Министерство транспорта Российской Федерации
Московская Государственная Академия Водного Транспорта

Потапова Е.В.

ОБЩАЯ ТЕОРИЯ СТАТИСТИКИ

Альтаир — МГАВТ
Москва
2011

Потапова Елена Владимировна

ОБЩАЯ ТЕОРИЯ СТАТИСТИКИ

Учебное пособие

Компьютерная верстка: Шершакова И.В.

Подписано в печать__.__.2011
Формат 60х90/16. Объем 9,62 п.л.
Заказ № __________ Тираж 100 экз.

Альтаир-МГАВТ
Московская государственная академия водного транспорта
117105 г. Москва, Новоданиловская набережная, д.2, корп.1

Министерство транспорта Российской Федерации
Московская Государственная Академия Водного Транспорта

Потапова Е.В.

ОБЩАЯ ТЕОРИЯ СТАТИСТИКИ

Учебное пособие

Альтаир – МГАВТ
Москва
2011

Потапова Е.В.

Общая теория статистики. Учебное пособие. М, Альтаир — МГАВТ, 2011, 155 с.

Пособие подготовлено в соответствии с требованиями государственного образовательного стандарта для экономических специальностей.

В пособии излагаются общие вопросы теории статистики. Рассматриваются метод группировок, расчет относительных и средних величин, показатели вариации, корреляционный и регрессионный анализ, анализ временных рядов, выборочное наблюдение и экономические индексы. В каждой главе приведены примеры с подробными решениями, широко используются графики и таблицы, приводятся вопросы для самоконтроля знаний.

Для студентов специальностей "Экономика и управление на предприятии (по отрасли водный транспорт)", "Управление персоналом", "Менеджмент организации".

Рецензент: доц. М.М. Галкина

Учебное пособие рассмотрено и одобрено на заседании кафедры "Экономика водного транспорта" " 31 " августа 2010г. (протокол № 1)

Рассмотрено и рекомендовано к изданию Учебно-методическим советом МГАВТ.

ОГЛАВЛЕНИЕ

Предисловие 5

1. Статистическое наблюдение 6
1.1. Основные понятия 6
1.2. Формы, виды и способы статистического наблюдения 9
1.3. Проверка достоверности собранных данных 12
Вопросы для самоконтроля 13

2. Сводка и группировка материалов статистического наблюдения . 14
Вопросы для самоконтроля 20
3. Статистическая таблица и ее элементы 21
Вопросы для самоконтроля 23
4. Графическое представление статистической информации 24
Вопросы для самоконтроля 33
5. Абсолютные и относительные величины 34
Вопросы для самоконтроля 39

6. Средние величины 40
6.1. Основные понятия 40
6.2. Степенные средние 40
6.2.1. Средняя арифметическая 41
6.2.2. Средняя гармоническая 43
6.2.3. Средняя геометрическая 46
6.2.4. Средняя квадратическая 46
6.3. Структурные средние 47
6.3.1. Мода 47
6.3.2. Медиана 49
Вопросы для самоконтроля 52

7. Ряды распределения и их характеристики 54
7.1. Основные понятия 54
7.2. Графическое изображение рядов распределения 57
7.3. Показатели центра распределения 62
7.4. Показатели вариации (колеблемости) признака и правило сложения дисперсий 62
7.4.1. Показатели вариации признака 62
7.4.2. Сложение дисперсий изучаемого признака 67
7.5. Показатели формы распределения 71
7.5.1. Ранговые характеристики 72
7.5.2. Показатели дифференциации 75
7.5.3. Кривые распределения, показатели асимметрии и эксцесса . . . 76
Вопросы для самоконтроля 80

8. Выборочное наблюдение 81
8.1. Виды схемы отбора 81
8.2. Средняя и предельная ошибки выборки 83
Вопросы для самоконтроля 88
9. Корреляционная связь и ее статистическое изучение 89
9.1. Понятие корреляции 89
9.2. Парная корреляция 91
9.2.1. Определение степени тесноты связи в случае линейной зависимости между признаками 91
9.2.2. Определение степени тесноты связи в случае нелинейной зависимости между признаками 93
9.2.3. Оценка существенности линейного коэффициента корреляции 95
9.2.4. Оценка существенности корреляционного отношения 97
9.2.5. Определение параметров линейной зависимости 98
9.2.6. Определение параметров нелинейной зависимости 101
9.2.7. Экономическая интерпретация результатов построения однофакторной экономической модели 105
9.3. Непараметрические методы оценки связи 107
Вопросы для самоконтроля 112
10. Ряды динамики 113
10.1. Основные понятия 113
10.2. Показатели рядов динамики 115
10.3. Средние показателя рядов динамики 119
10.4. Выявление и характеристика основной тенденции развития 122
10.5. Анализ сезонных колебаний 127
Вопросы для самоконтроля 131
11. Индексы и их использование в экономических исследованиях . . 132
11.1. Основные понятия 132
11.2. Индивидуальные индексы 133
11.3. Общие индексы 135
11.3.1. Общие индексы количественных показателей 135
11.3.2. Общие индексы качественных показателей 138
11.4. Построение системы индексов 140
11.5. Изучение динамики средних величин 141
11.6. Использование индексного метода в анализе взаимосвязей экономических явлений 145
Вопросы для самоконтроля 149
Приложение 151
Рекомендуемая литература 152

ПРЕДИСЛОВИЕ

Цель данного пособия — помочь студентам, изучающим дисциплину «Статистика», овладеть приемами и навыками статистической обработки и анализа экономической информации, научиться правильно интерпретировать полученные результаты и формулировать выводы.

Пособие включает в себя основные разделы курса «Общая теория статистики». В нем рассматриваются методы сбора статистической информации; особенности методологии проведения, сводки и группировки первичных данных; способы построения статистических таблиц и графиков; методы расчета относительных и средних величин; методы анализа вариации и частотных распределений; методы и показатели оценки взаимосвязи признаков; методология изучения динамики экономических явлений и процессов; виды и способы расчета экономических индексов.

Настоящее издание учебного пособия, в отличие от предыдущего (1-е издание — 2002 г.), переработано и дополнено новыми материалами. Заново написаны главы: «Графическое представление статистической информации», «Средние величины», «Ряды распределения», «Корреляционная связь и ее статистическое изучение», «Индексы и их использование в экономических исследованиях».

Пособие составлено в соответствии с учебной программой курса и предназначено для студентов экономических специальностей: «Экономика и управление на предприятии (по отрасли водный транспорт)», «Управление персоналом», «Менеджмент организации».

1. Статистическое наблюдение

1.1. *Основные понятия*

Любое статистическое исследование последовательно проходит следующие этапы:

I статистическое наблюдение;

II группировка и сводка материалов, собранных посредством статистического наблюдения;

III вычисление и анализ обобщающих статистических показателей относительных, средних величин, экономических индексов и т.д.

Статистическое наблюдение — это научно организованная работа по сбору массовых первичных данных о количественной стороне общественной жизни.

Статистическое наблюдение состоит из трех этапов:

– подготовительные работы,
– сбор необходимой информации,
– проверка достоверности собранных данных.

Чтобы успешно провести статистическое наблюдение, разрабатывается специальный план наблюдения, включающий в себя программно-методологические и организационные вопросы.

К **программно-методологическим вопросам** относятся: определение цели, объекта и единицы наблюдения, разработка программы наблюдения.

Цель статистического наблюдения формируется исходя из задач статистического исследования.

Объектом наблюдения называют совокупность единиц рассматриваемого явления, о которых должна быть собрана статистическая информация. Объектом наблюдения может быть, например, совокупность жителей страны, предприятий, коммерческих банков, высших учебных заведений и т.п.

Единицей наблюдения называют первичный элемент объекта наблюдения, который является носителем признаков, подлежащих учету. Так, например, при переписи населения единицей наблюдения является каждый отдельный человек. Однако, если ставится также задача определить численность и состав домохозяйств, то единицей наблюдения наряду с человеком будет являться каждое домохозяйство.

Единица наблюдения обладает множеством различных признаков. Именно значения различных признаков и регистрируются на стадии статистического наблюдения.

Признак — это объективная характеристика единицы совокупности, характерная черта или свойство, которое может быть определено или т измерено. Например, признаками человека являются пол, возраст, место жительства, профессия, среднемесячный доход и др. Признаками, характеризующими промышленное предприятие, являются выручка от реализации продукции, прибыль, стоимость основных фондов, численность персонала и др.

Значения признаков у разных единиц совокупности меняются, т.е. варьируют от единицы к единице. Статистика изучает явления, оценивая **варьирующие** признаки единиц совокупности.

Значение признака может задаваться на дату (численность населения, остатки продукции материалов на складах, остатки вкладов на банковских счетах) или за период, например, за месяц, квартал, полугодие, год и т.д. (количество родившихся или умерших за год, численность эмигрировавшего или иммигрировавшего населения за квартал, уровень заработной платы работников за месяц).

Значение признака у отдельных единиц совокупности называется **вариантой** и обозначается x_i.

Программа наблюдения — это перечень вопросов, на которые в процессе наблюдения должны быть получены ответы, либо это перечень признаков и показателей, подлежащих регистрации.

Основные требования к программе наблюдения:

1. Программа должна содержать только те вопросы, которые действительно необходимы для данного статистического исследования.
2. В программу следует включать лишь те вопросы, на которые можно получить точные ответы.
3. Нельзя включать в программу вопросы, которые могут вызвать подозрение, что ответы на них могут быть использованы во вред опрашиваемым.
4. Программу целесообразно строить так, чтобы ответами на одни вопросы можно было контролировать ответы на другие.

Вопросы программы статистического наблюдения и ответы на них находят отражение в **статистическом формуляре** (переписной лист, анкета, форма и т.д.). К статистическим формулярам составляется **инструкция**, где подробно разъясняется, как следует заполнять формуляр.

При подготовке и проведении статистического наблюдения необходимо также решить ряд **организационных вопросов**. К ним относятся: определение места и способа наблюдения, сроков проведения наблюдения, критический момент наблюдения, подбор и расстановка кадров и др.

Срок проведения наблюдения — это период, в течение которого должны быть получены сведения об изучаемом явлении.

Критический момент наблюдения — это момент, по состоянию на который происходит учет.

Так Всероссийская перепись населения 2010 г. проводилась с 14 по 25 октября 2010 г., а критическим моментом наблюдения был ноль часов 14 октября.

1.2. *Формы, виды и способы статистического наблюдения*

Различают формы, виды и способы статистического наблюдения (рис. 1.1.).

Рис. 1.1. Формы, виды и способы статистического наблюдения

Формы статистического наблюдения

Статистические данные можно получить различными путями. С организационной точки зрения различают три формы статистического наблюдения: отчетность, специально организованное статистические наблюдение и регистры.

Отчетность — это такая организационная форма наблюдения, при которой единицы наблюдения представляют сведения о своей деятельности по образцам установленной формы и в установленные сроки. Отчетные документы обязательно заверяются подписью лиц, ответственных за ее предоставление.

Специально организованное статистическое наблюдение представляет собой сбор данных посредством переписей, специально организованных обследований и учетов. Примером этой формы наблюдения могут служить переписи населения; выборочное обследование населения по проблемам занятости; регистрация цен для расчета индекса потребительских цен; всероссийская сельскохозяйственная перепись и др.

Регистры (регистровое наблюдение) — это форма непрерывного статистического наблюдения за процессами, имеющими фиксированное начало, стадию развития и фиксированный конец. При этой форме наблюдения данные об изучаемых объектах заносятся в статистический регистр. Каждая единица наблюдения характеризуется совокупностью показателей. Одни показатели остаются неизменными в течение всего времени наблюдения; другие, периодичность изменения которых неизвестна, обновляются по мере изменения; третьи — представляют собой динамические ряды показателей с заранее известным периодом обновления. Все эти показатели хранятся до полного завершения наблюдения за единицей исследуемой совокупности. Примерами таких форм наблюдения могут служить регистры населения (поименованный и регулярно обновляемый перечень жителей страны), регистр предприятий и организаций и др.

Виды статистического наблюдения

Виды статистического наблюдения классифицируются по времени регистрации фактов во времени и по степени охвата единиц совокупности.

По времени регистрации различают непрерывное, периодическое и единовременное наблюдение.

Непрерывное наблюдение ведется постоянно и регистрация фактов проводится систематически по мере их возникновения. Примером непрерывного наблюдения может служить, например, регистрация рождения, смерти и гражданского состояния.

Периодическое наблюдение это наблюдение, при котором регистрация фактов проводится через определенные интервалы времени. Примером периодического наблюдения могут служить переписи населения, текущая отчетность предприятий, регистрация цен на момент закрытия товарно-сырьевых и валютных бирж и т.д.

Единовременное наблюдение проводится однократно по мере необходимости или время от времени без соблюдения строгой периодичности.

По степени охвата единиц совокупности различают сплошное и несплошное наблюдение.

Сплошным называется наблюдение, при котором регистрируются все без исключения единицы изучаемой совокупности.

Несплошным называется наблюдение, когда учету подлежит только часть единиц исследуемого явления.

Различают три метода несплошного наблюдения:

– выборочный метод;
– метод основного массива;
– монографический метод.

Выборочным называется наблюдение, основанное на принципе случайного отбора единиц изучаемой совокупности, которые подлежат наблюдению.

Метод основного массива характеризуется тем, что отбирают наиболее крупные единицы наблюдения, в которых сосредоточена значительная доля всех подлежащих изучению фактов.

Так, в ряде европейских стран до 80% розничного товарооборота приходится на 10–15 крупнейших компаний — сетей супермаркета. Поэтому для определения основных тенденций в сфере розничных продаж достаточно осуществлять наблюдение за работой этих компаний.

Монографический метод используется для подробного описания единичных, но типичных объектов. К монографическому наблюдению относится, например, аудит сложившегося бизнеса.

Способы статистического наблюдения

Статистические материалы могут быть получены различными способами:

– **непосредственным наблюдением** осуществляется представителем статистических органом на основе личного осмотра, подсчета или измерения изучаемых признаков;

– **документальным способом наблюдения** — путем использования различного рода документов;
– **способом опроса** — путем регистрации ответов, даваемых опрашиваемыми лицами.

1.3. *Проверка достоверности собранных данных*

В практике организации статистического наблюдения важное место занимает контроль за точностью собранных данных. Ошибки наблюдения возникают по различным причинам и делятся на ошибки регистрации и ошибки репрезентативности.

Ошибки регистрации встречаются как при сплошном, так и при несплошном наблюдении и отражают расхождение между фактическим значением показателя и зарегистрированным в процессе статистического наблюдения. Они могут быть случайными или систематическими.

Случайные ошибки происходят по чисто случайным причинам. Они могут быть направлены в равной мере и в сторону увеличения и в сторону уменьшения значения показателя, и потому при большом массиве наблюдения не вызывают искажения итоговых результатов.

Систематические ошибки направлены в одну сторону: либо в сторону увеличения, либо в сторону уменьшении я значения показателя. Они могут быть преднамеренными (тенденциозными) и непреднамеренными.

В целях выявления и устранения ошибок статистического наблюдения применяют внешний, логический и арифметический контроль. При **внешнем контроле** выясняется, на все ли вопросы в формулярном бланке даны ответы.

Логический контроль позволяет путем логического сопоставления ответов на отдельные вопросы программы наблюдения выяснить допущенные ошибки.

Арифметический контроль основан на проверке взаимосвязанных показателей, отраженных в формуляре статистического наблюдения.

Ошибки репрезентативности могут возникнуть при несплошном наблюдении, когда выборочная совокупность недостаточно полно отражает состав генеральной совокупности и показатели, исчисленные по выборочной совокупности, не будут совпадать с показателями, вычисленными для всей совокупности.

Вопросы для самоконтроля

1. Какие этапы проходит любое статистическое исследование?
2. Дайте определение объекта и единицы статистического наблюдения. Приведите пример.
3. Что такое статистический признак? Приведите пример.
4. Что понимается под программой статистического наблюдения?
5. Какие требования предъявляются к программе наблюдения?
6. Что такое статистический формуляр?
7. Назовите разницу между понятиями «срок наблюдения» и «критический момент наблюдения».
8. Что включают в себя организационные вопросы статистического наблюдения?
9. Назовите формы, виды и способы статистического наблюдения.
10. Что такое ошибки регистрации? Какие бывают ошибки регистрации?
11. Что такое ошибки репрезентативности?
12. В чем заключается логический контроль статистической информации?
13. В чем заключается арифметический контроль статистической информации?

2. Сводка и группировка материалов статистического наблюдения

Чтобы полученные в результате статистического наблюдения сведения о каждой единице наблюдения могли быть использованы для характеристики изучаемой совокупности в целом, они должны быть научно обработаны, систематизированы, подсчитаны, обобщены. Это достигается путем группировки и сводки исходной информации.

Группировка — это расчленение единиц статистической совокупности на группы, однородные в каком-либо существенном отношении.

Следующим за группировкой этапом систематизации и обобщения данных является статистическая сводка.

Статистическая сводка (в узком смысле слова) это подсчет числа единиц в подгруппах и группах, выделенных при группировке и подведение итогов по количественным признакам.

Результаты группировки и сводки оформляются в виде статистических таблиц.

Признак, на основе которого производится группировка, называется **группировочным признаком** или **основанием группировки**.

Группировочные признаки могут быть атрибутивными (качественными) и количественными.

Атрибутивные признаки — это признаки, которые выражаются в виде понятий или наименований. Например, пол, профессия, образование, родной язык и др.

Количественные признаки — это признаки, значения которых имеют цифровое выражение. Например, возраст, рост человека, стаж работы, доходы и др.

Если изучаемый признак может принимать только два значения, то он называется **альтернативным.** Альтернативными чаще всего бывают качественные признаки. Например, признак «пол» имеет только два значения: мужской или женский. Альтернативный признак «качество» также может принимать два значения: либо продукция бракованная, либо нет.

