cdot 0,08 \cdot 0,92 \cdot 5000}{0,02^2 \cdot 5000 + 2^2 \cdot 0,08 \cdot 0,92} = \frac{1472}{2,2944} \cong 642.$$

Чтобы с вероятностью 0,954 можно было утверждать, что предельная ошибка доли брака не превысит 2%, необходимо из 5000 деталей отобрать 642.

**Задача 6.** Что произойдет с предельной ошибкой выборки, если:

- а) дисперсия уменьшится в 4 раза;
- б) численность выборки увеличить в 9 раз;
- в) вероятность исчисления изменится с 0,683 до 0,997.

Решение. Из формулы для расчета предельной ошибки выборки

$$\Delta = t \sqrt{\frac{\sigma^2}{n}}$$

видно, что она:

а) прямо пропорциональна корню квадратному из дисперсии. Следовательно, если дисперсия уменьшится в 4 раза, то предельная ошибка уменьшится в 2 раза;

б) обратно пропорциональна корню квадратному из численности выборки. Следовательно, если объем выборки увеличится в 9 раз, то предельная ошибка уменьшится в 3 раза;

в) прямо пропорциональна вероятности исчисления, т.е. при увеличении  $P$  с 0,683 ( $t=1$ ) до 0,997 ( $t=3$ ) ошибка увеличится в 3 раза.

## РАЗДЕЛ 7. СТАТИСТИЧЕСКОЕ ИЗУЧЕНИЕ СВЯЗИ

### 7.1. Виды и формы взаимосвязи между явлениями

Одной из важнейших задач статистики является изучение, измерение и количественное выражение взаимосвязи между явлениями общественной жизни, установленной на основе качественного анализа.

Различают два вида связей: функциональную и корреляционную, обусловленные двумя типами закономерностей: динамическими и статистическими.

Для явлений, в которых проявляются динамические закономерности, характерна жесткая, механическая причинность, которая может быть выражена в виде уравнения, четкой зависимости и т.д. Такая зависимость называется функциональной. При функциональной связи каждому значению одной величины (аргумента) соответствует одно или несколько вполне определенных значений другой величины (функции).

В общественных процессах, в которых проявляются статистические закономерности, нет строгой зависимости между причиной и результатом и обычно не представляется возможным выявить строгую зависимость.

Связь, при которой каждому значению аргумента соответствует не одно, а несколько значений функции и между аргументом и функциями нельзя установить строгой зависимости называется корреляционной. Корреляционная зависимость проявляется только в средних величинах и выражает числовое соотношение между ними в виде тенденции к возрастанию или убыванию одной переменной величины при возрастании или убывании другой.

По направлению различают *прямую* и *обратную* связи.

По аналитическому выражению корреляционная связь может быть *прямолинейной* и *криволинейной*.

### 7.2. Основные приемы изучения взаимосвязей

а) **Метод параллельных рядов.** Чтобы установить связь между явлениями достаточно расположить полученные в результате сводки и обработки материалы в виде параллельных рядов и сопоставить их между собой.

б) **Балансовый метод.** Для характеристики взаимосвязи между явлениями в статистике широко применяется балансовый метод. Сущность его заключается в том, что данные взаимосвязанных показателей изображаются в виде таблицы и располагаются таким образом, чтобы итоги между отдельными

ми частями были равны, т.е. чтобы был баланс. Балансовый метод используется для характеристики взаимосвязи между производством и распределением продуктов, денежными доходами и расходами населения и т.д.

в) **Метод аналитических группировок.** При наличии массовых статистических данных для изучения и измерения взаимосвязей социально-экономических явлений широко пользуются методом аналитических группировок. Аналитические группировки позволяют установить наличие связи между двумя и более признаками и ее направление. Метод группировок сочетается с методом средних и относительных величин.

г) **Дисперсионный анализ.** Аналитические группировки при всей своей значимости не дают количественного выражения тесноты связи между признаками. Эта задача решается при помощи дисперсионного и корреляционного анализов.

Дисперсионный анализ дает, прежде всего, возможность определить роль систематической и случайной вариаций в общей вариации и, следовательно, установить роль изучаемого фактора в изменении результативного признака. Для этого пользуются правилом сложения дисперсий.

### 7.3. Корреляционный анализ

#### Определение формы связи

Изучение взаимосвязей между признаками статистической совокупности заключается в определении формы и количественной характеристики связи, а также степени тесноты связи. Корреляционный анализ и решает эти две основные задачи.

Первая задача заключается в определении формы связи, т.е. в установлении математической формы, в которой выражается данная связь.

Предварительный этап при установлении формы связи заключается в теоретическом анализе изучаемого явления, а также в представлении искомой связи графически. График, построенный по исходным данным, позволяет приблизительно определить: есть ли какая-то связь между явлениями; ее направление (прямая или обратная); примерную тесноту связи (естественно, что при графическом анализе используются только две переменные).

Применение методов корреляционного анализа дает возможность выражать связь между признаками аналитически - в виде уравнения - и придавать ей количественное выражение.

Другими словами необходимо найти зависимость вида  $y=f(x)$ , причем в качестве функции  $f(x)$  могут быть

полином 1-го порядка -  $\hat{y} = a_0 + a_1x$

полином 2-го порядка -  $\hat{y} = a_0 + a_1x + a_2x^2$

степенная функция -  $\hat{y} = a_0 \cdot x^{a_1}$

гиперболическая функция -  $\hat{y} = a_0 + \frac{a_1}{x}$

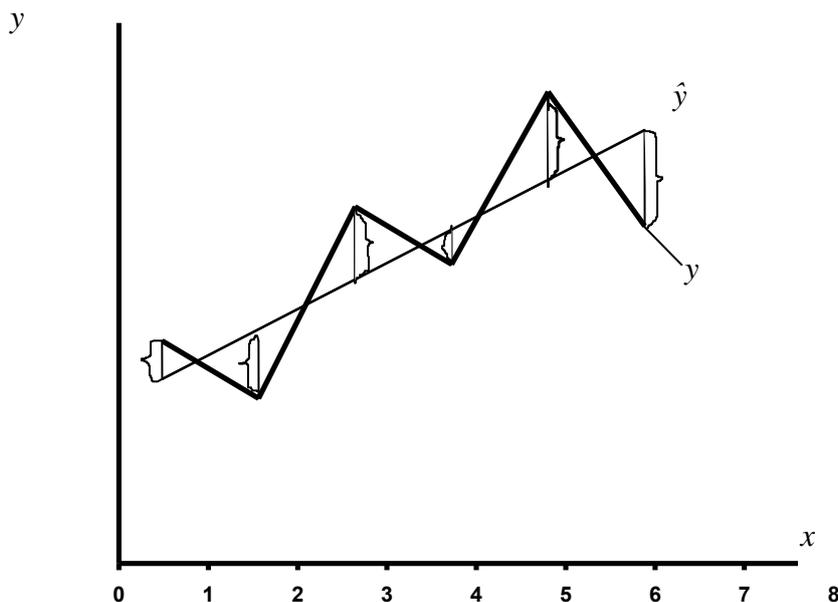
(могут быть использованы и другие виды функций).

Неизвестные параметры функций (аналитических уравнений связи) находятся методом наименьших квадратов, сущность которого в следующем: сумма квадратов отклонений фактических данных от выровненных должна быть наименьшей (см. рисунок):

$$(y_1 - \hat{y}_1)^2 + (y_2 - \hat{y}_2)^2 + \dots + (y_n - \hat{y}_n)^2 = S = \min$$

или

$$\sum (y - \hat{y})^2 \rightarrow \min$$



*Отклонение фактических уровней от выровненных*

### Измерение тесноты связи

При изучении корреляционной связи важно выяснить не только форму, но и тесноту связи между факторным и результативным признаком. Для этого (при прямолинейной связи) рассчитывается показатель, называемый парным линейным коэффициентом корреляции  $r_{xy}$ , вычисляемый по формуле

$$r_{xy} = \frac{\sum (x - \bar{x}) \cdot (y - \bar{y})}{\sqrt{\sum (x - \bar{x})^2 \cdot \sum (y - \bar{y})^2}}.$$

Коэффициент корреляции принимает значение от -1 до +1, причем если  $r_{xy} > 0$ , то корреляция прямая, если  $r_{xy} < 0$ , то корреляция обратная, а если  $r_{xy} = 0$ , то корреляция отсутствует полностью.

В зависимости от того, насколько  $r_{xy}$  приближается к единице, различают связь слабую, умеренную, заметную, высокую, тесную и весьма тесную.

Коэффициент корреляции может быть исчислен и по следующей формуле

$$r_{xy} = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\sigma_y \cdot \sigma_x},$$

где  $\sigma_y$  – среднее квадратическое отклонение результативного признака;

$\sigma_x$  – среднее квадратическое отклонение факторного признака.

Зная линейный коэффициент корреляции, можно определить и параметры уравнения регрессии вида  $y = a_0 + a_1 x$  потому что:

$$a_1 = r_{xy} \frac{\sigma_y}{\sigma_x} \quad \text{и} \quad a_0 = \bar{y} - a_1 \bar{x}.$$

Коэффициент корреляции  $r_{xy}$  применяется только в тех случаях, когда между явлениями существует прямолинейная связь. Если же связь криволинейная, то пользуются коэффициентом корреляции, вычисляемым по формуле

$$R = \sqrt{1 - \frac{\sum (y - y_T)^2}{\sum (y - \bar{y})^2}},$$

где  $y$  - исходные значения результативного показателя;

$y_T$  - теоретические значения;

$\bar{y}$  - среднее значение  $y$ .

Имея среднее значение дисперсий, коэффициент корреляции можно вычислить как

$$R = \sqrt{\frac{\sigma_B^2}{\sigma_0^2}} = \sqrt{1 - \frac{\sigma_{\text{ост}}^2}{\sigma_0^2}},$$

где  $\sigma_B^2 = \frac{\sum (y_T - \bar{y})^2}{n}$  – факторная (межгрупповая) дисперсия или дисперсия воспроизводимости;

дисперсия воспроизводимости;

$$\sigma_{\text{ост}}^2 = \frac{\sum (y_{\text{T}} - y)^2}{n} - \text{случайная (средняя из внутригрупповых) дисперсия}$$

или остаточная дисперсия;

$$\sigma_0^2 = \frac{\sum (y - \bar{y})^2}{n} - \text{общая дисперсия.}$$

Коэффициент корреляции по своему абсолютному значению находится в пределах от 0 до 1.

Если коэффициент корреляции возвести в квадрат и выразить в процентах, получим показатель, называемый коэффициентом детерминации

$$D = R^2 \cdot 100\%.$$

Он показывает, на сколько процентов изменение результативного фактора зависит от изменения факторного признака. Коэффициент детерминации является наиболее конкретным показателем, так как он отвечает на вопрос о том, какая доля в общем результате зависит от фактора, положенного в основании группировки.

#### 7.4. Множественная корреляция

Определение формы и тесноты связи между тремя и более параметрами называется множественной корреляцией. При множественной корреляции определение формы связи аналогично определению формы связи при парной корреляции, а само уравнение регрессии ищется в виде (как правило)

$$y = a_0 + a_1 x_1 + a_2 x_2 + \dots + a_n x_n.$$

При определении тесноты связи есть свои особенности. Теснота связи измеряется множественным коэффициентом корреляции, вид которого аналогичен коэффициенту корреляции при парной связи

$$R = \sqrt{1 - \frac{\sigma_{\text{ост}}^2}{\sigma_0^2}}.$$

Если изучается взаимодействие только трех факторов  $y = f(x, z)$ , то коэффициент множественной корреляции можно определить по формуле

$$r_{xy(z)} = \frac{r_{xy} - r_{zy} \cdot r_{xz}}{\sqrt{(1 - r_{zy}^2)(1 - r_{xz}^2)}}$$

где  $r_{xy}, r_{zy}, r_{xz}$  - парные коэффициенты корреляции.