В некоторых случаях к альтернативному виду можно привести и количественные признаки. Например, можно задать некоторое значение количественного признака и анализировать подмножества единиц совокупности с бóльшим или мéньшим значением. Например, выбрать из совокупности рабочих предприятия тех работников, у которых стаж свыше 5 лет. В этом случае совокупность рабочих будет характеризоваться альтернативным признаком: стаж либо свыше 5 лет, либо до 5 лет.

При группировке по атрибутивному признаку число групп определяется количеством соответствующих наименований значений признака.

При группировке по количественному признаку число групп определяется в зависимости от характера изменения признака.

Если количественный признак принимает только некоторые, чаще целые значения, т.е. является **дискретным** (например — количество детей в семье, комнат в квартире, тарифный разряд), то число групп соответствует количеству значений признака.

Непрерывные признаки — это признаки, которые в определенных пределах могут принимать любые значения (целые и дробные), например, стаж работы или возраст сотрудников.

При группировке данных с непрерывными признаками или дискретными признаками, имеющими большое число возможных значений, весь диапазон значений признака разбивается на интервалы (группы).

Величина (ширина) **интервала** — это разница между максимальным и минимальным значениями признака в группе.

При этом наименьшее значение признака в интервале называется **нижней границей интервала**, а наибольшее значение признака в интервале называется **верхней границей интервала**.

В статистической практике используют три вида интервалов: равные, неравные и специализированные.

Равные интервалы используются обычно в тех случаях, когда варьирующий (т.е. изменяющийся) признак в группе распределяется примерно равномерно.

Величину равных интервалов определяют путем деления разности максимального и минимального значений признака в совокупности на число образуемых групп (интервалов):

$$h = \frac{x_{\max} - x_{\min}}{n},$$

где h — величина (ширина) интервала;

$x_{\max}$, $x_{\min}$ — соответственно, максимальное и минимальное значение признака в изучаемой совокупности;

n — число образуемых групп.

Величину интервала h можно не округлять, но если округляют, то всегда до бóльшего целого числа, чтобы наибольшее значение признака совокупности попало в последний интервал.

Если число групп (n) заранее не задано или исследователь затрудняется с определением их количества, можно использовать формулу Стерджесса для определения числа групп для данной совокупности:

$$n = 1 + 3{,}22\lg N,$$

где n — число групп;

N — число единиц в изучаемой совокупности.

Пример. Сформировать 3 группы с одинаковыми интервалами для следующих значений:

2,1; 8,1; 10,5; 3,5; 2,2; 2,2; 6,4; 12,0; 15,1; 18,3; 10,1; 5,0

Сначала определим ширину интервала h, число групп нам задано — три;

$$h = \frac{18{,}3 - 2{,}1}{3} = 5{,}4 \approx 6.$$

Ширину интервала (5,4) округляем в бóльшую сторону, чтобы максимальное значение признака (18,3) попало в последний интервал.

Формируем интервалы и заносим результаты в таблицу 2.1.

Таблица 2.1

Интервал	Количество значений попадающих в интервал
2,1 — 8,1 (–)*)	6
8,1 — 14,1	4
14,1 — 20,1	2
Всего	**12**

*) При количественных группировках следует обращать внимание на правильное обозначение нижней и верхней границ каждой группы. Так, если не дать при данной группировке специальных указаний, то возникнут трудности при ее составлении и чтении — куда, например, отнести значение 8,1 — в первый или во второй интервал. Для устранения подобной неопределенности рядом с первым интервалом указывают символы «–» (т.е. исключая верхнюю границу) или «+» (т.е. включая верхнюю границу).

Неравные интервалы используются, как правило, в аналитических группировках (см. ниже). В этом случае ширина интервала подбирается таким образом, чтобы число единиц в образованных группах было примерно одинаково.

Специализированные интервалы используются в типологических группировках. Границы групп устанавливаются там, где намечается переход от одного качества к другому. Наметить точки перехода возможно только на основе теоретического анализа.

Статистические группировки преследуют цели выделения качественно однородных совокупностей, изучения структуры совокупности, исследования существующих зависимостей. Каждой из этих целей соответствует особый вид группировки: типологическая, структурная и аналитическая.

В **типологических группировках** первичная статистическая информация объединяется в однородные по качеству и условиям развития социально-экономические типы. Например, группировки хозяйствующих субъектов по формам собственности; разделение экономически активного населения на занятых и безработных; работников — на занятых преимущественно физическим и умственным трудом и др. Решение вопроса об основании группировки осуществляется на основе анализа сущности изучаемого социально-экономического явления (табл. 2.2).

Таблица 2.2

Динамика численности экономически активного и экономически неактивного населения города N в 2007—2008 гг., чел.

	2007 г.	2008 г.
Экономически активное население, — всего	68 079	70 355
в том числе		
– занятое население	60 021	61 261
– безработное население	8 058	9 094
Экономически неактивное население	41 264	39 862
Всего	**109 343**	**110 217**

Структурные группировки характеризуют состав однородной совокупности по какому-либо признаку. Их задача — показать долю (удельный вес) тех или иных значений признака в совокупности. Например, группировка населения по полу; работников предприятия — по стажу работы, квалификации; перевезенных грузов — по видам грузов и т.д. (табл. 2.3).

Таблица 2.3

Перевозки грузов судоходной компанией в 2008 г., тыс.т

	тыс.т	%
Перевезено грузов, — всего	118	100
в том числе:		
– сухогрузов	85	72
– наливных грузов	23	20
– лесных грузов в плотах	10	8

Аналитические группировки используют для выявления взаимосвязи между группировочным признаком и другим (или другими) признаками (табл. 2.4).

Таблица 2.4

Группировка рабочих автотранспортных предприятий по вооруженности труда основным капиталом

Вооруженность труда основным капиталом, тыс.руб./чел.	Годовая производительность труда работников, тыс.ткм	Текучесть кадров, %
123,0—125,0	25,20	7,9
125,0—127,0	25,96	6,3
127,0—129,0	31,47	8,9

В зависимости от числа положенных в основание группировки признаков различают простые и комбинированные (многомерные) группировки.

Простая группировка — это группировка выполненная по одному признаку (табл. 2.2, 2.3, 2.4).

Среди простых группировок особо выделяют ряды распределения.

Ряд распределения — это группировка, в которой для характеристики групп, **упорядоченных по возрастанию или убыванию**, применяется один показатель — численность группы. Другими словами, это ряд чисел, показывающий, как распределяются единицы совокупности по изучаемому признаку (табл. 2.5)

Таблица 2.5

Распределение оценок за тест в студенческой группе

(максимальное число — 10 баллов)

Баллы	Количество студентов, чел.
0	1
2	1
4	6
5	4
6	3
9	1
Всего	**16**

Сложная группировка — это группировка единиц совокупности по нескольким признакам (табл. 2.6).

Таблица 2.6

Распределение семей района N по количеству детей

Группы населения	Количество детей в семье	Количество семей, тыс.
1. Городское население	1 ребенок	42 312
	2 ребенка	14 104
	3 ребенка	8 050
	4 и более детей	6 054
Итого по группе 1	—	70 520
2. Сельское население	1 ребенок	1 538
	2 ребенка	6 152
	3 ребенка	4 614
	4 и более детей	3 076
Итого по группе 2	—	15 380
Всего	**—**	**85 900**

От группировок следует отличать классификацию.

Классификацией называется систематизированное распределение явлений и объектов на определенные группы, классы, разряды на основании их сходства и различия.

Отличительными чертами классификаций являются:

– в их основе лежит качественный (атрибутивный) признак;
– классификации стандартны и устанавливаются органами государственной и международной статистики;
– классификации более устойчивы, чем группировки, т.к. остаются неизменными в течение длительного периода времени.

Вопросы для самоконтроля

1. Что такое группировка?
2. Дайте определение группировочного признака.
3. Какие существуют разновидности группировочных признаков?
4. Что такое альтернативные признаки?
5. Как определяется величина интервала при группировке по количественному признаку?
6. Какие существуют виды группировок?
7. Что такое структурная группировка?
8. Что такое аналитическая группировка?
9. В чем заключается смысл простой и сложной группировки?
10. Что понимается под классификацией в статистике? Чем она отличается от группировки?

3. Статистическая таблица и ее элементы

Данные сводки и группировки материалов обычно излагаются в виде статистических таблиц.

Статистическая таблица — это форма рационального, наглядного изложения статистических данных о явлениях и процессах, изучаемых статистикой.

Различают скелет (остов) и макет таблицы. Если из статистической таблицы изъяты все слова и цифры, то получится **скелет таблицы**. Если записать заголовки граф и строк, то получится **макет таблицы**. (Рис. 3.1)

Сказуемое / Подлежащее	Заголовки граф				
1	2	3	4	5	6
Перечено (группы) единиц совокупности					

Рис. 3.1. Макет статистической таблицы

Статистическая таблица подобно предложению в грамматике имеет подлежащее и сказуемое.

Подлежащим называется перечень единиц или групп, на которые разбита статистическая совокупность (т.е. объект изучения). Подлежащее, как правило, размещается в строках левой части таблицы.

Сказуемое таблицы — это числовые показатели, с помощью которых характеризуется подлежащее.

В зависимости от построения подлежащего статистические таблицы принято делить на простые, групповые и комбинированные.

Простые — это такие статистические таблицы, в подлежащем которых нет группировок. Простые таблицы, как правило, используется в качестве справочного материала. Простые таблицы бывают:

перечневые — подлежащее — перечень единиц, составляющих объект изучения;

территориальные — дается перечень территорий, стран, городов и пр.;

хронологические — в подлежащем приводятся периоды или даты.

Для анализа исходной информации используют групповые и комбинационные таблицы.

Таблицы, которые содержат в подлежащем группировку по одному признаку, называются **групповыми**. (Рис. 3.2, табл. 2.2, 2.3, 2.4, 2.5).

Группы населения по возрасту	Численность населения, тыс.чел.	Численность населения, в % к итогу
0—4 5—9 10—14 … 70 и более		
Всего:		

Рис. 3.2. Макет групповой таблицы

Если группировка подлежащего проводится по двум и более признакам, таблица называется **комбинационной**. В этом случае все единицы распределяются на группы сначала по одному признаку, а затем внутри каждой из выделенных групп — на подгруппы по другому признаку. (Рис. 3.3; табл. 2.6)

Группы предприятий по величине прибыли, млн.руб.	Группы предприятий по численности промышленно-производственного персонала, чел.	Число предприятий
50—100	200—250	
	250—300	
	300—350	
Итого по группе:	—	
100—150	200—250	
	250—300	
	300—350	
Итого по группе:		
Итого по подгруппам	200—250	
	250—300	
	300—350	
Всего:		

Рис. 3.3. Макет комбинационной таблицы

Основные правила оформления таблиц

1. Таблица по возможности должна быть краткой. Не следует загружать ее излишними подробностями, затрудняющими анализ исследуемых явлений.

2. Название таблицы, заглавия строк подлежащего и граф сказуемого должны быть сформулированы точно, кратко и ясно и, если это требуется, должны иметь единицы измерения.
3. Слова в таблице пишутся полностью, без сокращений (за исключением общепринятых).
4. При оформлении таблиц обычно применяются такие условные обозначения:

знак (—) — когда явление отсутствует;

X — если явление не имеет осмысленного содержания;

0,0 — если численное значение признака меньше принятой в таблице точности;

(…) или (н.св.) — когда отсутствуют сведения о его размере.

Округленные числа приводятся в таблице с одинаковой степенью точности.

Вопросы для самоконтроля

1. Дайте определение статистической таблицы.
2. Что такое подлежащее статистической таблицы?
3. Что такое сказуемое статистической таблицы?
4. Назовите виды таблиц по характеру подлежащего.
5. Какие таблицы называются перечневыми? групповыми? комбинационными?
6. Назовите основные правила оформления таблиц.

4. Графическое представление статистической информации

Полученный в результате обработки статистический материал, расположенный в таблицах, часто нуждается в наглядном изображении с помощью построения статистических графиков.

Графиками в статистике называют наглядные изображения статистических величин в виде различных линий, геометрических фигур или географических картосхем.

В отличие от табличной формы изложения материала, графическое представление данных является менее точным (по графику сложно судить о величине явления с большой точностью), зато графики дают наглядное представление о соотношениях величин, частей или о динамике явления.

Каждый график состоит из графического образа и вспомогательных элементов.

Графический образ — это совокупность точек, линий и фигур, с помощью которых изображаются статистические данные.

Вспомогательными элементами графика являются:

1) ***поле графика*** — то пространство, в котором размещаются образующие график геометрические знаки;
2) ***пространственные ориентиры***, определяющие расположение геометрических знаков в поле графика.

   Пространственные ориентиры задаются системой координатных сеток или контурных линий, которые делят это поле на части;
3) ***масштабные ориентиры***. Они определяются системой масштабных шкал или специальными масштабными знаками;
4) ***экспликация графика***, состоит из:
   а) названия графика;
   б) смыслового значения каждого знака, применяемого на данном графике.

В зависимости от способов построения, целей использования и форм графического образа, графики можно классифицировать следующим образом.

I. **По способу построения** различают диаграммы и статистические карты.

Диаграммы представляют собой чертеж, на котором статистические показатели изображаются в виде линий или геометрических фигур.

Статистические карты используются для характеристики распределения явления на определенной территории.

II. **По целя использования:**

а) **диаграммы** подразделяются на:

- диаграммы структуры (секторные, столбиковые и полосовые диаграммы структуры);
- диаграммы сравнения (столбиковые и полосовые диаграммы сравнения, знаки Варзара, направленные диаграммы, пиктограммы, круговые, квадратные);
- диаграммы динамики и выполнения плана (столбиковые, полосовые, линейные, квадратные, круговые, радиальные);
- диаграммы оценки вариации (полигон, гистограмма, кумулята, огива);

б) **статистические карты** подразделяются на:

- картограммы (точечные, фоновые);
- картодиаграммы.

Диаграммы структуры предназначены для наглядного изображения структуры изучаемой статистической совокупности. Их назначение — показать, как соотносятся между собой отдельные части целого. При построении диаграмм **структуры** абсолютные величины пересчитываются в проценты.

Диаграммы сравнения используют для того, чтобы показать, как соотносятся между собой различные сопоставимые объекты. Чтобы сравнение было корректным, ***период времени*** за который (или по состоянию на который) берутся данные ***должен быть один и тот же***.

Диаграммы динамики и выполнения плана характеризуют развитие явления во времени (или по сравнению с плановыми показателями).

Диаграммы оценки вариации служат для графического изображения рядов распределения и их характеристик. Построение этих диаграмм более подробно представлено в главе 7 (п. 7.2).

Картограмма представляет собой географическую карту, на которой штриховкой, окраской различной яркости или точками изображается интенсивность размещения социально-экономического явления в пределах каждой единицы территориального деления.

Картодиаграмма — это сочетание географической карты с диаграммой. С помощью картодиаграмм можно отразить более сложные статистико-географические сопоставления по сравнению с картодиаграммами.

Способы построения диаграмм:

1. **Секторные диаграммы** используются, главным образом, для изображения структуры совокупности. Они представляют собой графические изображения статистических данных при помощи секторов круга.

Для построения диаграммы исходные данные сначала переводятся в проценты: вся совокупность принимается за 100% и определяется удельный вес каждой части совокупности. «Работающим» геометрическим параметром в секторной диаграмме служит величина угла между радиусами: 1% принимается на диаграмме равным 3,6°, а сумма всех углов, составляющая 360°, приравнивается к 100%.

На примере таблицы 4.1 построена секторная диаграмма (рис. 4.1).

Таблица 4.1

Перевозки грузов судоходной компанией в 2005 и 2008 гг.

	Объем перевозок, тыс.т		Структура перевозок, %	
Всего перевезено	231,0	240,0	100,0	100,0
в том числе – наливные грузы	69,3	36,0	30,0	15,0
– сухогрузы	150,1	204,0	65,0	85,0
– лесные грузы	11,6	—	5,0	—

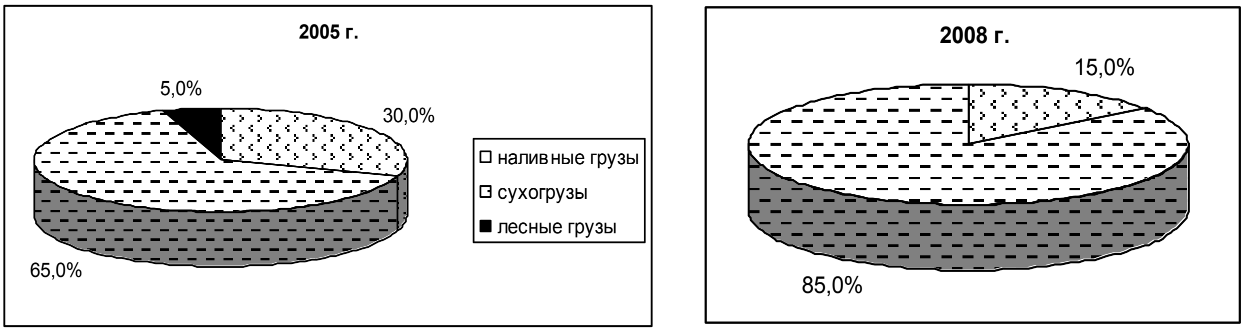


Рис. 4.1. Динамика структуры перевозок грузов судоходной компанией в 2005 и 2008 гг.

2. **Столбиковые диаграммы структуры**. При построении этих диаграмм исходные данные также предварительно выражаются в процентах. Длина прямоугольника, выражающего всю совокупность, принимается равной 100% (рис. 2.4). Ширина столбиков одинаковая.