Множественный коэффициент корреляции находится в пределах от 0 до 1.

Множественный коэффициент детерминации, равный квадрату  $R$ , выраженный в процентах, характеризует долю вариации результативного признака  $Y$  под воздействием всех изучаемых факторных признаков.

Поскольку факторные признаки действуют не изолированно, а по взаимосвязи, то может возникнуть задача определения тесноты связи между результативным признаком и одним из факторных при постоянных значениях прочих факторов. Она решается при помощи частных коэффициентов корреляции. Например, при линейной связи  $y=f(x,z)$  частный коэффициент корреляции между  $x$  и  $y$  при постоянном  $z$  вычисляется по следующей формуле

$$r_{xy(z)} = \frac{r_{xy} - r_{zy} \cdot r_{xz}}{\sqrt{(1 - r_{zy}^2)(1 - r_{xz}^2)}}$$

Частный коэффициент корреляции при изучении зависимости  $Y$  от  $Z$  при постоянном  $X$  определяется по формуле

$$r_{zy(x)} = \frac{r_{zy} - r_{xy} \cdot r_{xz}}{\sqrt{(1 - r_{xy}^2)(1 - r_{xz}^2)}}$$

Парные коэффициенты корреляции, как правило, выше частных. Это объясняется тем, что факторы взаимно коррелируют между собой.

При значительном количестве факторов частный коэффициент корреляции можно получить по формуле

$$R_{yx(z...v)} = \sqrt{\frac{R_{yxz...v}^2 - R_{yz...v}^2}{1 - R_{yz...v}^2}},$$

где  $R_{yxz...v}$  – коэффициент множественной корреляции;  $R_{yz...v}$  – коэффициент множественной корреляции результативного фактора ( $y$ ) со всеми за исключением исследуемого.

## 7.5. Простейшие методы измерения тесноты связи

Измерение тесноты связи между факторами с помощью корреляционно-регрессионного и дисперсионного анализов сопряжено с большими вычислительными трудностями. Для ориентировочной оценки степени тесноты связи существуют приближённые методы, не требующие трудоемких расчетов. К ним относятся: *коэффициент корреляции знаков Фехнера, коэффициент корреляции рангов, коэффициент ассоциации и коэффициент взаимной сопряженности*.

### 1. Коэффициент корреляции знаков.

Основан на сопоставлении знаков отклонений от средней и подсчете числа случаев совпадения и несовпадения знаков. Коэффициент корреляции знаков определяется по формуле

$$i = \frac{U - V}{U + V},$$

где  $U$  - число пар с одинаковыми знаками отклонений  $x$  и  $y$  от  $\bar{x}$  и  $\bar{y}$ ;

$V$  - число пар с разными знаками отклонений  $x$  и  $y$  от  $\bar{x}$  и  $\bar{y}$ .

Коэффициент корреляции знаков колеблется от -1 до +1. Этот показатель исчисляется очень просто, но именно в силу этого он не очень точен.

**2. Коэффициент корреляции рангов.** Этот показатель вычисляется не по первичным данным, а по рангам (порядковым номерам), которые присваиваются всем значениям изучаемых признаков, расположенным в порядке их возрастания. Если значения признака совпадают, то определяется средний ранг путем деления суммы рангов на число совпадающих значений. Коэффициент корреляции рангов определяется по формуле

$$\rho = \lambda - \frac{6 \sum d^2}{n(n^2 - 1)},$$

где  $d^2$  - квадрат разности рангов для каждой единицы  $d=x-y$ .

Коэффициент корреляции рангов также колеблется в пределах от -1 до +1.

### 3. Коэффициент ассоциации.

Коэффициент ассоциации применяется для установления меры связи между двумя качественными альтернативными признаками. Для его вычисления строится комбинационная 4-клеточная таблица:


которая выражает связь между двумя альтернативными явлениями.

Коэффициент ассоциации рассчитывается по формуле

$$A_c = \frac{ad - bc}{\sqrt{(a+b)(c+d)(a+c)(b+d)}}.$$

Коэффициент ассоциации тоже колеблется в пределах от -1 до +1.

### 4. Коэффициент взаимной сопряженности

В тех случаях, когда требуется установить связь между качественными признаками, каждый из которых состоит из трех и более групп, применяется коэффициент взаимной сопряженности. Для определения степени тесноты

связи вычисляется специальный показатель, который называется коэффициентом взаимной сопряженности. Он определяется по формуле:

$$C = \sqrt{\frac{\chi^2}{n \cdot (m_1 - 1)(m_2 - 1)}}$$

где  $n$  - число единиц совокупности;

$m_1$  и  $m_2$  - число групп по первому и второму признаку;

$\chi^2$  - показатель абсолютной квадратичной сопряженности Пирсона.

Методика применения всех четырех коэффициентов показана при решении типовых задач.

### **Решение типовых задач**

**Задача 1.** По данным о месячной заработной плате 10 рабочих трех разных профессий (токарь, слесарь и кузнец) вычислены: общая дисперсия заработной платы  $\sigma_0^2 = 1636$  и средняя из внутригрупповых дисперсий  $\overline{\sigma^2} = 1140$ . Вычислить корреляционное отношение.

Решение. Корреляционное отношение вычисляется по формуле

$$\eta^2 = \frac{\delta^2}{\sigma_0^2}.$$

Следовательно, сначала необходимо найти межгрупповую дисперсию

$$\delta^2 = \sigma_0^2 - \overline{\sigma^2} = 1636 - 1140 = 496.$$

Подставляя это значение в вышеприведенную формулу, получим:

$$\eta^2 = \frac{496}{1636} = 0,303.$$

**Задача 2.** Вычислено следующее уравнение множественной регрессии по сельскохозяйственным предприятиям области, характеризующее связь между размерами предприятия по посевной площади  $x$  га, числу коров  $z$  шт., доходами предприятия  $y$  млн. руб.:  $y = 290 + 3,12x + 0,65z$ . Требуется проанализировать параметры уравнения регрессии.

Решение. Увеличение посевной площади на один гектар приводит к повышению доходов в среднем на 3,12 млн руб.; увеличение числа коров на 1 голову приводит к повышению доходов на 0,65 млн руб.

**Задача 3.** по данным задачи №2 получены следующие парные коэффициенты корреляции  $r_{xy} = 0,835$ ;  $r_{zy} = 0,724$ ;  $r_{xz} = 0,900$ .

Вычислить коэффициент множественной корреляции и детерминации.

Решение.

$$R_{yxz} = \sqrt{\frac{r_{xy}^2 + r_{zy}^2 - 2 \cdot r_{xy} \cdot r_{zy} \cdot r_{xz}}{1 - r_{xz}^2}},$$

$$R_{yxz} = \sqrt{\frac{0,835^2 + 0,724^2 - 2 \cdot 0,835 \cdot 0,724 \cdot 0,90}{1 - 0,9^2}} = 0,843.$$

Таким образом, между доходами, посевной площадью и числом коров существует тесная связь. Коэффициент детерминации в нашем случае равен

$$D = R_{yxz}^2 \cdot 100\% = 0,84^2 \cdot 100 = 70\%.$$

Это значит, что 70% всей вариации денежных доходов предприятий объясняется действием рассматриваемых факторов.

**Задача 4.** По данным задачи №3 вычислите коэффициент корреляции между  $y$  и  $x$  и сделайте выводы.

**Решение.** Частный коэффициент корреляции характеризует тесноту связи между результативным признаком и одним из факторных при постоянных значениях прочих факторов, т. е.

$$r_{xy(z)} = \frac{r_{xy} - r_{zy} \cdot r_{xz}}{\sqrt{(1 - r_{zy}^2)(1 - r_{xz}^2)}},$$

где  $r_{xy}$ , - парные коэффициенты корреляции. Подставляя в формулу их значения, получим:

$$r_{xy(z)} = \frac{0,835 - 0,724 \cdot 0,9}{\sqrt{(1 - 0,724^2)(1 - 0,9^2)}} = 0,596.$$

Следовательно, связь между денежными доходами и посевной площадью при исключении влияния других факторов заметная.

Этот же результат можно получить, используя общую формулу коэффициента частной корреляции

$$r_{xy(z)} = \sqrt{\frac{r_{yxz}^2 - r_{yz}^2}{1 - r_{yz}^2}}.$$

Подставляя имеющиеся значения, получим:

$$r_{xy(z)} = \sqrt{\frac{0,843^2 - 0,724^2}{1 - 0,724^2}} = 0,596.$$

**Задача 5.** По 10 предприятиям имеются следующие данные, млн руб. (табл.7.1).

Таблица 7.1

Вы- пуск продук- ции	,5	,0	,0	,0	,0	,5	,0	,0	,0	0,0
Фонд заработной платы	,6	,9	,5	,8	,25	,31	,11	,05	,0	,5

Определите коэффициент корреляции знаков и сделайте выводы о тесноте связи.

Решение. Для вычисления коэффициента корреляции знаков составим вспомогательную таблицу (табл. 7.2), из которой получим число совпадений знаков от средней. В нашем случае оно равно 8, число несовпадений знаков отклонений от средней – 2.

Таблица 7.2

№ пп	Выпуск продукции	Фонд з/пл	Знаки отклонения от средней	
			$(x - \bar{x})$	$(y - \bar{y})$
1	1,0	0,11	---	---
2	1,5	0,31	---	---
3	2,0	0,25	---	---
4	2,5	0,6	---	---
5	3,0	0,8	---	---
6	4,0	1,05	---	+
7	5,0	0,9	+	---
8	6,0	1,5	+	+
9	8,0	2,0	+	+
10	10,0	2,5	+	+
Ит ого	43,0	10,0		
Ср едняя	4,3	1,0		

Подставляя эти значения в формулу коэффициента корреляции, получим:

$$i = \frac{U - V}{U + V} = \frac{8 - 2}{8 + 2} = 0,6.$$

Связь между выпуском продукции и фондом заработной платы прямая и заметная.

**Задача 6.** По 10 предприятиям имеются следующие данные (табл 7.3):

Таблица 7.3

Среднесписочная численность	00	50	100	400	750	20	000	00	300	000
Выпуск продукции	,0	,2	,5	,0	0,0	,8	,0	,0	,5	1,0

Решение. Коэффициент корреляции рангов вычисляется по рангам (порядковым номерам), которые присваиваются всем единицам совокупности, расположенным в порядке возрастания признака. Для его вычисления составим вспомогательную таблицу (табл 7.4).

Таблица 7.4

п/п	Выпуск продукции $x$	Численность работников $y$	Ранги		Разность рангов $d$	Квадрат разности рангов $d^2$
			по $x$	по $y$		
	5,0	800	4	3	1	1
	4,2	950	3	4	-1	1
	6,5	1100	6	6	0	0
	8,0	1400	7	7	0	0
	10,0	1750	9	8	1	1
	1,8	520	1	2	-1	1
	6,0	1000	5	5	0	0
	2,0	400	2	1	1	1
	9,5	2300	8	10	-2	4
	11,0	2000	10	9	1	1
0						
ТОГО						10

Подставляя соответствующие данные в формулу коэффициента, получим:

$$\rho = 1 - \frac{6 \sum d^2}{n \cdot (n^2 - 1)} = 1 - \frac{6 \cdot 10}{10 \cdot (100 - 1)} = 0,94,$$

то есть связь между выпуском продукции и среднесписочной численностью работников весьма тесная.