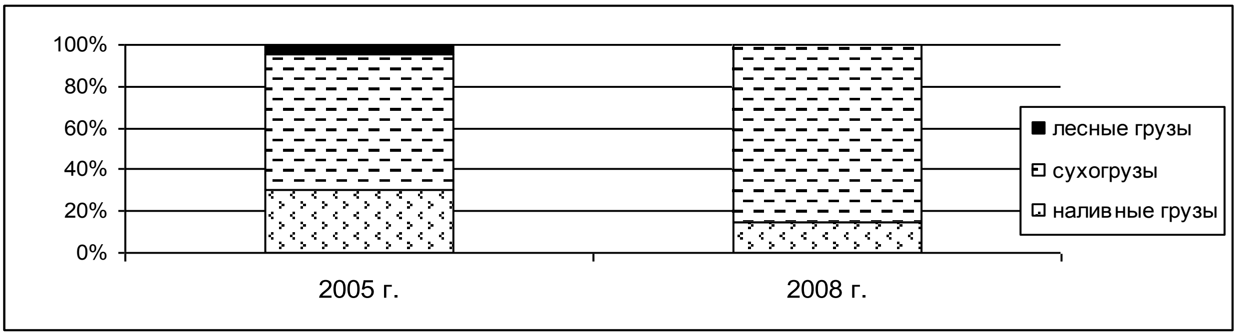


Рис. 4.2. Структура перевозок грузов судоходной компанией в 2005 и 2008 гг.

3. **В полосовых диаграммах структуры** масштабная шкала расположена по горизонтали (рис. 4.3).

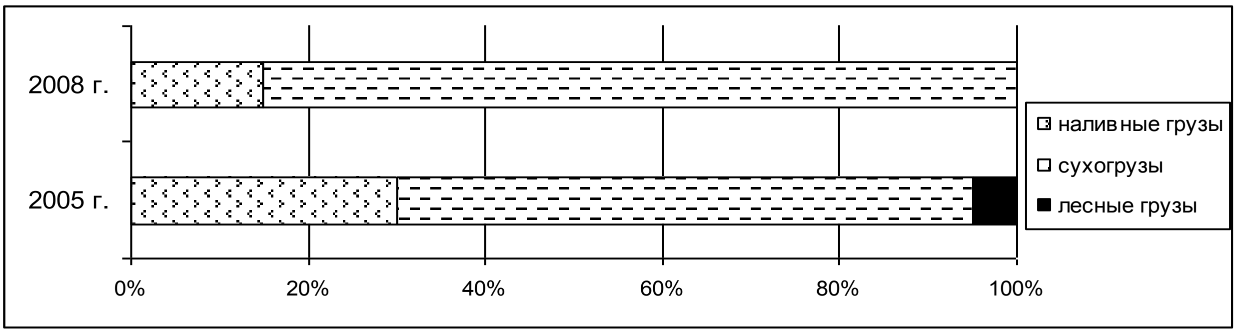


Рис. 4.3. Структура перевозок грузов судоходной компанией в 2005 и 2008 гг.

4. При построении **столбиковых диаграмм сравнения** (или **динамики**) каждый столбик изображает величину отдельного уровня исследуемого статистического ряда. Сравнение уровней показателей возможно потому, что они

выражены в одной единице измерения. Величина каждого столбика по вертикали соответствует размеру изображаемого на графике показателя. Основания столбиков одинаковые.

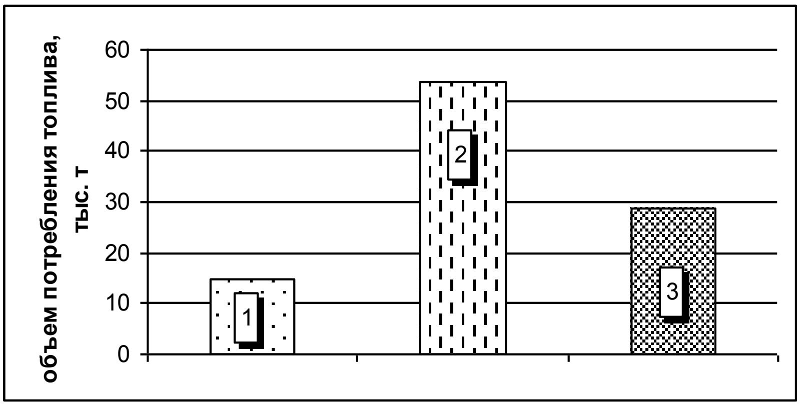
На примере табл. 4.2 построена:

– столбиковая диаграмма **сравнения** потребления топлива организациями транспорта в 2008 г. (рис. 4.4);
– столбиковая диаграмма **динамики** потребления автомобильного бензина в 2004—2008 гг. (рис. 4.5).

Таблица 4.2

Потребление топлива организациями транспорта в регионе N, тыс.т

	2004	2005	2006	2007	2008
Автомобильный бензин	19,1	18,0	17,1	15,2	14,8
Дизельное топливо	47,1	49,5	49,3	53,9	53,8
Уголь и продукты переработки угля	45,9	44,0	39,7	34,3	28,6



Условные обозначения
1 — автомобильный бензин
2 — дизельное топливо
3 — уголь и продукты переработки угля

Рис. 4.4. Потребление топлива организациями транспорта в регионе N в 2008 г.

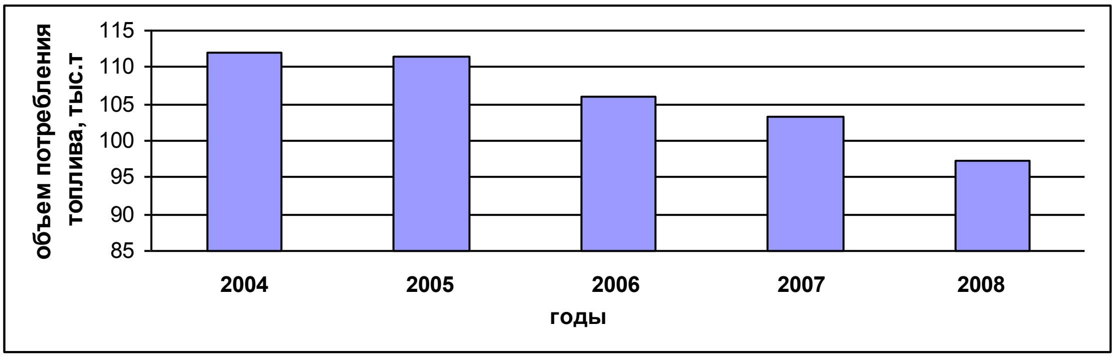


Рис. 4.5. Динамика потребления автомобильного бензина предприятиями транспорта региона N в 2004—2008 гг.

5. **Полосовые (ленточные) диаграммы сравнения** (или динамики) являются разновидностью столбиковых диаграмм сравнения. В отличие от последних, масштабная шкала у них расположена по горизонтали.

6. **Знак Варзара**. Это вид диаграмм является разновидностью столбиковых диаграмм сравнения. Он используется для одновременного сравнения трех величин, одна из которых является основанием прямоугольника, вторая высотой, а третья — площадью прямоугольника.

Например, произведение объема перевезенного груза на расстояние перевозки дает грузооборот. Если одну сторону диаграммы брать пропорционально объему перевозки, а другую — расстоянию перевозки, то площадь прямоугольника и представляет собой знак Варзара, т.е. грузооборот.

По данным табл. 4.3. построены знаки Варзара (рис. 4.6).

Таблица 4.3

Перевозки грузов и грузооборот внутреннего водного транспорта общего пользования в 2005 г.

	2005 г
Перевезено грузов — всего, млн.т из них: в международном сообщении	108,3 29,3
Грузооборот — всего, млрд. т.км из них: в международном сообщении	70,9 43,8
Средняя дальность перевозки 1 т груза — всего, км из них: в международном сообщении	654,7 1494,9

Рис. 4.6. Знаки Варзара

7. **Направленные диаграммы** являются разновидностью полосовых диаграмм сравнения. Они отличаются от обычных двусторонним расположением полос, направленные в разные стороны.

Эти диаграммы удобны в тех случаях, когда отдельные объекты сравнения характеризуются противоположными по знаку показателями.

Направленные диаграммы широко используются в демографической статистике при изображении половозрастной структуры населения (рис. 4.7).

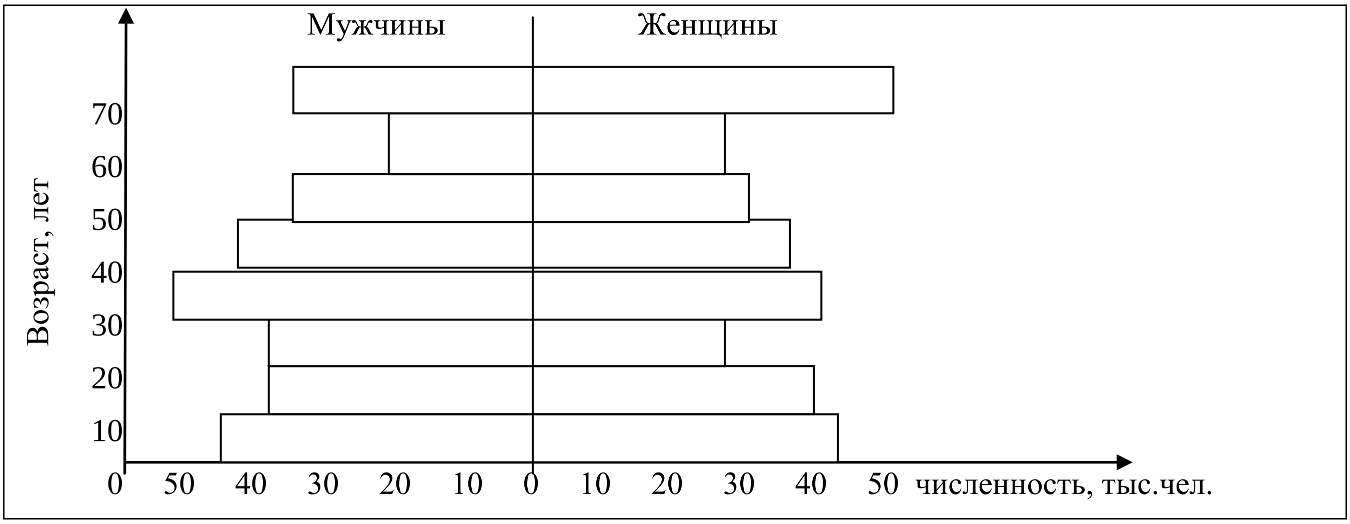


Рис. 4.7. Распределение населения города N России по полу и возрасту в 2008 г.

8. **Пиктограммы** (условные обозначения). В пиктограммах небольшой символ может представлять как один, так и несколько предметов.

Пример 1. В студенческой группе 12 девушек и 7 юношей. В этом случае диаграмма сравнения может выглядеть следующим образом:

Девушки: ☺☺☺☺☺☺☺☺☺☺☺☺

Юноши: ☺☺☺☺☺☺☺

Символ ☺ представляет одного студента.

Пример 2. Численность сотрудников пяти автотранспортных предприятий (АТП) можно представить следующим образом:

АТП 1: △△△△

АТП 2: △△

АТП 3: △△△△△

АТП 4: △△△

АТП 5: △△△△△△△

Символ △ представляет 10 сотрудников.

9. При построении **круговых** или **квадратных** диаграмм исходят из того, что площадь круга (или квадрата) должна быть равна величине изображаемого показателя. Следовательно, радиус круга $R = \sqrt{S/\pi}$, а сторона квадрата $a = \sqrt{S}$, где S — величина изучаемого показателя.

Круговая и квадратная диаграммы являются достаточно распространенными. Однако, на наш взгляд, эти виды графиков не дают наглядного соотношения между изучаемыми явлениями. Например, на рис. 4.8 изображены квадраты А и Б площадью 5 и 10 ед. соответственно. Сторона квадрата А равна $\sqrt{5} = 2{,}24$; сторона квадрата Б: $\sqrt{10} = 3{,}16$.

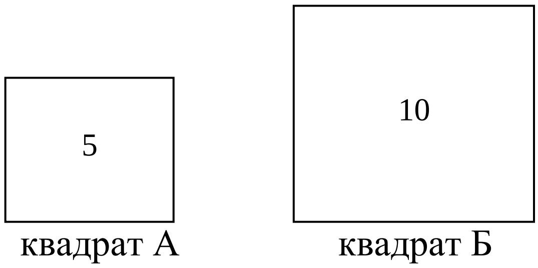


Рис. 4.8. Квадратная диаграмма

Площадь квадрата А точно в два раза больше квадрата Б, однако зрительно, квадрат А не выглядит в два раза больше, т.к. его сторона лишь в 1,45 раза больше стороны квадрата Б, то есть создается, так называемый «оптический обман». Сказанное справедливо и для круговых диаграмм.

10. **Линейные диаграммы** используются для изображения ***динамики явлений***. На оси абсцисс откладывают периоды или даты. На оси ординат — величины изображаемого явления в выбранном масштабе (рис. 4.9).

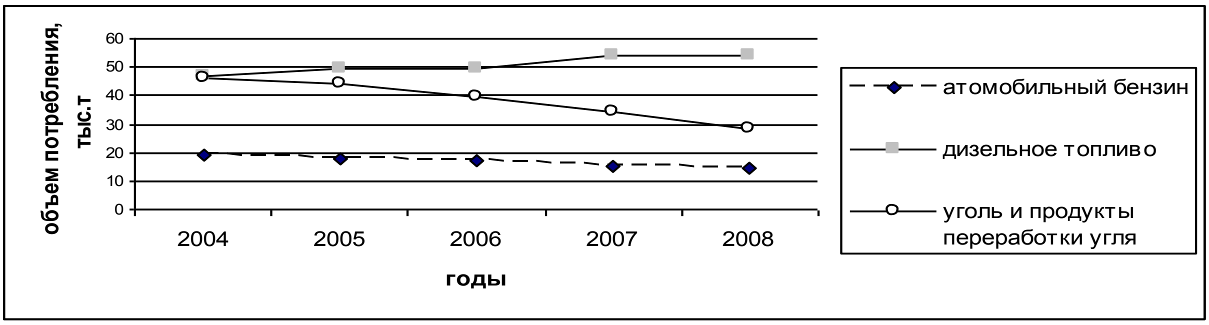


Рис. 4.9. Динамика потребления топлива организациями транспорта в регионе N

11. **Радиальные диаграммы** обычно применяются для иллюстрации сезонных колебаний (гл. 10). Их разделяют на замкнутые и спиральные. **Замкнутые** радиальные диаграммы изображают внутригодичный цикл динамики какого-либо одного года. Спиральные радиальные диаграммы показывают внутригодичный цикл динамики за ряд лет.

Радиальные диаграммы строятся в системе полярных координат. Точкой отсчета является центр круга. Масштабные шкалы являются радиусами круга, расположенными на равных расстояниях друг от друга. Количество шкал соответствует количеству периодов времени. Первая шкала располагается вертикально в верхней половине круга. Затем по часовой стрелке откладываются другие шкалы. На шкалах отличаются соответствующие периодам времени значения показателя и соединяются отрезками.

В замкнутой диаграмме декабрь одного года соединяется с январем этого же года, в спиральной диаграмме — с январем следующего года.

Пример замкнутой радиальной диаграммы приведен на рис. 4.10.

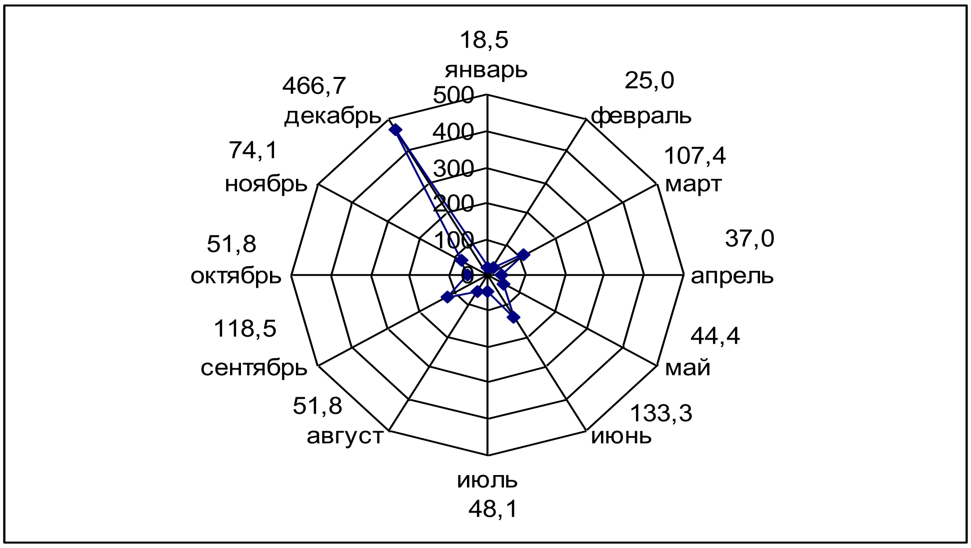


Рис. 4.9. Сезонность ввода в действие жилых домов строительными предприятиями города N (индексы сезонности, %)

12, 13. **Фоновые и точечные картограммы**. На **фоновых** картограммах распределение изучаемого явления на территории изображается различной раскраской территориальных единиц с разной густотой цвета (или штриховкой различной интенсивности). На **точечной** картодиаграмме величина исследуемого показателя изображается в виде точек. Точка соответствует определенному значению, принятому для характеристики единицы совокупности (или определенному количеству единиц).

Вопросы для самоконтроля

1. В чем заключается назначение статистических графиков?
2. Что такое поле графика? Пространственные ориентиры графика?
3. Что называется экспликацией графика?
4. Чем отличаются диаграммы от статистических карт?
5. Назначение структурных диаграмм? Назовите способы построения структурных диаграмм.
6. В чем назначение диаграмм сравнения? Назовите разновидности диаграмм сравнения.
7. Как строятся знаки Варзара?
8. Как строятся радиальные диаграммы?
9. В чем отличие картодиаграммы от картограммы?

5. Абсолютные и относительные величины

Изучая массовые явления и процессы статистика в своих выводах опирается на числовые данные, полученные в конкретных условиях места и времени. Эти данные называются **статистическими показателями**.

В отличие от признака, статистический показатель получается расчетным путем. Это может быть простой подсчет единиц совокупности или более сложные расчеты.

По форме выражения показатели могут быть представлены абсолютными, относительными и средними величинами.

Результаты статистического наблюдения регистрируются в форме абсолютных величин.

Абсолютными величинами называются показатели, выражающие размер или объем того или иного явления в определенное время и на определенной территории.

В статистике все абсолютные величины являются именованными числами, то есть выражаются в единицах измерения, присущих тем или иным общественным явлениям. Единицы измерения могут быть натуральные и **стоимостные (денежные)**.

Натуральные единицы измерения в свою очередь могут быть **простыми** (метры, тонны, штуки и т.д.) и **сложными** (грузооборот транспорта измеряется в тонно-километрах, затраты труда на производство продукции в человеко-часах и т.д.).

Стоимостные (денежные) единицы измерения используются для характеристики статистических показателей в денежном выражении.

Относительными величинами называют величины, выражающие количественные соотношения между социально-экономическими явлениями и их признаками.

Относительные величины **являются отношениями двух абсолютных величин**.

Величину, которую сравнивают, то есть числитель дроби, называют **текущим (или сравниваемым)** показателем.

Величина, с которой производится сравнение, то есть знаменатель дроби, называется **базой сравнения** или **основанием**.

Логической формулой относительной величины является дробь:

$$\text{Относительная величина} = \frac{\text{абсолютная величина } 1(\textit{текущий показатель})}{\text{абсолютная величина } 2(\textit{база сравнения})}$$

В статистике различают несколько видов относительных показателей.

Отношение одноименных величин:

1. относительные величины динамики (*ОВД*);
2. относительные величины планового задания (*ОВПЗ*)

   или просто: относительные величины плана (ОВП);
3. относительные величины выполнения плана (*ОВВП*);
4. относительные величины структуры ($ОВС_{тр}$);
5. относительные величины координации (*ОВК*);
6. относительные величины сравнения (*ОВС*);

Отношение разноименных величин:

7. относительные величины интенсивности (*ОВИ*).

1. **Относительные величины динамики (*ОВД*)**

Динамикой в статистике называют изменение социально-экономического явления во времени. ОПД характеризуют направление и интенсивность изменения показателей во времени.

Относительные показатели динамики также принято называть ***коэффициентами роста***, а если они указываются в процентах - то ***темпами роста***.

В зависимости от того, что принято за базу сравнения, ОВД могут рассчитываться двумя способами.

I способ. За базу сравнения принимается какой-либо один период времени (обычно начальный) и все остальные показатели сравниваются с этим периодом. Так как база (основание) показателя при этом не меняется, то такие показатели называются показателями, рассчитанными на ***постоянной базе*** сравнения (или ***базисными***).

II способ. За базу сравнения принимается период, *предшествующий* изучаемому. В этом случае каждый показатель сравнивается с предыдущим, то есть база сравнения меняется при переходе от одной ОВД к другой, поэтому такие показатели называются показателями, рассчитанными на ***переменной базе*** сравнения (или ***цепными***).

Пример. Имеются данные о величине показателя в различные периоды времени:

2006	2007	2008	2009	2010
y_1	y_2	y_3	y_4	y_5

Система *коэффициентов роста с постоянной базой* сравнения (то есть *базисных*) будет выглядеть следующим образом:

$$K'_{p1} = \frac{y_2}{y_1}; \quad K'_{p2} = \frac{y_3}{y_1}; \quad K'_{p3} = \frac{y_4}{y_1}; \quad K'_{p4} = \frac{y_5}{y_1}.$$

Если полученные базисные коэффициенты роста умножить на 100%, то получим *базисные темпы роста*.

Система *коэффициентов роста с переменной базой* сравнения (т.е. *цепных*), соответственно, примет вид:

$$K_{p1} = \frac{y_2}{y_1}; \quad K_{p2} = \frac{y_3}{y_2}; \quad K_{p3} = \frac{y_4}{y_3}; \quad K_{p4} = \frac{y_5}{y_4}.$$

Умножив цепные коэффициенты роста на 100%, получаем *цепные темпы роста* изучаемого показателя.

2. **Относительные величины планового задания (*ОВПЗ*)** характеризуют отношение планируемого уровня показателя ($y_{пл}$) к фактически достигнутому уровню того периода, по сравнению с которым намечается увеличение или уменьшение показателя (y_1):

$$ОВПЗ = \frac{y_{пл}}{y_1} \cdot 100\%.$$

3. **Относительные величины выполнения плана (*ОВВП*)** используются для контроля за ходом выполнения планов предприятия (фирмы). Процент выполнения плана определяется отношением фактически достигнутого уровня в плановом периоде (y_2) к показателю, установленному планом ($y_{пл}$):

$$ОВВП = \frac{y_2}{y_{пл}} \cdot 100\%.$$

Относительные величины динамики, планового задания и выполнения плана связаны соотношением:

$$ОВД = ОВПЗ \times ОВВП.$$

4. **Относительные величины структуры (*ОВС$_{тр}$*)** характеризуют доли (удельные веса) составных элементов (или группы) в общем итоге. Их считают в процентах или в долях единицы.

$$ОВС_{тр}(d_i) = \frac{i - я\, часть\, совокупности}{вся\, совокупность} \cdot 100\%.$$

$\Sigma d_i = 100\%$ (или 1, если $ОВС_{тр}$ считались в долях единицы).

Пример: в студенческой группе, состоящей из 25-ти человек, 18 студентов изучает английский язык, 7 — немецкий язык. Относительные показатели структуры:

$$d_1 = \frac{18}{25} \cdot 100\% = 72\%$$

$$d_2 = \frac{7}{25} \cdot 100\% = 28\%$$

$$72\% + 28\% = 100\%,$$

таким образом 72% студентов изучают английский язык и 28% — немецкий.

5. **Относительные величины координации (*ОВК*)** отражают соотношение двух частей единого целого. Они показывают, сколько единиц одной части совокупности приходится на 1, 100, 1000 и больше единиц другой части совокупности, взятой за базу сравнения.

В качестве базы сравнения обычно выбирают часть, имеющую наибольший удельный вес в совокупности или являющуюся приоритетной с экономической точки зрения.

Пример 1: часть собственных средств фирмы составляет 70%, а привлеченных — 30%. Тогда относительная величина координации: $OBK = \frac{30}{70} = 0{,}43$. Это означает, что на единицу собственных средств приходится 0,43 привлеченных.

Пример 2: Этот пример решается аналогично, если вместо удельных весов собственных и привлеченных средств даны абсолютные величины.

6. **Относительные величины сравнения (*ОВС*)** — это результат сопоставления одноименных показателей, относящихся к разным совокупностям за один и тот же период времени.

Пример. Имеются данные о количестве подключенных терминалов сотовой связи в 2006 г. (табл. 5.1).

Таблица 5.1

на конец 2006 г., штук

Наименование субъекта РФ	Число подключенных терминалов сотовой связи на 1000 человек населения
г. Москва и Московская область	15554,4
Ярославская область	1132,2
Тверская область	1066,8
Тамбовская область	848,9
Воронежская область	842,7
…	

Сравним данные по г. Москве и Московской области, занимающих среди субъектов РФ первое место, и Тамбовской области:

$$OBC = \frac{1554{,}4}{848{,}9} = 1{,}83.$$

По количеству подключенных в 2006 г. терминалов г. Москва и Московская область превосходят Тамбовскую область в 1,83 раза, или на 705,5 терминалов (1554,4—848,9).

8. **Относительные величины интенсивности (*ОВИ*)** характеризуют степень распространенности или развития того или иного явления в определенной среде. Эти величины могут быть получены как отношение разноименных величин, определенным образом взаимосвязанных (плотность населения на кв. км, производство той или иной продукции на душу населения и т.п.). Все эти величины являются именованными.

Вопросы для самоконтроля

1. В чем отличие понятий «статистический показатель» и «статистический признак»?
2. Какие величины относятся к абсолютным величинам?
3. Какие величины называются относительными величинами?
4. Что такое «база сравнения»?
5. Назовите виды относительных величин.
6. Назовите способы расчета относительных величин динамики.
7. Какая взаимосвязь существует между относительными величинами динамики, планового задания и выполнения плана?
8. Что характеризуют относительные величины сравнения? координации?
9. Как рассчитываются показатели сравнения?
10. Какие относительные величины являются именованными?

6. Средние величины

6.1. *Основные понятия*

Рассмотренные в предыдущей главе относительные величины позволяют одновременно сравнивать между собой только две абсолютные величины.

Однако часто возникают ситуации, когда необходимо получить обобщающую характеристику всех единиц, входящих в изучаемую совокупность, то есть среднюю величину.

Средней величиной в статистике называется обобщающий показатель, который характеризует типичный уровень варьирующего (изменяющегося) признака в расчете на единицу совокупности в конкретных условиях места и времени.

Средняя величина — всегда именованная. Она имеет ту же размерность, что и признак у единиц изучаемой совокупности. Средняя величина обозначаются символом $\bar{x}$.

В статистике используются различные виды средних величин, которые подразделяются на две категории:

– степенные средние;
– структурные средние.

6.2. *Степенные средние*

К этой группе средних величин относятся:

– средняя арифметическая $(\bar{x})$;
– средняя гармоническая $(\bar{x}_{гарм})$;
– средняя геометрическая $(\bar{x}_{геом})$;
– средняя квадратическая $(\bar{x}_{кв})$.

Рассмотрим более подробно некоторые виды степенных средних.

6.2.1. Средняя арифметическая $(\bar{x})$

Средняя арифметическая $(\bar{x})$ является наиболее распространенным видом средних величин. Она вычисляется как сумма всех значений, деленная на количество значений.

Для небольшого числа значений, или если данные не сгруппированы, средняя арифметическая рассчитывается по формуле **простой средней арифметической**:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{\sum x}{n}^{*)},$$

где x_i или x — индивидуальные значения признака. В ряде распределения отдельное значение признака называется **вариантой;**

n — число единиц изучаемой совокупности.

Пример: Найдите среднюю арифметическую следующих чисел: 5, 3, 0, 1, 8.

$$\bar{x} = \frac{5+3+0+1+8}{5} = \frac{17}{5} = 3{,}4$$

Если отдельные значения признака повторяются, то исходные данные лучше сначала упорядочить в виде ряда распределения, а затем воспользоваться формулой **взвешенной средней арифметической**:

$$\bar{x} = \frac{\sum x_i f_i}{\sum f_i},$$

где x_i — индивидуальные значения признака;

f_i — частота повторения индивидуального значения признака x_i (или «вес» признака).

Пример: Найти среднюю арифметическую чисел:2, 4, 8, 3, 3, 0, 2, 2, 0, 2, 3, 3, 3.

---

*)для удобства и простоты записи знак $\sum_{i=1}^{n}$ в дальнейшем заменен знаком $\sum$.

Построим ряд распределения (табл. 6.1.)

Таблица 6.1

x_i	f_i	$x_i f_i$
0	2	0
2	4	8
3	5	15
4	1	4
8	1	8
Всего	13	35

Рассчитаем среднюю арифметическую.

$$\bar{x} = \frac{\sum x_i f_i}{\sum f_i} = \frac{35}{13} = 2{,}69.$$

Аналогичный результат можно было бы получить суммируя все значения и разделив их затем на количество значений, т.е. используя формулу простой средней арифметической:

$$\bar{x} = \frac{\sum x}{n} = \frac{0+0+2+2+2+2+3+3+3+3+3+4+8}{13} = 2{,}69.$$

Средняя арифметическая обладает рядом свойств. Важнейшие из них следующие:

1. Произведение средней на сумму частот всегда равно сумме произведения индивидуальных значений признака на соответствующие частоты:

$$\bar{x}\sum f_i = \sum x_i f_i .$$

2. Сумма всех отклонений индивидуальных значений признака от среднего значения равна нулю:

$$\sum (x_i - \bar{x}) f_i = 0.$$

3. При увеличении или уменьшении всех частот в одно и то же число раз средняя арифметическая не меняется.
4. При увеличении или уменьшении каждого значения признака в одно и то же число раз средняя арифметическая увеличивается или уменьшается во столько же раз.

5. При уменьшении или увеличении каждого значения признака на какое-либо произвольное число, средняя арифметическая уменьшается или увеличивается на это же число.

6.2.2. Средняя гармоническая $\left(\bar{x}_{гарм}\right)$

Средняя гармоническая простая $\left(\bar{x}_{гарм}\right)$ — это обратная к средней арифметической из обратных значений признаков. Ее вычисляют, когда необходимо осреднение обратных индивидуальных значений признаков путем их суммирования (например, в случаях определения средних расходов времени, труда или материалов на единицу продукции).

Средняя гармоническая взвешенную $\left(\bar{x}_{гарм}\right)$ вычисляют тогда, когда известны данные об общем объеме признака ($w = xf$), а также индивидуальные значения признака (x), неизвестной является лишь частота (f).

Формулы средней гармонической (простой и взвешенной) имеют вид:

простая: $$\bar{x}_{гарм} = \frac{n}{\sum \frac{1}{x_i}} = \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \ldots + \frac{1}{x_n}};$$

взвешенная: $$\bar{x}_{гарм} = \frac{\sum w_i}{\sum \frac{w_i}{x_i}} = \frac{w_1 + w_2 + \ldots + w_n}{\frac{w_1}{x_1} + \frac{w_2}{x_2} + \ldots + \frac{w_n}{x_n}}.$$

При расчете одного и того же показателя в зависимости от исходной информации могут быть использованы либо средняя арифметическая, либо средняя гармоническая.

Рассмотрим это на следующем примере.

Пример: по данным таблицы 6.2 определите среднюю цену на товар М.

Таблица 6.2

	Цена за единицу товара М, руб., x_i	Выручка от продажи товара М в магазинах города, тыс.руб., w_i
Магазин А	300	300,0
Магазин Б	280	350,0
Магазин В	350	339,5
Всего	—	989,5

Расчет средней цены можно записать отношением:

$$\text{Средняя цена за единицу товара} = \frac{\text{Выручка от продажи всех товаров}}{\text{Общее количество проданных товаров}}$$

Сумма выручки, то есть показателя, который находится в числителе, известна.

Для определения неизвестной величины — количества проданных товаров — необходимо отдельно по каждому магазину разделить выручку на цену, тогда:

$$\bar{x}_{гарм} = \frac{\sum w_i}{\sum \frac{w_i}{x_i}} = \frac{989500}{\frac{300000}{300} + \frac{35000}{280} + \frac{339500}{350}} = 307{,}3 \text{ руб.}$$

Средняя цена товара М — 307,3 руб.

Как видно из примера, средняя гармоническая является превращенной формой средней арифметической. Вместо гармонической средней можно было бы использовать среднюю арифметическую, но для этого сначала необходимо определить веса отдельных значений признака.

В рассмотренном примере частоты (то есть количество проданных каждым магазином товаров) будут равны:

по магазину А: $$f_1 = \frac{w_1}{x_1} = \frac{300000}{300} = 1000 \text{ ед.};$$

по магазину Б: $$f_2 = \frac{w_2}{x_2} = \frac{3500000}{280} = 1250 \text{ ед.};$$

по магазину В: $$f_3 = \frac{w_3}{x_3} = \frac{339500}{350} = 970 \text{ ед.}$$

Тогда средняя цена товара будет равна:

$$\bar{x} = \frac{\sum x_i f_i}{\sum f_i} = \frac{300 \cdot 1000 + 280 \cdot 1250 + 350 \cdot 970}{1000 + 1250 + 970} = 307{,}3 \text{ руб.}$$

В том случае, если объемы явлений (то есть произведения по каждому признаку $w_1 = w_2 = w_n$) **равны, применяется средняя гармоническая простая.**

Пример. Предположим, в фирме, специализирующейся на упаковке и отправке товаров, заняты два работника. Первый из них затрачивает на обработку заказа 5 минут, второй — 15 минут. Каковы средние затраты времени на 1 заказ, если общее время работы у них одно и то же? Предположим, что оба работают 1 час.

Общая формула средней будет выглядеть следующим образом:

$$\text{Среднее время на выполнение одного заказа} = \frac{\text{Общие затраты времени на выполнение всех заказов}}{\text{Общее количество выполненных заказов}}$$

Тогда $\bar{x}_{гарм} = \dfrac{60+60}{\dfrac{60}{5}+\dfrac{60}{15}} = 7{,}5$ мин.

В данном примере рассчитывать среднее значение по формуле средней арифметической простой — неверно, так как в течение одного и того же промежутка времени — 1 часа рабочим будет выполнено различное количество заказов.

Проверим эти расчеты: если считать среднее время по средней арифметической простой, то среднее время выполнения заказа будет равно $\bar{x} = (5+15):2 = 10$ мин. По свойству средней арифметической мы знаем, что произведение средней на сумму частот будет равно сумме произведений отдельных значений признака на соответствующие частоты: $\bar{x}\sum f_i = \sum x_i f_i$.

Рассчитаем частоты (f_i), то есть определим, сколько заказов может выполнить каждый работник за один час работы. Первый работник: $f_1 = \dfrac{60}{5} = 12$ заказов, второй работник: $f_2 = \dfrac{60}{15} = 4$ заказа. Оба выполняют за час 16 заказов.

Таким образом:

$10 \cdot 16 \neq 5 \cdot 12 + 15 \cdot 4$ (первое свойство средней арифметической):

$160 \neq 120$.

Значит расчет с помощью средней арифметической простой, в данном случае не верен.

6.2.3. Средняя геометрическая $(\bar{x}_{геом})$

Средняя геометрическая $(\bar{x}_{геом})$ используется для расчета среднего коэффициента роста в рядах динамики.

Если между коэффициентами роста равные промежутки времени, то используется формула простой геометрической:

$$\bar{x}_{геом} = \sqrt[n]{x_1 \cdot x_2 \cdot \ldots \cdot x_n} = \sqrt{\prod (x_i)}.$$

Если коэффициенты роста относятся к периодам различной продолжительности, то средний коэффициент роста за весь период определяется по формуле взвешенной геометрической:

$$\bar{x}_{геом} = \sqrt[\sum f_i]{x_1^{f_1} \cdot x_2^{f_2} \cdot \ldots \cdot x_n^{f_n}} = \sqrt[\sum f_i]{\prod \left(x_i^{f_i}\right)}.$$

Более подробно этот показатель рассмотрен в главе 10.