**Задача 7.** По данным табл. 7.5 вычислить коэффициент ассоциации.

Таблица 7.5

Зависимость выполнения плана рабочими от наличия специальной подготовки

	Прошли спецподготовку	Не прошли спецподготовку	Итого
Выполнили план	70 ( <i>a</i> )	20 ( <i>b</i> )	90 ( <i>a+b</i> )
Не выполнили план	2 ( <i>c</i> )	8 ( <i>d</i> )	10 ( <i>c+d</i> )
Всего	72 ( <i>a+c</i> )	28 ( <i>b+d</i> )	100

Решение. Коэффициент ассоциации используется для установления степени тесноты связи между двумя качественными альтернативными признаками, состоящими из двух групп, и определяется по следующей формуле

$$A = \frac{ad - bc}{\sqrt{(a+b)(c+d)(a+c)(b+d)}}.$$

Подставим соответствующие данные из табл.7.5 в формулу и получим:

$$A = \frac{70 \cdot 8 - 20 \cdot 2}{\sqrt{90 \cdot 10 \cdot 72 \cdot 28}} = 0,39.$$

Как видно, между выполнением плана и наличием специальной подготовки существует прямая и заметная связь.

## РАЗДЕЛ 8. РЯДЫ ДИНАМИКИ

### 8.1. Понятие о динамических рядах и их видах

Процессы и явления общественной жизни, которые изучаются статистикой, находятся в постоянном движении и изменении. В процессе развития меняются размеры, состав, объем, структура конкретных общественных явлений. Эти изменения статистика выражает при помощи различных статистических показателей.

Статистические данные, характеризующие изменения явлений во времени, называются динамическими (хронологическими или временными) рядами. Такие ряды имеют огромное значение для выявления и изучения складывающихся закономерностей в явлениях общественной жизни.

Довольно часто имеющиеся динамические ряды несопоставимы в силу изменения круга объектов учета, территориальных границ, изменения масштаба единиц измерения и т.д. В этом случае для преобразования несопоставимых динамических рядов в сопоставимые используют различные приемы, основные из которых следующие: прямой пересчет данных, пересчеты при помощи ключей и смыкание рядов.

В зависимости от характера изучаемых величин различают три вида динамических рядов: моментные, интервальные и ряды средних.

**Моментными рядами** называются ряды статистических величин, характеризующие размеры изучаемого явления на определенные даты или моменты времени. Примером могут служить данные о среднесписочной численности работающих по состоянию на первое число каждого месяца.

Отличительной особенностью моментных рядов является то, что они не подлежат суммированию.

**Интервальными рядами** называются ряды статистических показателей, характеризующих размеры изучаемого явления за определенные промежутки (периоды, интервалы) времени. Интервальные ряды можно суммировать.

**Ряды средних величин** - это ряды, характеризующие изменения средних уровней изучаемого явления во времени. Как и моментные, ряды средних величин не подлежат суммированию.

### Вычисление средней динамического ряда

Средняя, вычисленная из уровней динамического ряда, называется хронологической средней. Способы ее расчета зависят от вида динамического ряда.

а) для интервальных рядов средняя исчисляется по формуле средней арифметической, причем при равных интервалах применяется средняя арифметическая простая, а при неравных - средняя арифметическая взвешенная.

б) для моментных рядов средняя рассчитывается по формуле

$$\bar{y} = \frac{\frac{y_1}{2} + y_2 + \dots + y_{n-1} + \frac{y_n}{2}}{n - 1},$$

т.е. средняя хронологическая моментного ряда равна сумме всех уровней ряда, поделенной на число членов ряда без одного, причем первый и последний члены ряда берутся в половинном размере.

Если интервалы между периодами неравные, то применяется средняя арифметическая взвешенная, а в качестве весов берутся отрезки времени между датами, к которым относятся парные средние смежных значений уровня.

### **8.2. Основные показатели, используемые при анализе динамических рядов**

Динамические ряды анализируются при помощи ряда показателей, определяющих характер, направление, интенсивность количественных изменений во времени. К ним относятся: уровень ряда, средний уровень, абсолютный прирост, темп роста, коэффициент роста, темп прироста, коэффициент опережения, абсолютное значение одного процента прироста.

**Уровнем ряда** называется абсолютная величина каждого члена динамического ряда. Различают начальный (величина первого члена ряда), конечный (последнего), средний уровень ряда.

**Средний уровень** определяется в зависимости от вида динамического ряда.

**Абсолютный прирост** характеризует размер увеличения или уменьшения изучаемого явления за определенный период времени. Он определяется как разность между данным уровнем и предыдущим или начальным. Уровень, который сравнивается, называется текущим, а уровень с которым производится сравнение, называется базисным. Если каждый уровень ряда сравнивается с предыдущим, то получаются цепные показатели. Если же все уровни

ряда сравниваются с одним и тем же, первоначальным уровнем, то полученные показатели называются базисными.

Абсолютный прирост определяется по формулам;

цепной:  $-\Delta_i = y_i - y_{i-1}$ ; базисный:  $-\Delta = y_i - y_0$ ,

где  $y_i$  – текущий уровень ряда;  $y_{i-1}$  – уровень предшествующий;  $y_0$  – начальный уровень ряда.