6.2.4. Средняя квадратическая $(\bar{x}_{кв})$

Средняя квадратическая $(\bar{x}_{кв})$ используется в тех случаях, когда необходимо найти среднюю для величин, выраженных в виде квадратных функций, например, средние диаметры колес, труб, средние стороны квадратов и т.д. Также она используется для расчета среднего квадратического отклонения (σ) и дисперсии (σ^2), являющихся показателями колеблемости значений признака в совокупности.

Средняя квадратическая простая: $\bar{x}_{кв} = \sqrt{\frac{x_i^2}{n}}$,

средняя квадратическая взвешенная: $\bar{x}_{кв} = \sqrt{\frac{\sum x_i^2 f_i}{\sum f_i}}$.

При расчете различных степенных средних по одним и тем же данным средние не будут одинаковы.

Чем выше степень средней, тем больше ее величина.

6.3. *Структурные средние*

К структурным средним величинам относятся мода (M_o) и медиана (M_e).

Этот вид средних величин обычно используется в тех случаях, когда имеющейся информации недостаточно для расчета степенных средних или расчет степенных средних нецелесообразен. В отличие от степенных средних мода и медиана не зависят от крайних значений признака в совокупности.

6.3.1. Мода

Мода (M_o) — это такое значение признака, которое чаще всего повторяется в ряду распределения. Мода применяется при экспертных оценках, при определении наиболее ходовых размеров одежды, обуви и др., что учитывается при планировании.

Пример. В группе из 12-ти студентов получены следующие оценки за экзамен:

3, 3, 4, 4, 2, 3, 5, 5, 4, 4, 4, 4

Модой будет являться оценка 4 балла, так как она встречается чаще всего — 6 раз (2 балла встречаются 1 раз, 3 балла — 3 раза, 5 баллов — 2 раза).

В ряду распределения модой будет значение признака, имеющего наибольшую частоту.

Пример. Имеются данные о составе и количестве судов судоходной компании (табл. 6.3).

Таблица 6.3

Состав флота судоходной компании

Грузоподъемность судов, т	800	1000	1200	1500	1800	3000	Всего
Число судов данной грузоподъемности, ед.	3	1	2	3	8	1	18

Модой в данном случае будет грузоподъемность 1800 т, так как частота повторений этого признака наибольшая — 8 раз.

Определение моды в интервальном ряду распределения

В интервальном ряду распределения моду можно определить двумя способами:

1) графическим способом (построением гистограммы)

2) расчетным путем.

Пример. В таблице 6.4 приведены результаты опроса 30 сотрудников фирмы о времени, затрачиваемом ими утром на дорогу до работы. Определите моду данного распределения.

Таблица 6.4

Время, мин	0—20	20—40	40—60	60—80	80—100	Всего
Количество сотрудников, чел.	1	5	7	12	5	30

1. Графический способ

Построим гистограмму распределения (рис. 6.2).

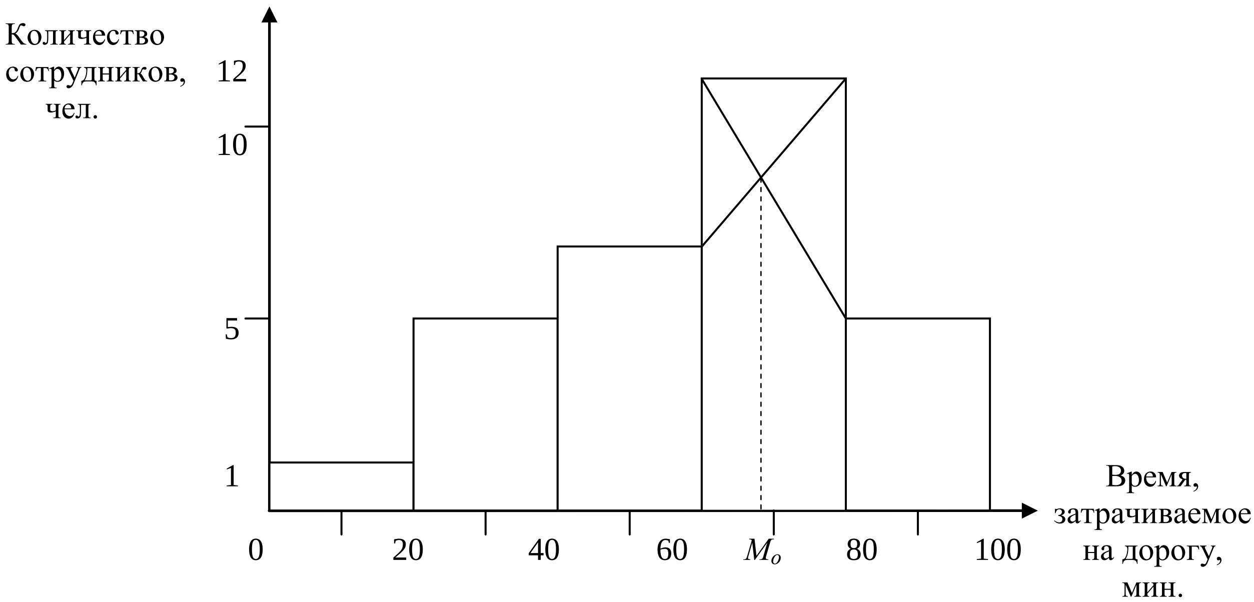


Рис. 6.2. Гистограмма распределения времени, затрачиваемого сотрудниками на дорогу

Определим модальный интервал. **Модальным интервалом** называется интервал, которому соответствует столбец графика с максимальной высотой. В нашем примере это интервал 60—80.

Для определения моды проводятся диагонали через вершины самого высокого столбца (модального интервала), как показано на гистограмме. Модой будет являться абсцисса точки пересечения этих двух линий.

2. Расчетный способ:

$$M_o = x_{M_o} + h\frac{\left(f_{M_o} - f_{M_o-1}\right)}{\left(f_{M_o} - f_{M_0-1}\right)+\left(f_{M_o} - f_{M_0+1}\right)},$$

где x_{M_0} — нижняя граница модального интервала;

h — ширина интервала;

f_{M_o-1} — частота интервала, предшествующего модальному;

f_{M_o} — частота модального интервала;

f_{M_o+1} — частота интервала, следующего за модальным.

Для нашего примера:

$$M_o = 60 + 20\frac{12-7}{(12-7)+(12-5)} = 68{,}3\text{мин.}$$

Мода и средняя величина по-разному характеризуют совокупность. Мода определяет размер признака, свойственный, хотя и значительной части, но все же не всей совокупности.

Мода по своему обобщающему значению менее точна по сравнению со средней арифметической, характеризующей совокупность в целом с учетом всех без исключения элементов совокупности.

6.3.2. Медиана

Медианой (*M_e*) называют такое значение признака, которое находится в середине упорядоченного (по возрастанию или убыванию) ряда значений, т.е. одна половина значений будет меньше медианы, а другая — больше.

Медиана используется при статистическом контроле качества продукции и технологического процесса на промышленных предприятиях, широко применяется в демографической статистике.

Определение медианы в дискретном ряду распределения

Если число значений n в ряду распределения *нечетное*, то медианой будет являться серединное (центральное) значение, т.е. $\frac{1}{2}(n+1)$-е значение.

Пример. Найти медиану чисел

2, 0, 4, 8, 5

Сначала упорядочим данные, представив их в виде ряда распределения: 0, 2, 4, 5, 8.

Имеется 5 значений, то есть n = 5.

Медианой является $\frac{1}{2}(5+1)=3$-е значение, которое равно 4. Следовательно, M_e = 4.

Если число значений в ряду распределения n *четное*, то порядковый номер медианы будет иметь дробное значение. В этом случае в качестве медианы принимают среднюю арифметическую из двух серединных значений.

Пример. Найти медиану чисел

8, 5, 7, 1, 3, 9, 2, 10

Упорядочим данные:

1, 2, 3, 5, 7, 8, 9, 10

Имеется 8 значений, то есть n = 8.

Порядковый номер медианы будет равен: $\frac{1}{2}(8+1)=4{,}5$. Следовательно, медиана находится между 4-м и 5-м значениями.

4-е значение равно 5;

5-е значение равно 7.

Медиана определяется как средняя арифметическая этих значений:

$$M_e=\frac{5+7}{2}=6.$$

Если данные представлены в табличной форме, то медиану удобнее находить по накопленным частотам. Место медианы в ряду распределения (то есть ее порядковый номер) определяется также: $\frac{1}{2}(n+1)$.

Пример. Найти медиану ряда распределения, приведенного в таблице 6.5.

Таблица 6.5

Значение признака, x	0	1	2	5	9	10	Всего
Частота повторений, f	4	6	8	12	7	3	40

Суммарная частота равна 40, то есть n = 40. Порядковый номер медианы будет $\frac{1}{2}(40+1)=20{,}5$, значит, сама медиана будет находиться между 20-м и 21-м по счету значением признака x.

Как определить, какое именно значение находится на 20-м, а какое на 21-м месте? Для этого рассчитаем накопленные частоты (т.е. будем последовательно суммировать все частоты f_i) — табл. 6.6.

Таблица 6.6

Значение признака, x_i	0	1	2	5	9	10	Всего
Частота повторений, f_i	4	6	8	12	7	3	40
Накопленные частоты, f_i^H	4	10 (4 + 6)	18 (10+8)	30 (18+12)	37 (30+7)	40 (37+3)	—

Из таблицы можно легко увидеть, что

первые 4 значения, то есть с 1-го по 4-е все равны 0;

следующие 6 значений, то есть с 5-го по 10-е все равны 1;

следующие 8 значений, то есть с 11-го по 18-е все равны 2;

следующие 12 значений, то есть с 19-го по 30-е все равны 5;

следующие 7 значений, то есть с 31-го по 37-й все равны 9;

следующие 3 значения, то есть с 38-го по 40-й все равны 10.

Следовательно, 20-е значение равно 5; 21-е также равно 5. Медиана — это средняя арифметическая между ними: $\frac{5+5}{2}=5$.

Определение медианы в интервальном ряду распределения

Для определения медианы в интервальном ряду сначала находится медианный интервал.

Медианным интервалом называется первый интервал, накопленная частота которого равна или превышает полусумму частот.

Внутри медианного интервала медиана определяется по формуле:

$$M_e = x_{M_e} + h\frac{0{,}5\sum f - f_{M_e}^H}{f_{M_e}},$$

где x_{M_e} — нижняя граница модального интервала;

h — ширина интервала;

$0{,}5\sum f$ — полусумма частот;

$f_{M_e}^{H}$ — накопленная частота интервала, предшествующего медианному;

f_{M_e} — частота медианного интервала.

Пример. Используя данные табл. 6.7, определить медианное время, затрачиваемое сотрудниками фирмы на дорогу до работы.

Таблица 6.7

Время, мин, x_i	0—20	20—40	40—60	60—80	80—100	Всего
Количество сотрудников, чел., f_i	1	5	8	12	4	30
Накопленная частота, f_i^H	1	6	14	26	30	—

Определим медианный интервал (то есть интервал, в котором находится медианное значение):

полусумма частот равна: $0{,}5\sum f = 0{,}5 \cdot 30 = 15$.

Первые три интервала не подходят, так как их накопленные частоты (соответственно, 1, 6 и 14) меньше 15. Накопленная частота четвертого интервала равна 26. 26 > 15, следовательно, медианные интервал — [60—80].

$$M_e = 60 + 20\frac{0{,}5 \cdot 30 - 14}{12} = 61{,}7 \text{ мин.} \approx 62 \text{ мин.}$$

Таким образом, половина сотрудников фирмы тратит на дорогу в среднем до 62 мин., другая половина — от 62 мин. до 100 мин.

Вопросы для самоконтроля

1. Что представляет собой средняя величина?
2. Назовите виды средних величин.
3. Напишите формулу средней арифметической и приведите пример расчета средней по формуле средней арифметическое простой; средней арифметической взвешенной.
4. Назовите основные свойства средней арифметической.

5. В каких случаях используется средняя гармоническая? Приведите формулы расчета средней гармонической простой и взвешенной.
6. В каких случаях используется средняя геометрическая и как она вычисляется?
7. В каких случаях используется средняя квадратическая? Формулы для ее вычисления.
8. Что называется модой? Как рассчитывается мода в дискретных рядах распределения?
9. Определение моды в интервальных рядах распределения.
10. Что называется медианой? Как рассчитывается медиана в дискретных рядах распределения?
11. Как определяется медиана в интервальном ряда распределения?

7. Ряды распределения и их характеристики

7.1. *Основные понятия*

Как уже отмечалось в главе 2, одной из разновидностей простых группировок являются ряды распределения.

Ряды распределения представляет собой упорядоченную по какому-либо признаку совокупность единиц.

В зависимости от того, какой признак положен в основу группировки, ряды распределения делят на:

1. **Атрибутивные ряды распределения,** когда группировочный признак является качественным (табл. 7.1):

Таблица 7.1

Наличие речных и озерных судов России (на конец 2005 г.)

Суда	Количество, ед.
Всего в т.ч.	21 532
– пассажирские и грузопассажирские	1 957
– сухогрузные	2 120
– наливные	603
– буксирные	8 529
– рейдовые	176
– вспомогательные	8 147

2. **Вариационные ряды распределения**, когда в основе группировки лежит количественный признак.

В дальнейшем будет идти речь только о вариационных рядах распределения, которые для краткости будем называть просто рядами распределения.

Пример 1: построить ряд распределения для следующих чисел: 2, 8, 4, 3, 10, 5.

Ряд распределения будет выглядеть так:

– упорядоченный по возрастанию: 2, 3, 4, 5, 8, 10;

или

– упорядоченный по убыванию: 10, 8, 5, 4, 3, 2.

Если значения признака повторяются, то ряд распределения удобней представить в виде таблицы, состоящей из двух граф. Первая графа — сгруппированные значения признака, вторая графа — численность групп (графа частот).

Пример 2. Построить ряд распределения для следующих значений

2, 8, 2, 4, 4, 3, 4, 4, 2, 2, 8, 4

Ряд распределения в табличной форме будет выглядеть следующим образом (табл. 7.2):

Таблица 7.2

Ряд распределения

Значения признака, x_i	Частота повторений, f_i
2	4
3	1
4	5
8	1
Всего	**11**

Числовые значения признака, встречающиеся в изучаемой совокупности, называются **вариантами**. В первом примере это числа 2, 3, 4, 5, 8 и 10, а во втором примере — это числа 2, 3, 4, и 8. Их принято обозначать символом x_i.

Значения, показывающие численность выделенных групп, называются **частотами**. Во втором примере это числа 4, 1 и 5. В рядах распределения частоты обозначаются символом f_i.

Частоты, выраженные в виде относительных величин (долей единиц или процентов) называют **частостями** f_i. $\sum f_i' = 1$ или $\sum f_i' = 100\%$ (табл. 7.3)

В ряде случаев при анализе рядов распределения необходимо рассчитывать накопленные частоты.

Накопленные частоты $\left(f_i^H\right)$ — это частоты, образующиеся путем последовательного прибавления к частоте первой варианты последующих частот. Накопленные частоты показывают, сколько единиц совокупности имеют значение признака не больше, чем данное значение. Например, по данным табл. 7.3 можно сказать, что значения (варианты) 2 и 3 встречаются не больше 5-ти раз из 11-ти.

В табл. 7.3 приведен пример расчета частостей и накопленных частот.

Таблица 7.3

Ряд распределения

Варианты, x_i	Частоты, f_i	Частости,		Накопленные частоты, $\left(f_i^H\right)$
		в процентах, %	в долях единиц	
2	4	$36{,}4\left(\frac{4}{11}\cdot 100\%\right)$	0,364(4:100)	4
3	1	$9{,}1\left(\frac{1}{11}\cdot 100\%\right)$	0,091(1:11)	5(4+1)
4	5	$45{,}4\left(\frac{5}{11}\cdot 100\%\right)$	0,454(5:11)	10(5+5)
8	1	$9{,}1\left(\frac{1}{11}\cdot 100\%\right)$	0,091(1:11)	11(10+1)
Всего	11	100	1,000	—

В зависимости от способа построения ряды распределения могут быть дискретными или интервальными.

Дискретный ряд распределения — это ряд, построенный по признакам, имеющим прерывные изменения. Например, ряд распределения семей по количеству детей в семье; рабочих — по тарифному разряду, квартир — по количеству комнат и т.п. Примером дискретного ряда распределения могут служить данные табл. 7.2.

Интервальный ряд распределения — это ряд, построенный для непрерывных признаков, то есть таких, которые могут принимать в определенном интервале любые значения, в том числе и дробные. Например, ряд распределения служащих по стажу работы, возрасту; величине дохода и т.п. Примером интервального ряда распределения могут служить данные табл. 7.4.

7.2. *Графическое изображение рядов распределения*

Для изображения рядов распределения используют 4 вида графиков: полигон распределения, гистограмму распределения, кумуляту и огиву.

Полигон распределения используются для изображения дискретных рядов распределения. По оси X откладывают варианты; по оси Y — соответствующие им частоты. Полученные значения соединяют ломаной линией. Крайние точки полученной ломаной соединяются с лежащими на оси X следующими (меньшими или большими) возможными, но практически не наблюдающимися значениями признака, частота которых, очевидно, равна 0.

Пример. Построить полигон распределения для данных табл. 7.2

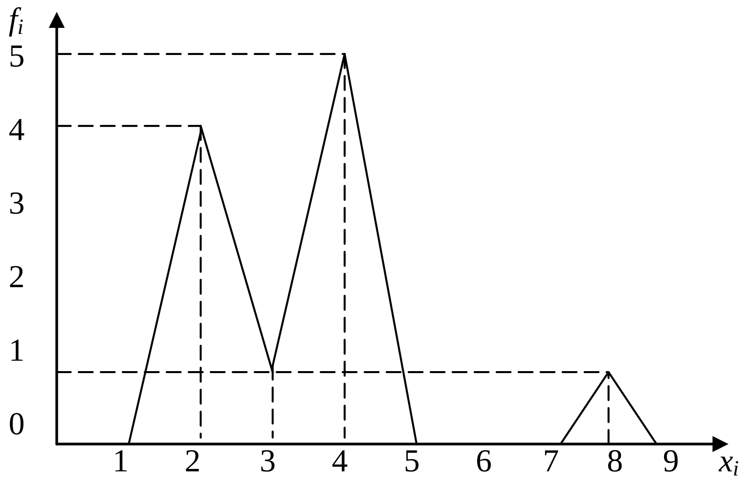


Рис. 7.1. Полигон распределения дискретного ряда

Гистограмма распределения служит для изображения интервальных рядов распределения. По оси X откладываются интервалы, а по оси Y — соответствующие частоты. По данным табл. 7.4 на рис. 7.2 построена гистограмма распределения.

Таблица 7.4

Интервальный ряд распределения

Группы	Частоты, f_i	Накопленные частоты, $\left(f_i^H\right)$
2—5 (–)	2	2
5—8	4	6
8—11	9	15
11—14	5	20
Всего	**20**	—

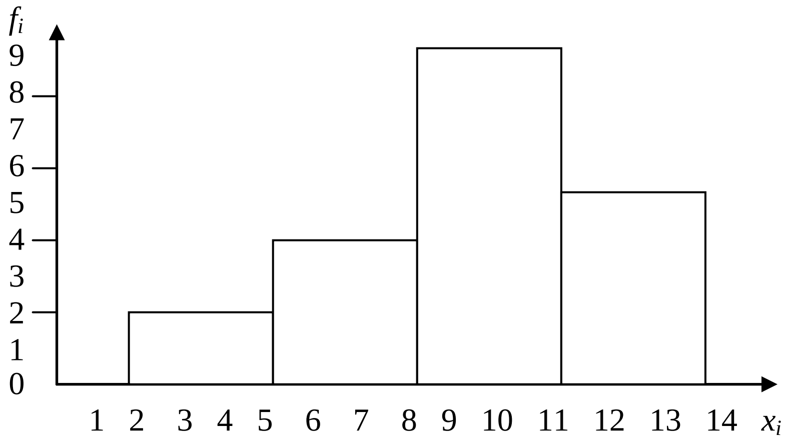


Рис. 7.2. Гистограмма распределения интервального ряда

Кумулята строится для характеристики процесса концентрации явления. По оси X откладываются варианты или интервалы, а по оси Y — **накопленные** частоты. Полученные точки соединяют ломаной линией.

При построении кумуляты по **интервальным рядам** распределения накопленные частоты отмечаются в верхних границах интервалов.

Пример 1. Построить кумуляту распределения для данных табл. 7.3 (дискретный ряд).

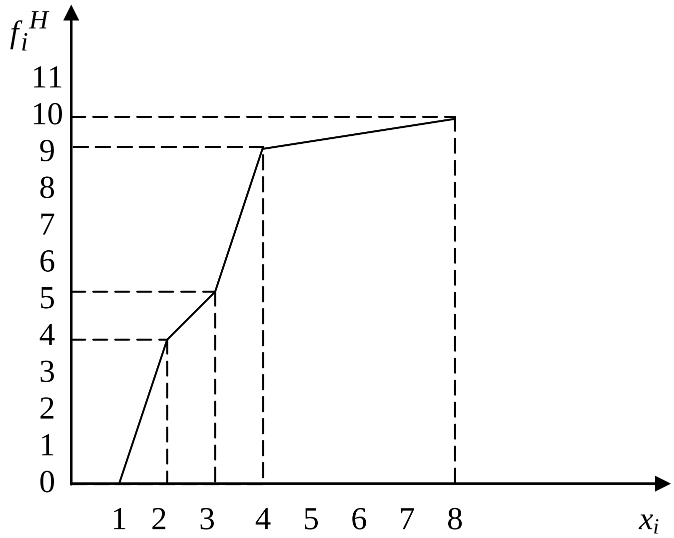


Рис. 7.3. Кумулята распределения дискретного ряда

Пример 2. Построить кумуляту распределения по данным табл. 7.4 (интервальный ряд).

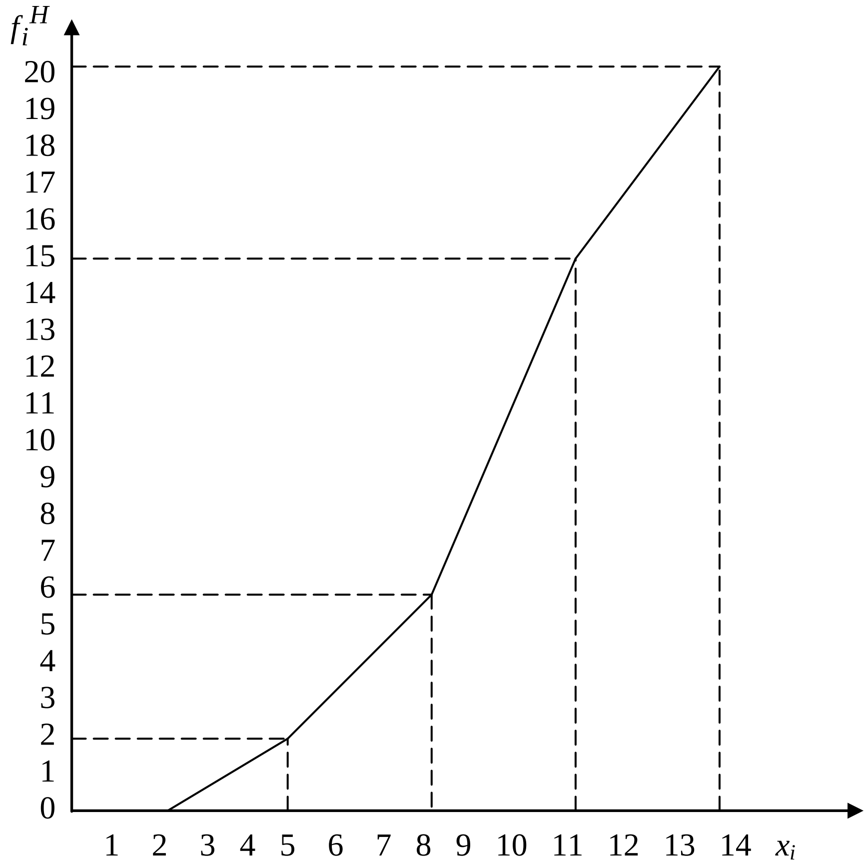


Рис. 7.4. Кумулята распределения интервального ряда

Огива получается, если поменять местами оси кумуляты.

При изучении процессов концентрации (концентрации капитала; концентрации производства; концентрации доходов населения и др.) **используется графическое изображение вариационного ряда в виде кривой Лоренца**.

Для ее построения и значения признака и частоты выражаются в относительных величинах (долях или процентах к итогу). Затем рассчитываются их накопленные значения. При построении графика на оси X откладываются накопленные частоты, а на оси Y — накопленные относительные величины значений признака. Соединив все точки прямыми линиями, получают кривую линию, характеризующую степень неравномерности распределения (кривую Лоренца). Линия, соединяющая нижний левый угол графика с верхним правым (диагональ четырехугольника), является линией равномерного распределения. Чем больше кривая Лоренца отличается от диагонали, тем больше неравномерность в распределении значений признака в совокупности.

Пример. Используя данные табл. 7.5 графически представим процесс концентрации с помощью кривой Лоренца.

Таблица 7.5

Распределение туристических фирм города N по численности работающих и расчет данных для построения кривой Лоренца

Группы турфирм по численности сотрудников, чел.	Середина интервала, чел. x_i	Количество турфирм, f_i	$x_i f_i$	Удельный вес группы (частность)		Накопленные частости	
				по числу сотрудников фирмы, $\frac{x_i f_i}{\sum x_i f_i}$	по числу фирм, $f_i' = \frac{f_i}{\sum f_i}$	по числу сотрудников фирмы, q_i	по числу фирм, p_i
1—5 (+)	3	9	27	0,06	0,24	0,06	0,24
6—10	8	10	80	0,20	0,27	0,26	0,51
11—15	13	8	104	0,25	0,22	0,51	0,73
16—20	18	6	108	0,26	0,16	0,77	0,89
21—25	23	4	92	0,23	0,11	1,000	1,00
Всего	—	**37**	**411**	**1,00**	**1,00**	—	—

Графическое изображение степени концентрации турфирм по численности сотрудников представлено на рис. 7.5.

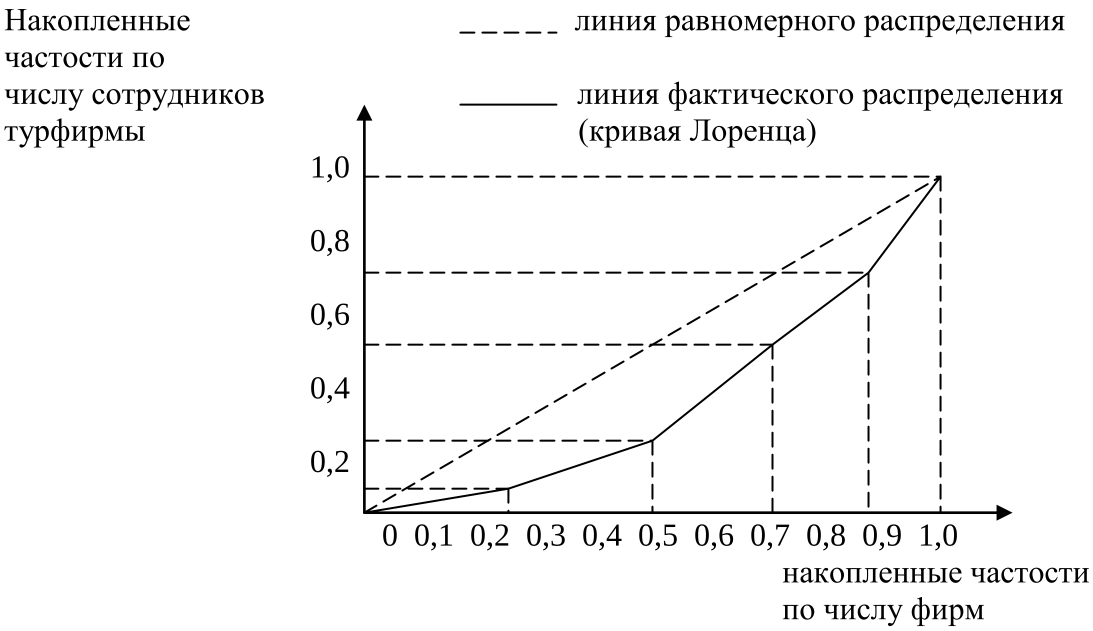


Рис. 7.5. Графическая интерпретация степени концентрации численности работников туристических фирм

Пунктирная линия на графике (линия равномерного распределения) означает, что определенной доли туристических фирм соответствует такая же доля численности сотрудников (т.е. 10% турфирм имели бы 10% сотрудников от общего числа, 20% турфирм — соответственно 20% сотрудников и т.д.). Чем больше отклонение кривой Лоренца от линии равномерного распределения, тем более неравномерно распределен объем изучаемого показателя между единицами (или группами) статистической совокупности.

Графический метод может использоваться для сопоставления кривых Лоренца, характеризующих концентрацию явлений в различных совокупностях, или при анализе процесса концентрации какого-либо показателя в динамике.

Такие сопоставления достаточно широко распространены в социально-экономической статистике, например, при изучении распределения объема доходов между различными группами населения, при анализе степени концентрации банковского капитала и т.д.

Количественная оценка степени концентрации значений признака дается с помощью **коэффициента концентрации (индекса Джини)**, который рассчитывается на основе кривой Лоренца.

Коэффициент концентрации (индекс Джини) — G рассчитывается по формуле:

$$G = \sum_{i=1} p_i \cdot q_{i+1} - \sum_{i=1} p_{i+1} \cdot q_i \,,$$

где p_i — накопленные частости (доли) единиц распределения $f_i^{н}$;

q_i — накопленные частости (доли) единиц суммарного показателя.

Коэффициент концентрации может принимать значения от 0 до 1; при равномерном распределении $G = 0$; чем больше степень концентрации, больше величина индекса.

Пример. Рассчитаем индекс Джини для рассмотренного выше примера (табл. 7.5).

$$G = \left[0{,}24 \cdot 0{,}260 + 0{,}51 \cdot 0{,}51 + 0{,}73 \cdot 0{,}77 + 0{,}89 \cdot 1{,}00\right] -$$
$$- \left[0{,}51 \cdot 0{,}06 + 0{,}73 \cdot 0{,}26 + 0{,}89 \cdot 0{,}51 + 1{,}00 \cdot 0{,}77\right] = 0{,}33.$$

Полученное значение свидетельствует о средней степени неравномерности распределения сотрудников.

Для анализа вариационных рядов распределения используются три группы показателей:

1. показатели центра распределения;
2. показатели вариации (колеблемости);
3. показатели формы распределения.

7.3. *Показатели центра распределения*

Показатели центра распределения служат для характеристики среднего значения признака в вариационном ряду распределения. К ним относятся: средняя арифметическая, мода и медиана.

Расчет этих показателей рассмотрен в главе 6.

7.4. *Показатели вариации (колеблемости) признака и правило сложения дисперсий*

7.4.1. Показатели вариации признака

Средняя величина даёт обобщающую характеристику совокупности по изучаемому варьирующему признаку и показывает типичный для данных условий уровень этого признака. Однако средняя величина не полностью характеризует изучаемую совокупность. Например, средняя величина в двух совокупностях может быть одинакова, но в одном случае все индивидуальные признаки отличаются от нее мало, в другом — эти отличия велики. Иначе говоря, в одном случае вариация признака мала, а в другой — велика.

Пример: Сравнить результаты тестирования трех параллельных студенческих групп. Распределение баллов, полученных студентами, приведено на рис. 7.6.

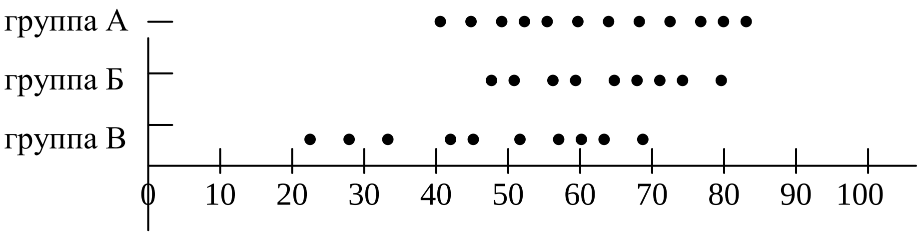


Рис. 7.6. Диаграмма распределения баллов, полученных по результатам тестирования

Сравнивая результаты можно сказать, что студенты группы А в целом набрали более высокие баллы, но разброс баллов достаточно велик (т.е. они сильно варьируются). В группе Б баллы также высоки, но они более компактны по сравнению с оценками группы А. Группа В получила в целом более низкие баллы, чем предыдущие группы и разброс баллов достаточно велик.

Таким образом, различия между группами можно анализировать по двум факторам:

1. среднему баллу;
2. колеблемости (или вариации) баллов.

Для характеристики величины колебаний значений признаков в совокупности рассчитывают абсолютные и относительные показатели вариации.

Абсолютные показатели:

1) размах вариации;
2) среднее линейное отклонение;
3) среднее квадратическое отклонение;
4) дисперсия.

Относительные показатели:

1) коэффициент осцилляции;
2) относительное линейное отклонение;
3) коэффициент вариации.

Абсолютные показатели вариации

Размах вариации (*R*) представляет собой разность между наибольшим ($x_{\max}$) и наименьшим ($x_{\min}$) значениями признака.

$$R = x_{\max} - x_{\min}$$

Пример. Найдите размах следующего ряда значений:

$$2, 8, 7, 4, 2, 1$$

$$R = 8 - 1 = 7.$$

Этот показатель отражает вариацию только крайних значений признака и не учитывает степени колеблемости основной массы единиц совокупности.

Более точными характеристиками вариации признаков являются показатели колеблемости относительно *среднего уровня* признака:

– среднее линейное отклонение;

– среднее квадратическое отклонение;

– дисперсия.

Среднее линейное отклонение и среднее квадратическое отклонение показывают на сколько в среднем отклоняются индивидуальные значения признака от среднего значения. Они выражаются именованными числами.

Среднее линейное отклонение $\left(\bar{d}\right)$ определяется по формулам:

$$\bar{d} = \frac{\sum |x_i - \bar{x}| f_i}{n}$$ (если данные не сгруппированы) или

$$\bar{d} = \frac{\sum |x_i - \bar{x}| f_i}{\sum f_i}$$ (если данные представлены в виде вариационного ряда распределения)

где x_i — конкретные значения признака;

$\bar{x}$ — среднее значение признака (средняя арифметическая);

n — число значений;

f_i — частоты.

С помощью этого показателя можно анализировать, например, состав работающих, ритмичность производства, равномерность поставок материалов и т.п., однако этот показатель затрудняет применение методов математической статистики (его нельзя поставить в соответствие с каким-либо вероятностным законом), поэтому в статистических исследованиях обычно используют среднее квадратическое отклонение.

Среднее квадратическое отклонение (σ) имеет вид:

— для несгруппированных данных:

$$\sigma = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n}}$$

— для вариационного ряда распределения:

$$\sigma = \sqrt{\frac{\sum (x_i - \bar{x})^2 f_i}{\sum f_i}}.$$

Среднее квадратическое отклонение по величине всегда больше линейного отклонения. Соотношение $\sigma / \bar{d}$ зависит от наличия в совокупности резких, выделяющихся отклонений. Чем больше сильных отклонений, тем больше это соотношение. Для нормального закона распределения соотношение $\sigma / \bar{d}$ = 1,2.