**Темпом роста** называется отношение данного уровня к предыдущему или начальному, выраженному в процентах. Темпы роста бывают цепными и базисными и вычисляются по формулам

цепной:  $K_i = \frac{y_i}{y_{i-1}} \cdot 100$ ; базисный:  $-K = \frac{y_i}{y_0} \cdot 100$ ;

Если темпы роста выражены в виде простых отношений (база-1), то полученные показатели называются коэффициентами роста.

**Темпом прироста** называется отношение абсолютного прироста к предыдущему или начальному членам ряда, выраженным в процентах;

цепной:  $T_i = \frac{\Delta_i}{y_{i-1}} \cdot 100$ ; базисный  $-T = \frac{\Delta}{y_0} \cdot 100$ .

Темп прироста также может быть рассчитан как:

цепной:  $T_i = K_i - 1$ ; базисный  $-T = K - 1$ .

Для характеристики темпов роста и прироста в среднем за весь период вычисляют средний темп роста и прироста. Средний темп, коэффициент роста и прироста определяются по формулам средней геометрической.

Для определения средней из средних коэффициентов или темпов роста за неодинаковые промежутки времени применяется средняя геометрическая взвешенная

$$\bar{K} = \sum m_i \sqrt[m_i]{\prod K_i^{m_i}}$$

где  $m$  - продолжительность отрезков времени.

**Коэффициентом опережения** называется отношение базисных темпов роста двух динамических рядов за одинаковые отрезки времени

$$K_{оп} = \frac{K'}{K''},$$

где  $K'$  - базисные темпы роста первого ряда;  $K''$  - базисные темпы роста второго ряда.

В тех случаях, когда темпы роста по двум сравниваемым рядам динамики неизвестны, а имеются средние темпы роста за одинаковый период времени, коэффициент опережения рассчитывается по формуле

$$K_{\text{оп}} = \frac{\bar{K}'^n}{\bar{K}''^n},$$

где  $\bar{K}'$  – средний темп роста первого ряда динамики,  $\bar{K}''$  – второго, а  $n$  – число лет в периоде.

Отношение абсолютного прироста к темпу прироста представляет собой **абсолютное значение одного процента прироста** и определяется по формуле

$$A \% = \frac{\Delta_i}{T_i} = \frac{y_{i-1}}{100},$$

где  $A \%$  - абсолютный прирост;  $T_i$  - цепной темп прироста;  $y_{i-1}$  - уровень, предшествующий  $y_i$ .

Из формулы видно, что абсолютное значение одного процента прироста равно одной сотой части предшествующего уровня.

### **8.3. Важнейшие приемы обработки и анализа динамических рядов**

Существуют различные приемы обработки динамических рядов:

#### **а) Приведение рядов к одному основанию.**

Для выявления связи или различия в динамике двух или нескольких рядов их можно привести к одному основанию. Для этого показатели каждого ряда выражаются в процентах к первому или любому другому члену ряда.

#### **б) Разбивка ряда на короткие периоды.**

Для выявления тенденции данных колеблющихся рядов их разбивают на более короткие периоды, а затем определяют средний уровень по каждому периоду.

#### **в) Сглаживание способом скользящих (подвижных) средних.**

Сущность его заключается в том, что по конкретным уровням ряда рассчитываются сглаженные, скользящие средние, которые получаются из подвижных сумм путем последовательного сдвига на одну дату суммируемых показателей. Затем подвижные суммы делят на число дат, получая, таким образом, скользящие или подвижные средние. Например, складывают три первых члена ряда, а их среднюю относят ко второму периоду, затем складывают 2-й, 3-й и 4-й члены ряда, а их среднюю относят к третьему периоду и т.д.

#### г) **Метод аналитического выравнивания динамических рядов.**

Сущность метода состоит в том, что основная тенденция выражается в виде функции  $y=f(x)$ , где за параметр  $x$  принимается время  $t$ .

### **8.4. Корреляционный анализ и сезонные колебания в рядах динамики**

Для изучения связи в рядах динамики применяется и корреляционный анализ. Однако его применение связано с определенными трудностями, потому что в динамических рядах уровни независимы друг от друга.

Зависимость между каждым предыдущим и последующим членами динамического ряда называется автокорреляцией. Корреляция между уровнями динамических рядов будет правильно отражать связь между явлениями только при условии устранения автокорреляции. Для этого существует ряд способов.

Первый способ состоит в том, что ищется связь не между уровнями рядов, а между первыми, вторыми и т. д. разностями (т. е. из каждого последующего уровня ряда вычитается значение предыдущего - первые разности и т. д.). В этом случае коэффициент корреляции вычисляется по формуле

$$r = \frac{\sum \Delta x \cdot \sum \Delta y}{\sqrt{\sum \Delta x^2 \cdot \sum \Delta y^2}}.$$

Второй способ исключения автокорреляции состоит в том, что сопоставляются отклонения от тренда (основной тенденции) по изучаемым рядам: для каждого динамического ряда проводится аналитическое выравнивание, затем находятся отклонения от найденной основной тенденции и уже потом, используя эти отклонения в качестве искомым переменных, определяют связь.

Для многих явлений общественной жизни характерны внутригодовые повторяющиеся колебания, которые называются сезонными. Они наблюдаются в различных отраслях народного хозяйства: при производстве большинства видов сельхозпродуктов, их переработки, в строительстве, транспорте, торговле и т. д.

Для выявления и измерения интенсивности сезонных колебаний пользуются индексами сезонности, причем индексы сезонности могут вычисляться по разному.

1. Если средний годовой уровень сезонного явления остается от года к году относительно неизменным, применяется метод простых средних. Он со-

стоит в определении простой средней за одни и те же месяцы всего изучаемого периода и в сопоставлении их со средней за весь изучаемый период.

2. Когда уровень явления проявляет тенденцию к росту или снижению, применяют метод помесечных отношений. Он заключается в том, что в начале вычисляются по каждому году процентные отношения между показателями за каждый данный и предшествующий месяцы, а затем из полученных отношений определяется среднеарифметическое.