Дисперсия (σ^2) — представляет собой квадрат отклонений индивидуальных значений признака от средней арифметической:

$$\sigma^2 = \frac{\sum (x_i - \bar{x})^2}{n} \text{ (для несгруппированных данных)}$$

$$\text{или } \sigma^2 = \frac{\sum (x_i - \bar{x})^2 f_i}{\sum f_i} \text{ (для вариационного ряда).}$$

Пример. Определите среднее линейное отклонение, среднее квадратическое отклонение и дисперсию по исходным данным табл. 7.6.

Таблица 7.6

x_i	f_i	$x_i f_i$	$x_i - \bar{x}$	$\lvert x_i - \bar{x} \rvert f_i$	$(x_i - \bar{x})^2$	$(x_i - \bar{x})^2 f_i$
3	2	6	–7	–14	196	392
11	3	33	1	3	9	27
21	1	21	11	11	121	121
	6	60	—	28	—	540

Средняя арифметическая равна:

$$\bar{x} = \frac{\sum x_i f_i}{\sum f_i} = \frac{60}{6} = 10.$$

Среднее линейное отклонение:

$$\bar{d} = \frac{\sum \lvert x_i - \bar{x} \rvert f_i}{\sum f_i} = \frac{28}{6} = 46{,}7.$$

Среднее квадратическое отклонение:

$$\sigma = \sqrt{\frac{\sum (x_i - \bar{x})^2 f_i}{\sum f_i}} = \sqrt{\frac{540}{6}} = 9{,}49.$$

Дисперсия:

$$\sigma^2 = 90.$$

Относительные показатели вариации

При сравнении колеблемости различных признаков в одной и той же совокупности или при сравнении одного и того же признака разных совокупностей используются **относительные показатели вариации**.

Они рассчитываются как отношение соответствующих абсолютных показателей колеблемости к средней арифметической и обычно выражаются в процентах.

Формулы для расчета относительных показателей вариации следующие:

Коэффициент осцилляции:

$$K_R = \frac{R}{\bar{x}} \cdot 100\%.$$

Относительное линейное отклонение:

$$K_{\bar{d}} = \frac{\bar{d}}{\bar{x}} \cdot 100\%.$$

Коэффициент вариации:

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

Наиболее часто в статистической практике используется коэффициент вариации.

Коэффициент вариации (как и относительное линейное отклонение) показывает, *на сколько процентов в среднем* отклоняются индивидуальные значения признака от средней арифметической.

Коэффициент вариации также используют для характеристики однородности изучаемой совокупности: если коэффициент вариации не превышает 33%, то совокупность считается однородной.

7.4.2. Сложение дисперсий изучаемого признака

Изучая дисперсию интересующего нас признака по всей совокупности, мы не можем определить, насколько сильно влияет на вариацию этого признака тот или иной фактор. Например, производительность труда работников зависит от стажа работы, квалификации, уровня механизации труда, мотивации работников и других факторов.

Дисперсия производительности труда, рассчитанная по всей совокупности, показывает вариацию производительности под действием всех факторов (стажа, квалификации, механизации и др.).

Вариация признака, вызванная влиянием *всех факторов*, называется **общей дисперсией** ($\sigma_о^2$):

$$\sigma_o{}^2 = \frac{\sum (x_i - \bar{x}_o)^2 f_i}{\sum f_i},$$

где $\bar{x}_o$ — общая средняя арифметическая для всей изучаемой совокупности.

Если необходимо выяснить, насколько сильно влияет на вариацию изучаемого признака тот или иной фактор в отдельности, всю совокупность данных группируют по этому фактору. Вариация признака, вызванная фактором, положенным в основу группировки, определяется с помощью **межгрупповой (факторной)** дисперсии (δ^2):

$$\delta^2 = \frac{\sum (\bar{x}_i - \bar{x}_o)^2 n_i}{\sum n_i},$$

где $\bar{x}_i$ — средняя по отдельной группе;

n_i — число единиц в группе.

Вариацию, вызванную влиянием остальных факторов в каждой отдельной группе, характеризует внутригрупповая (остаточная) дисперсия $\left(\sigma_i^2\right)$:

$$\sigma_i^2 = \frac{\sum \left(x - \bar{x}_i\right)^2 f}{\sum f},$$

где f — число единиц в отдельной группе.

По совокупности в целом вариация значений признака под влиянием прочих факторов оценивается с помощью **средней из внутригрупповых дисперсий** $\left(\bar{\sigma}^2\right)$**:**

$$\bar{\sigma}^2 = \frac{\sum \sigma_i^2 n_i}{\sum n_i}.$$

Между общей дисперсией σ_o^2, средней из внутригрупповых дисперсий $\bar{\sigma}^2$ и межгрупповой дисперсией δ^2 существует соотношение, которое называется **правилом сложения дисперсий**. Согласно этому правилу, **общая дисперсия равна сумме средней из внутригрупповых и межгрупповой дисперсий**:

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

Эмпирическое корреляционное отношение (η), рассчитанное на базе правила сложения дисперсий, **показывает, какая часть (доля) общей дисперсии складывается под влиянием признака, положенного в основу группировки**.

$$\eta = \sqrt{\frac{\delta^2}{\sigma_o^{\ 2}}}.$$

Этот показатель лежит в интервале [0, 1]. Для качественной оценки силы связи между изучаемым (результативным) признаком и признаком, положенным в основу группировки (факторным) можно использовать следующую таблицу 7.7.

Таблица 7.7

Значения η	Оценка силы связи
0—0,3 (+)	слабая
0,3—0,7	средняя
0,7—1	тесная

Пример. Используя данные табл. 7.8 исследовать зависимость между численностью работающих (признак-фактор) и объемом выполненных работ (результативный признак).

Таблица 7.8

Объем выполненных работ проектно-изыскательными организациями

№ предприятия	Число работающих, x_i, чел.	Объем выполненных работ, y_i, млн.руб.
1	20	9,61
2	79	50,11
3	39	17,32
4	62	28,77
5	47	19,34
6	20	15,12
7	31	19,34
8	39	16,05
9	39	18,34
10	64	31,22
11	36	19,00
12	51	29,12

Произведем группировку организаций по числу работающих, выделив четыре группы. Результаты группировки представлены в табл. 7.9.

Таблица 7.9

Группировка проектно-изыскательских организаций

по числу работающих

№ группы	Численность работающих, чел.	Количество организаций, ед.	Объем выполненных работ, млн.руб.
1	20—35 (+)	3	9,61; 15,12; 19,34
2	35—50	5	17,32; 19,34; 16,05; 18,34; 19,00
3	50—65	3	28,77; 31,22; 29,12
4	65—80	2	50,11
Итого		**12**	

1. Рассчитаем общую дисперсию σ_o^2 (по данным табл. 7.8):

$$\bar{x} = \frac{9{,}61 + 50{,}11 + 17{,}32 + \ldots + 19{,}0 + 29{,}12}{12} = 22{,}78 \text{ млн.руб.}$$

$$\sigma_o{}^2 = \frac{(9{,}61 - 22{,}78)^2 + (50{,}11 - 22{,}78)^2 + (17{,}32 - 22{,}78)^2 + \ldots + (29{,}12 - 22{,}78)^2}{12} =$$
$$= 104{,}93$$

2. Рассчитаем групповые (частные) средние (табл. 7.9).

$$\bar{x}_1 = \frac{9{,}61 + 15{,}12 + 19{,}34}{3} = 14{,}69 \text{ млн.руб.}$$

$$\bar{x}_2 = \frac{17{,}32 + 19{,}34 + 16{,}05 + 18{,}34 + 19{,}00}{5} = 18{,}01 \text{ млн.руб.}$$

$$\bar{x}_3 = \frac{28{,}77 + 31{,}22 + 29{,}12}{3} = 29{,}70 \text{ млн.руб.}$$

$$\bar{x}_4 = 50{,}11 \text{ млн. руб.}$$

3. Определим межгрупповую дисперсию:

$$\delta^2 = \frac{(14{,}69 - 22{,}78)^2 \cdot 3 + (18{,}01 - 22{,}78)^2 \cdot 5 + (29{,}70 - 22{,}78)^2 \cdot 3 + (50{,}11 - 22{,}78)^2 \cdot 1}{12} =$$

4.

$$= 100{,}07$$

Рассчитаем внутригрупповые (остаточные) дисперсии:

$$\sigma_1^2 = \frac{(9{,}61 - 14{,}69)^2 + (15{,}12 - 14{,}69)^2 + (19{,}34 - 14{,}69)^2}{3} = 15{,}87$$

$$\sigma_2^2 = \frac{(17{,}32 - 18{,}01)^2 + (19{,}34 - 18{,}01)^2 + (16{,}05 - 18{,}01)^2 + (18{,}34 - 18{,}01)^2}{5} +$$
$$+ \frac{(19{,}00 - 18{,}01)^2}{5} = 1{,}44$$

$$\sigma_3^2 = \frac{(28{,}77 - 29{,}70)^2 + (31{,}22 - 29{,}70)^2 + (29{,}12 - 29{,}70)^2}{3} = 1{,}17$$

$$\sigma_4^2 = 0.$$

5. Вычислим среднюю из внутригрупповых дисперсий:

$$\bar{\sigma}^2 = \frac{15{,}87 \cdot 3 + 1{,}44 \cdot 5 + 1{,}17 \cdot 3 + 0 \cdot 1}{12} = 4{,}86.$$

6. Проверяем расчеты, определяя общую дисперсию по правилу сложения дисперсий:

100,07 + 4,86 = 104,93 — верно (см. п. 1).

7. Рассчитываем эмпирическое корреляционное отношение:

$$\eta = \sqrt{\frac{100{,}07}{104{,}93}} = 0{,}98.$$

Полученная величина свидетельствует о том, что фактор, положенный в основу группировки (то есть численность работающих), сильно влияет на объем работ, выполняемых организациями.

7.5. *Показатели формы распределения*

При анализе рядов распределения большое практическое значение имеет изучение формы распределения. Другими словами, важно знать, существует ли какая-то закономерность в распределении частот с увеличением (или уменьшением) значений признака или нет.

Приблизительное представление о форме распределения дают графики распределения — полигон и гистограмма.

Для более точной характеристики особенностей формы распределения используют:

1. ранговые характеристики;
2. показатели дифференциации;
3. кривые распределения, показатели асимметрии и эксцесса распределения.

7.5.1. Ранговые характеристики

Ранговые характеристики — это варианты, занимающие в ранжированном ряду распределения определенное место. К ним относят: квартили (Q); децили (D) и перцентили (P).

Квартили — это значения признака, которые делят ранжированный (упорядоченный) ряд на четыре равные по численности части.

Таких квартилей будет три: первая (нижняя) квартиль (Q_1), вторая квартиль (она же медиана) — Q_2 и (верхняя) третья квартиль Q_3.

Вычисление квартилей аналогично вычислению медианы.

В **дискретном ряду распределения** порядок расчета квартилей следующий:

1. Определяют место квартили (N_Q) в ряду распределения:

$$N_{Q1} = \frac{1}{4}(n+1); \quad N_{Q2} = \frac{2}{4}(n+1); \quad N_{Q3} = \frac{3}{4}(n+1),$$

где n — общее число единиц в совокупности.

2. Рассчитывают численные значения квартилей.

Если данные сгруппированы, то значения находят по накопленным частотам.

Пример 1. Определите первую и третью квартили по 9 значениям:

2, 4, 7, 3, 8, 9, 12, 9, 6

Прежде всего, необходимо упорядочить значения:

2, 3, 4, 6, 7, 8, 9, 9, 12

Нижняя квартиль (Q_1) это $\frac{1}{4}(9+1) = 2{,}5$-е значение. Поэтому мы используем среднюю арифметическую из 2-го и 3-го значений, которые равны 3 и 4. Таким образом $Q_1 = \frac{3+4}{2} = 3{,}5$.

Верхняя квартиль (Q_3), это $\frac{3}{4}(9+1) = 7{,}5$-е значение. Используем среднюю арифметическую из 7-го и 8-го значений, которые равны 9 и 9. Таким образом $Q_4 = 9$.

Вторая квартиль соответствует медиане: это $\frac{1}{2}(9+1)=5$-е по счету значение в ряду распределения, ему соответствует варианта со значением 7. $Q_2 = M_e = 7$.

Пример 2: Определите первую и третью квартили по 8 значениям:

12, 5, 4, 23, 17, 32, 27, 15

Прежде всего, необходимо упорядочить значения:

4, 5, 12, 15, 17, 23, 27, 32

Нижняя квартиль (Q_1) это $\frac{1}{4}(8+1)=2{,}25$-е значение. Следовательно, надо определить значение, которое находится на расстоянии 0,25 от 2-го значения до 3-го значения.

2-е значение равно 5.

3-е значение равно 12.

Расстояние от 5 до 12 равно 7.

Таким образом, нижняя квартиль Q_1 равна $5 + 0{,}25 \cdot 7 = 6{,}75$.

Верхняя квартиль — это ¾ (8 + 1) = 6,75-е значение. Надо определить значение, которое находится на расстоянии 0,75 от 6-го значения до 7-го значения.

6-е значение равно 23.

7-е значение равно 27.

Расстояние от 23 до 27 равно 4. Таким образом, верхняя квартиль Q_3 равна $23 + 0{,}75 \cdot 4 = 26$.

Расчет квартилей в интервальном ряду распределения также аналогичен расчету медианы в интервальном ряду:

1. Определяется место квартили (N_Q) в ряду:

$$N_{Q1} = \frac{1}{4}(n+1); \quad N_{Q3} = \frac{3}{4}(n+1),.$$

2. По накопленным частотам определяются квартильные интервалы.

3. Рассчитываются квартили по следующим формулам:

$$Q_1 = x_{Q1} + h\frac{\frac{1}{4}\sum f - f^H_{Q1-1}}{f_{Q1}}; \qquad Q_3 = x_{Q3} + h\frac{\frac{3}{4}\sum f - f^H_{Q3-1}}{f_{Q3}}; \quad ,$$

где x_{Q1}, x_{Q3} — нижние границы соответствующих квартильных интервалов;

h — ширина интервала;

Σf — сумма всех частот;

f^{H}_{Q1-1}, f^{H}_{Q3-1} — накопленные частоты, предшествующие соответствующим квартильным интервалам;

f_{Q1}, f_{Q3} — частота квартильных интервалов.

При расчете квартилей можно пользоваться частостями.

Пример. Вычислить первую и третью квартили по данным табл. 7.10.

Таблица 7.10

Интервал	f_i	F_i^H
3,7—4,6 (–)	2	2
4,6—5,5	4	6
5,5—6,4	6	12
6,4—7,3	5	17
7,3—8,2	3	20
Итого	**20**	—

Сначала определим квартильные интервалы, предварительно рассчитав накопленные частоты.

Первая квартиль будет находиться в интервале, накопленная частота которого равна или больше ¼ суммы частот (т.е. ¼ от 20-ти). Это будет интервал 4,6—5,5. Тогда:

$$Q_1 = 4{,}6 + 0{,}9\frac{\frac{1}{4}\cdot 20 - 2}{4} = 5{,}27.$$

Третья квартиль будет находиться в интервале, накопленная частота которого равна или больше ¾ суммы частот (т.е. ¾ от 20-ти). Это интервал 6,4—7,3. Значит:

$$Q_3 = 6{,}4 + 0{,}9\frac{\frac{3}{4}\cdot 20 - 12}{5} = 6{,}94.$$

Децили (*D*) — это значения признака, которые делят ранжированный ряд распределения на десять равных по численности частей.

Расчет децилей в дискретном и интервальном рядах осуществляется аналогично расчету квартилей.

При расчете децилей сначала определяют места девяти децилей (N_D):

$$N_{D1} = \frac{1}{10}(n+1); \quad N_{D2} = \frac{2}{10}(n+1);\ldots \quad N_{D9} = \frac{9}{10}(n+1).$$

В дискретном ряду численные значения децилей определяют по накопленным частотам.

В интервальном ряду сначала определяют децильный интервал (по накопленным частотам), затем находят численные значения децилей по формулам:

$$D_1 = x_{D1} + h\frac{\frac{1}{10}\sum f - f^H_{D1-1}}{f_{D1}};$$

$$D_2 = x_{D2} + h\frac{\frac{2}{10}\sum f - f^H_{D2-1}}{f_{D2}};$$

и т.д.,

где x_{D1}, x_{D2} — нижние границы децильных интервалов;

h — ширина интервала;

f^H_{D1-1}, f^H_{D2-1} — накопленные частоты, предшествующих децильным;

f_{D1}, f_{D2} — частота децильных интервалов.

Перцентили (*P*) — варианты, которые делят ранжированный ряд распределения на сто равных по численности частей. Определяются перцентили так же, как и квартили.

7.5.2. Показатели дифференциации

К показателям дифференциации относятся коэффициент фондовой дифференциации и коэффициент децильной дифференциации.

Коэффициент фондовой дифференциации ($K_Ф$) показывает, во сколько раз средний уровень 10% единиц совокупности, имеющих наибольшие значения, больше среднего уровня 10% единиц, имеющих наименьшие значения:

$$K_{Ф} = \frac{\bar{x}_{наиб}}{\bar{x}_{наим}},$$

где $\bar{x}_{наиб}$ — средний уровень признака из 10% наибольших значений признака;

$\bar{x}_{наим}$ — средний уровень признака из 10% наименьших значений признака.

Этот показатель можно рассчитывать по первичным данным.

В ряду распределения определяется **коэффициент децильной дифференциации (K_D)** по формуле:

$$K_D = \frac{D_9}{D_1},$$

где D_9, D_1 — соответственно девятая и первая дециль.

Этот коэффициент показывает, во сколько раз минимальный уровень признака у 10% единиц, имеющих наибольшие значения, больше максимального уровня признака у 10% единиц совокупности, имеющих наименьшие значения признака.

Показатели дифференциации широко используются, например, для оценки степени дифференциации доходов населения.

В 2006 г. наиболее высокий разрыв в начисленной средней заработной плате отмечался в оптовой торговле, финансовой деятельности и сельском хозяйстве. Фондовый коэффициент дифференциации составил в этих видах экономической деятельности, соответственно, 32,7; 31,4 и 21,3.