### Решение типовых задач

**Задача 1.** По следующим данным вычислить среднемесячные остатки материалов за полугодие (табл 8.1):

Таблица 8.1

Дата	1/1	1/2	1/3	1/4	1/5	1/6	1/7
Остатки на начало месяца, у.д.е.	64.8	46.0	28.0	36.0	23.8	21.4	10.2

Решение. В нашей задаче даны остатки материалов на определенные моменты в ремени (1/1, 1/2, и т.д.), промежутки между которыми равны. В этом случае средняя исчисляется по формуле средней хронологического ряда:

$$\bar{y} = \frac{\frac{y_1}{2} + y_2 + \dots + \frac{y_n}{2}}{n-1} = \frac{\frac{464,8}{2} + 446 + 428 + 436 + 423,8 + 421,4 + \frac{410,2}{2}}{7-1} = 432,1$$

у.д.е.

**Задача 2.** По следующим данным определите среднее поголовье коров по совхозу за год (табл 8.2):

Таблица 8.2

Дата	1.01.95	1.07.95	1.11.95	1.01.96
Поголовье	648	720	672	804

Решение. В задаче дано поголовье коров за определенный момент, даты, причем интервалы между датами не равны (6, 4 и 2 месяца). В таких случаях средняя хронологическая моментного ряда рассчитывается как средняя арифметическая взвешенная, где в качестве весов применяются отрезки времени между датами, к которым относятся парные средние смежные значения уровня. Определим прежде всего средние смежные значения уровней.

$$\frac{648 + 720}{2} = 684; \quad \frac{720 + 672}{2} = 696; \quad \frac{672 + 804}{2} = 738 .$$

Взвесим их теперь на отрезке времени между датами

$$\bar{y} = \frac{\sum y \cdot f}{\sum f} = \frac{684 \cdot 6 + 696 \cdot 4 + 738 \cdot 2}{6 + 4 + 2} = 697 .$$

**Задача 3.** Производство продукции по предприятию за 1990 – 1994 гг. характеризуется следующими данными (табл. 8.3):

Таблица 8.3

Годы	1990	1991	1992	1993	1994
Производство продукции, тыс.шт.	4140	4557	5030	5423	5426

Определить:

- 1) начальный, конечный и средний уровни ряда;
- 2) абсолютные приросты по годам, к 1990 г. и среднегодовой абсолютный прирост;
- 3) темп роста и прироста по годам и к 1990 г.;
- 4) абсолютное значение одного процента прироста;
- 5) среднегодовой темп роста и прироста за период 1990 – 1994 гг.

Решение.

1. Начальный уровень (величина первого члена ряда) - 4140, конечный - 5426. Средний уровень ряда определяется по формуле простой средней арифметической, так как ряд периодический

$$\bar{y} = \frac{4140 + 4557 + 5030 + 5423 + 5426}{5} = 4915,2 \text{ тыс. шт.}$$

Вычисленные основные показатели данного динамического ряда сведим в таблицу 8.4.

2. Абсолютный прирост показывает, насколько изменился текущий уровень по сравнению с предыдущим или базисным и определяется как разность двух уровней

$$\Delta_2 = 4557 - 4140 = 417; \quad \Delta_{94/90} = 5426 - 4140 = 1286 .$$



					ные	ные	ные	
	4	-	-	-				-
990	140							
	4	4	417	1				41,4
991	557	17		10,1	10,1	0,1	0,1	
	5	4	890	1				45,6
992	030	73		10,4	21,5	0,4	1,5	
	5	3	128	1				50,3
993	423	93	3	07,8	31,0	,8	1,0	
	5	3	128	1				54,2
994	426		6	00,06	31,1	,06	1,1	

**Задача 4.** Среднегодовой темп роста заработной платы рабочих завода за 1990-91 гг. составил 104%, а за 1992-1994 гг. -106%. Определить среднегодовой темп роста заработной платы на заводе за 1990 – 1994 гг.

Решение. Средняя из средних темпов роста за неодинаковые промежутки времени рассчитывается по формуле средней геометрической взвешенной, причем весами являются сами промежутки времени, следовательно:

$$\bar{K}_i = \sum^{m_i} \sqrt[m_i]{\prod K_i^{m_i}} = \sqrt[5]{104^2 \cdot 106^3} = 105,2\%.$$

### КОНТРОЛЬНЫЕ ВОПРОСЫ И ТЕМЫ ДЛЯ ПОДГОТОВКИ К ЗАЧЕТУ И ЭКЗАМЕНУ

1. Предмет и метод статистики.
2. Что такое закономерность? Динамические и статистические закономерности и их особенности.
3. Что такое статистическая совокупность, единица совокупности? Понятие вариации.
4. В чем сущность и значение закона больших чисел?
5. Какие науки являются теоретической основой статистики?
6. Чем объясняется расчленение статистической науки на отдельные отрасли? Почему изучение статистической науки начинается с общей теории статистики?
7. Сущность и виды учета.
8. На какие этапы подразделяется статистическое исследование?
9. Дайте определение статистического наблюдения. В чем его сущность?
10. Какие две организационные формы статистического наблюдения вы знаете?

11.Что такое объект наблюдения, единица наблюдения и единица совокупности?

12.Какое время в статистике называется объективным и в чем его отличие от субъективного?

13.Что такое критический момент?

14.Классификация видов и способов статистического наблюдения.

15.В чем состоит логический и арифметический контроль материалов статистического наблюдения?

16.Классификация ошибок наблюдения.

17.Случайные и систематические ошибки и их особенности.

18.Дайте понятие сводки.

19.Статистические группировки и их виды.

20.Общее понятие классификации, виды классификаций.

21.Какие признаки называются атрибутивными, а какие - количественными?

22.Что представляют собой ряды распределения и для чего они служат?

23.В чем отличие вариационного ряда от атрибутивного?

24.Что такое частоты, частости, варианты, объем ряда?

25.Как подразделяются вариационные ряды распределения?

26.Что такое полигон распределения, гистограмма распределения?

27.Что такое плотность распределения?

28.Что такое кумулятивная кривая и кривая Лоренца?

29.Из каких основных элементов состоит статистическая таблица?

30.Чем статистическая таблица отличается от обычной таблицы?

31.Какие виды таблиц по характеру подлежащего вы знаете?

32.Классификация таблиц по строению сказуемого.

33.Что такое абсолютные величины, их виды и способы выражения?

34.Какой учет называется натуральным, условно-натуральным, трудовым и стоимостным?

35.Что такое относительные величины?

36.Виды относительных величин.

37.Сущность средней величины.

38.Как вычисляется средняя арифметическая простая?

39.В каких случаях применяется взвешенная средняя арифметическая?

40.Как вычисляется средняя интервального ряда?

41. Как вычисляется средняя из групповых или частных средних?
42. Особенности вычисления средней из относительных величин.
43. Свойства средней арифметической.
44. Средняя гармоническая и способы ее вычисления.
45. Средняя геометрическая и способы ее вычисления.
46. Средняя квадратическая, когда она применяется?
47. Мажорантность средних.
48. Что такое мода, когда и для чего она применяется и как вычисляется?
49. Отличие моды и средней величины.
50. Что такое медиана, когда и для чего она применяется?
51. Что такое вариации признака и каковы показатели вариации?
52. Основные свойства дисперсии.
53. Как определяется дисперсия альтернативного признака?
54. От каких факторов зависит общая вариация признака?
55. Что такое «правило сложения дисперсий», его сущность?
56. Что такое межгрупповая дисперсия и что она характеризует?
57. Что такое средняя из внутригрупповых дисперсий и ее назначение?
58. Что такое момент распределения, виды моментов и их назначение?
59. Что такое кривые распределения и их виды и характеристики?
60. Какими показателями характеризуется форма кривой распределения, и как они вычисляются?
61. Что такое моделирование ряда распределения?
62. Какие этапы моделирования вы знаете?
63. Что такое «степень свободы» - как понятие?
64. Критерии согласия, их виды, для чего они применяются?
65. Выборочное наблюдение: виды и способы, общие понятия.
66. Что такое «ошибка выборки» и почему применяется термин «средняя ошибка выборки»?
67. Что такое предельная ошибка и как она вычисляется?
68. Особенности вычисления предельных ошибок при различных видах и схемах отбора.
69. Почему ошибка типического отбора, как правило, меньше ошибки собственно-случайного отбора?
70. Как определяется необходимая численность выборки?

- 71.Функциональная и корреляционная связь: их особенности.
- 72.Простейшие методы установления тесноты связи.
- 73.В чем сущность балансового метода и дисперсионного анализа?
- 74.Что такое корреляционное отношение?
- 75.Каковы основные задачи корреляционного анализа?
- 76.Каковы задачи регрессионного анализа?
- 77.Каков экономический смысл коэффициентов регрессии?
- 78.Что такое коэффициент корреляции и как он вычисляется?
- 79.Что такое коэффициент детерминации?
- 80.Что такое множественная корреляция?
- 81.Сущность и методы вычисления частных коэффициентов.
- 82.Какие методы (кроме корреляционно-регрессионного анализа) вы еще знаете?
- 83.Что такое динамический ряд?
- 84.Какие приемы применяются для приведения рядов в сопоставимый вид?
- 85.Классификация динамических рядов.
- 86.Расчет средней динамического ряда.
- 87.Основные показатели, применяемые для анализа динамических рядов.
- 88.Что такое коэффициент опережения?
- 89.Важнейшие приемы обработки и анализа динамических рядов.
- 90.Дисперсионный и корреляционный анализ в рядах динамики.
- 91.Сезонные колебания.

### **Библиографический список**

#### **Основной**

1. Яблочник А.Л.. Общая теория статистики. - М., 1976.
2. Общая теория статистики/Под ред. Пасхавера И.С. - М., 1976.
3. Общая теория статистики/Под ред. Спирина А.А., Башиной О.Э.- М., 1994.
4. Елисеева И.И., Юзбашев М.М. Общая теория статистики. - М., 1995.
5. Венецкий И.Г., Венецкий В.И. Основные математико-статистические понятия и формулы в экономическом анализе. - М., 1979.

### **Дополнительный**

1. Статистический словарь. - М., 1989.
2. Ежов А.И. Вычисление рядов распределения. - М., 1973.
3. Липкин М.И. Кривые распределения в экономических исследованиях. - М., 1972.
4. Головач А.В. и др. Критерии математической статистики в экономических исследованиях. - М.: 1973.
5. Пасхавер И.С. Закон больших чисел и статистические закономерности - М., 1974.
6. Дружинин Н.К. Выборочное наблюдение и эксперимент. - М., 1977.

Методические указания к контрольной работе  
по общей теории статистики

Составители

КРАЕВ Владимир Николаевич

КОВАЛЬЧУК Ирина Владимировна

Ответственный за выпуск – зав. кафедрой доцент В.Н. Краев

Редактор А.П. Володина

Корректор И.В. Паньшин

Изд. лиц. №020275 от 13.11.96. Подписано в печать 23.05.98.

Формат 60x84.16. Бумага для множит. техники. Гарнитура Таймс.

Печать офсетная. Усл. печ. л. 3,95. Уч.-изд. л.4.23. Тираж 100 экз. С – 8.

Заказ

Владимирский государственный университет.

Подразделение оперативной полиграфии

Владимирского государственного университета.

Адрес университета и подразделения оперативной полиграфии:

600026, Владимир, ул. Горького, 87.