Самый низкий коэффициент был зафиксирован в отраслях, связанных с производством и распределением электрической энергии, газа и воды — 11,5. На транспорте коэффициент фондовой дифференциации равен 17,6.

7.5.3. Кривые распределения, показатели асимметрии и эксцесса

Рассмотренные в п. 7.2 графические способы изображения вариационного ряда — полигон и гистограмма дают представление о фактической закономерности распределения единиц совокупности по величине варьирующего признака, в котором отражаются как общие, так и случайные условия, определяющие распределение. Эти кривые называются **эмпирическими кривыми распределения**.

Если бы мы имели возможность увеличивать число наблюдений (в пределах одной и той же совокупности) при одновременном уменьшении величины интервала между значениями признака, закономерность характерная для данного распределения, стала бы проявляться более четко, а ломаная линия, представляющая полигон частот (или гистограмму), приближалась бы к некоторой плавной линии, в пределе превращающейся в кривую линию. Эта кривая, отражающая закономерность изменения частот в ряду распределения, в чистом, «идеальном» виде, называется **теоретической кривой распределения**.

В практике статистических исследований приходится иметь дело с различными теоретическими видами распределений. Одним из наиболее распространенных является, так называемое, **нормальное распределение**. При **нормальном распределении** существует следующая взаимосвязь между изменением частот и значений варьирующего признака: **с увеличением значений варьирующего признака частоты сначала увеличиваются, а затем уменьшаются**.

Общий вид кривой нормального распределения представлен на рис. 7.7.

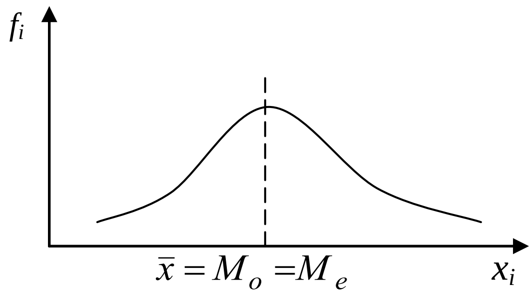


Рис. 7.7. Кривая нормального распределения

Кривая нормального распределения обладает рядом свойств, которые используются в экономической практике при анализе рядов распределения:

1. Кривая симметрична относительно максимальной ординаты, которая соответствует значению $\bar{x}$. При нормальном распределении значения средней арифметической, моды и медианы совпадают: $\bar{x} = M_o = M_e$.

2. Кривая асимптотически приближается к оси абсцисс, продолжаясь в обе стороны до бесконечности. То есть, чем больше значения отклоняются от $\bar{x}$, тем реже они встречаются.
3. Кривая имеет две точки перегиба, находящиеся на расстоянии $\pm\sigma$ от $\bar{x}$.
4. Для кривой нормального распределения действует **правило трех сигм**: при нормальном распределении:

68,3% всех значений признака совокупности находятся в пределах от $\bar{x}-\sigma$ до $\bar{x}+\sigma$;

95,4% всех значений признака в совокупности попадают в интервал от $\bar{x}-2\sigma$ до $\bar{x}+2\sigma$;

99,7% всех значений признака в совокупности лежат в пределах от $\bar{x}-3\sigma$ до $\bar{x}+3\sigma$.

Нормальное распределение возможно лишь в тех случаях, когда на величину признака влияет большое число случайных величин.

Для характеристики формы фактического ряда распределения используются показатели асимметрии и эксцесса.

Асимметрия присуща многим распределениям. Она характеризует «скошенность» распределения и возникает тогда, когда случайные влияния действуют в одном направлении сильнее, чем в другом.

Если у кривой распределения левая ветвь больше правой — это **левосторонняя асимметрия** и для нее характерно следующее соотношение: $M_o < M_e < \bar{x}$; если правая ветвь больше, то речь идет о правосторонней асимметрии: $M_o > M_e > \bar{x}$ (рис. 7.8).

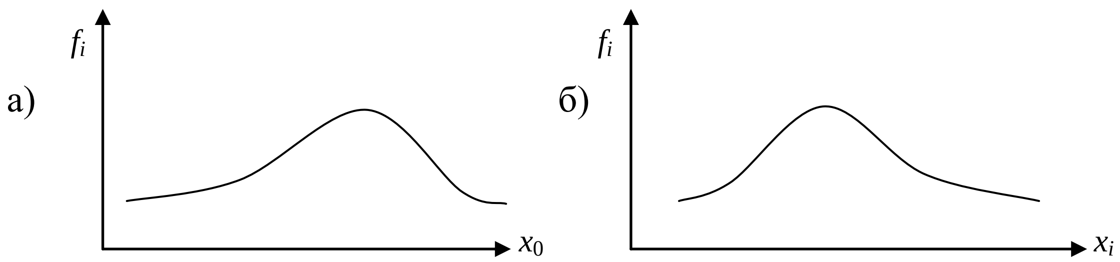


Рис. 7.8. Асимметричные ряды распределения

а) с левосторонней асимметрией; б) с правосторонней асимметрией

Для сравнительного анализа степени асимметрии у нескольких распределений рассчитывают **коэффициент асимметрии (A_S)**:

$$A_S = \frac{\mu_3}{\sigma_3},$$

где μ_3 — центральный момент третьего порядка:

$$\mu_3 = \frac{\sum (x_i - \bar{x})^3 f_i}{\sum f_i}.$$

Если $A_S > 0$, то асимметрия правосторонняя, если $A_S < 0$ — левосторонняя. Принято считать, что если $|A_S| > 0{,}5$, то асимметрия значительная; если $0{,}25 \le |A_S| \le 0{,}5$ — условно незначительна; если $|A_S| < 0{,}25$ — асимметрия незначительная и распределение можно считать симметричным.

Для симметричных распределений рассчитывается показатель эксцесса. **Эксцесс (E_S)** характеризует островершинность или плосковершинность распределения:

$$E_S = \frac{m_4}{\sigma_4} - 3,$$

где m_4 — центральный момент четвертого порядка:

$$m_4 = \frac{\sum (x_i - \bar{x})^4 f_i}{\sum f_i}.$$

У островершинных кривых $E_S > 0$; у плосковершинных кривых $E_S < 0$.

Вопросы для самоконтроля

1. Что такое ряд распределения?
2. Приведите пример атрибутивного ряда распределения; вариационного ряда.
3. Назовите основные элементы ряда распределения.
4. Назовите способы графического изображения рядов распределения.
5. Для чего используют кривую Лоренца? Как ее строят?
6. Назовите абсолютные показатели вариации. Их назначение?
7. Назовите относительные показатели вариации. Их экономическое толкование?
8. В чем заключается правило сложения дисперсий?
9. Что характеризует межгрупповая дисперсия?
10. Напишите формулу для расчета средней из внутригрупповых дисперсий.
11. Что показывает эмпирическое корреляционное отношение?
12. Что такое квартили? Децили? Перцентили?
13. Что показывает коэффициент фондовой дифференциации?
14. Что показывает коэффициент децильной дифференциации?
15. Каковы особенности кривых нормального распределения?
16. В чем заключается правило трех сигм?
17. Что такое асимметрия распределения?
18. Что такое эксцесс?

8. Выборочное наблюдение

Выборочное наблюдение является наиболее совершенным и широко применяемым способом несплошного статистического наблюдения.

При проведении несплошного наблюдения нельзя получить абсолютно точные данные, так как обследованию подвергается не вся *генеральная совокупность*, а только ее часть — *выборочная совокупность.* Однако, в отличие от других способов отбора (основного массива и монографического), выборочный метод позволяет математически оценить погрешность (*ошибку*) выборки и рассчитать пределы возможной ошибки выборочного наблюдения.

Ошибки, свойственные выборочному наблюдению называют *ошибками репрезентативности или представительства.*

8.1. Виды и схемы отбора

Размер ошибки выборки и методы ее определения зависят от вида и схемы отбора.

Различают четыре вида отбора единиц совокупности:

1. Случайный
2. Механический
3. Типический
4. Серийный (гнездовой).

При **собственно случайном способе** отбор единиц из генеральной совокупности производится по таблице случайных чисел.

Примером случайного отбора могут служить тиражи выигрышей, когда из общего количества выпущенных билетов в случайном порядке наугад выбирается определенная часть номеров, на которые приходится выигрыши.

Механический отбор заключается в том, что единицы изучаемого явления предварительно располагаются в определенном порядке по какому-либо принципу (по алфавиту, местоположению и т.п.) и затем механически, через определенный интервал, проводится отбор.

Например, если надо провести 25%-ную выборку студентов, то составляется список их фамилий по алфавиту и механически отбирается каждый четвертый студент.

Типический отбор используется в тех случаях, когда изучаемая совокупность неоднородна. Сначала единицы генеральной совокупности распределяются на однотипные группы по какому-либо признаку, существенно отличающему одну группу от другой. Затем внутри каждой группы одним из способов (собственно случайным или механическим) проводится отбор необходимого числа единиц для последующего изучения. Достоинство типологического способа отбора заключается в том, что он обеспечивает пропорциональное попадание в выборочную совокупность единиц из всех групп.

Например, необходимо провести типический отбор 1000 студентов из 10000, обучающихся на пяти факультетах института. Для этого их сначала группируют в однородные группы по факультетам, а затем из каждой группы отбирают случайным образом столько студентов, чтобы их численность была пропорциональна удельному весу численности каждого факультета.

При **серийном (гнездовом) отборе** выборочная совокупность образуется путем отбора целых серий (гнезд), а не индивидуальных единиц совокупности. В отобранных сериях все единицы совокупности подвергаются сплошному наблюдению, а сам отбор серий может осуществляться путем механического или случайного отбора.

Серийная выборка широко применяется в торговле, поскольку многие товары для их транспортировки, хранения и продажи упаковываются в пачки, коробки, ящики. При контроле качества товара рациональнее проверить несколько отдельных упаковок, чем из всех упаковок отбирать единицы.

Основной формой выборочного наблюдения является простая случайная выборка. Другие формы организации выборочного наблюдения представляют ее дальнейшее развитие и видоизменение в целях повышения репрезентативности наблюдения.

Точность выборки зависит и от схемы отбора. Выборка может быть проведена по схеме **возвратной (повторной)** выборки и в форме **безвозвратной (бесповторной)** выборки.

В первом случае встретившийся номер не исключается из списка и может быть выбран повторно. Во втором случае выбранные номера вычеркиваются из списка, и каждая данная единица может быть включена в выборочную совокупность только один раз. Бесповторный отбор дает более точные результаты.

При выборке из конечной совокупности на практике применяется способ бесповторной выборки.

8.2. Средняя и предельная ошибки выборки

При проведении выборочного наблюдения основной задачей является определение характеристик генеральной совокупности. Выборочные средние (или другие выборочные характеристики), как правило, не могут точно совпадать с искомыми характеристиками генеральной совокупности, а будут отклоняться от последних в ту или иную сторону. Отклонение значения показателя выборочной совокупности от размеров его во всей изучаемой совокупности представляет собой **ошибку репрезентативности (ошибку выборки)**.

При соблюдении принципа случайного отбора ошибка выборки определяется прежде всего численностью выборки. Чем больше численность выборки (при прочих равных условиях), тем меньше величина ошибки выборки.

Естественно, что если численность выборки довести до численности генеральной совокупности, то обследование становится сплошным и вопрос об ошибке выборки отпадает.

Для определения отклонения характеристик выборочной совокупности от генеральной совокупности применяются следующие показатели:

1. Средняя ошибка выборки. Она показывает, в каких пределах в ту или иную сторону можно ожидать отклонения среднего значения признака всей совокупности от среднего признака, рассчитанного по выборочной совокупности.

Расчет средней ошибки *простой случайной выборки* проводится по следующим формулам:

- при повторном способе отбора единиц:

$$\mu_{\bar{x}} = \sqrt{\frac{S^2}{n}},$$

где S^2 — дисперсия варьирующего признака, рассчитанная по выборочной совокупности (выборочная дисперсия);

n — численность единиц выборочной совокупности.

- при бесповторном способе отбора единиц:

$$\mu_{\bar{x}} = \sqrt{\frac{S^2}{n}(1-\frac{n}{N})},$$

где N – численность единиц генеральной совокупности.

2. Предельная (максимально возможная) ошибка выборки — это такая величина отклонения выборочной средней от генеральной, вероятность превышения которой, вследствие случайных причин в условиях данной выборки, очень мала.

Предельная ошибка *простой случайной выборки* определяется по формулам:

- при повторном способе отбора единиц:

$$\Delta_{\bar{x}} = t\mu_{\bar{x}} = t\sqrt{\frac{S^2}{n}},$$

где t — коэффициент кратности ошибки (коэффициент доверия), зависящий от вероятности, с которой можно гарантировать, что предельная ошибка выборки не превысит t-кратную среднюю ошибку.

- при бесповторном способе отбора единиц:

$$\Delta_{\bar{x}} = t\mu_{\bar{x}} = t\sqrt{\frac{S^2}{n}(1-\frac{n}{N})}$$

Предельная ошибка выборки $\Delta_{\bar{x}}$ может быть в каждом конкретном случае меньше μ, равна ей или больше ее. вероятность той или иной величины ее при достаточно большом n определяется как функция от t, т.е. $p = f$(t).

Существуют таблицы, определяющие зависимость вероятности, с которой гарантируется результат (p) от коэффициента доверия t. Ниже приведена краткая выдержка из этой таблицы (табл. 8.1).

Таблица 8.1

Распределение вероятности (p) в зависимости от коэффициента доверия t

T	Вероятность	*T*	Вероятность	*t*	Вероятность
1,0	0,683	1,7	0,911	2,4	0,984
1,1	0,729	1,8	0,928	2,5	0,988
1,2	0,77	1,9	0,943	2,6	0,991
1,3	0,806	2,0	0,954	2,7	0,993
1,4	0,838	2,1	0,964	2,8	0,995
1,5	0,866	2,2	0,972	2,9	0,996
1,6	0,890	2,3	0,979	3,0	0,997

Легко определить по таблице, что при $t = 1$ $p = f(1) = 0{,}683$.

Это значит, что в 683 случаях их 1000 характеристика генеральной совокупности будет отличаться от характеристики выборочной совокупности не больше, чем на величину μ, но в остальных 317 случаях она может отличаться и в большей степени.

Формулы для определения ошибки простой случайной выборки дают возможность не только определить эти ошибки, но и рассчитать предварительно, какую необходимо взять численность выборки, чтобы ошибка выборки не превышала определенные заданные размеры. При этом могут решаться две задачи.

1. **Определение пределов генеральных характеристик** с заданной степенью надежности (доверительной вероятностью) на основе показателей, полученных по данным выборки.

Доверительные интервалы для генеральной средней –

$$\bar{x} = \tilde{x} \pm \Delta_{\bar{x}};$$

$$\tilde{x} - \Delta_{\bar{x}} \leq \bar{x} \leq \tilde{x} + \Delta_{\bar{x}},$$

где $\tilde{x}$ – среднее значение признака в выборочной совокупности (выборочная средняя);

$\bar{x}$ – среднее значение признака в генеральной совокупности (генеральная средняя).

2. **Определение необходимого объема выборки**, который с заданной вероятностью обеспечивает заданную точность выборки:

- при повторном способе отбора единиц:

$$n = \frac{t^2 S^2}{\Delta_{\bar{x}}^2};$$

- при бесповторном способе отбора единиц:

$$n = \frac{t^2 N S^2}{\Delta_{\bar{x}}^2 N + t^2 S^2}.$$

Пример: Результаты простого случайного повторного наблюдения показали, что стоянки судов в порту по причине ремонта составили следующее количество часов (табл. 8.2).

Таблица 8.2

Ряд распределения судов по продолжительности стоянок в порту

Продолжительность стоянки, час.	3	4	5	6	7	8	9	10	11	13
Количество случаев	4	6	7	6	4	3	4	3	2	1

Определите:

1. Среднюю продолжительность стоянки судов, гарантируя результат с вероятностью 0,954 (95,4%).

2. Необходимую численность выборки при определении средней продолжительности стоянки, чтобы с вероятностью 0,99 (99%) предельная ошибка выборки не превышала 0,5 ч.

Решение

1. В таблице 8.3. приведены вспомогательные расчеты для определения средней продолжительности стоянки и дисперсии.

Таблица 8.3

Вспомогательная таблица для расчета

средней продолжительности стоянки и дисперсии

x_i	f_i	$x_i f_i$	$x_i - \bar{x}$	$(x_i - \bar{x})^2$	$(x_i - \bar{x})^2 f_i$
3	4	12	-3,5	12,25	49,00
4	6	24	-2,5	6,25	37,50
5	7	35	-1,5	2,25	15,75
6	6	36	-0,5	0,25	1,50
7	4	28	0,5	0,25	1,00
8	3	24	1,5	2,25	6,75
9	4	36	2,5	6,25	25,00
10	3	30	3,5	12,25	36,75
11	2	22	4,5	20,25	40,50
13	1	13	6,5	42,25	42,25
	40	**260**			**256,0**

$$\tilde{x} = 6{,}5 \text{ час.} \qquad S^2 = \frac{256}{40} = 6{,}4.$$

Средняя ошибка выборки составит:

$$\mu_{\bar{x}} = \sqrt{\frac{6{,}4}{40}} = \sqrt{0{,}16} = 0{,}4 \text{ ч.}$$

Найдем предельную ошибку выборки с вероятностью 0,954, то есть при $t = 2$. $\Delta_{\bar{x}} = \pm t\mu_{\bar{x}} = \pm 2 \cdot 0{,}4 = \pm 0{,}8$.

Выборочное обследование позволяет сделать следующий вывод: с вероятностью 0,954 можно утверждать, что во всей совокупности стоянок по причине ремонта средняя продолжительность стоянки не больше, чем

$\tilde{x} + \Delta_{\bar{x}} = 6{,}5 + 0{,}8 = 7{,}3$ ч. и не меньше, чем $\tilde{x} - \Delta_{\bar{x}} = 6{,}5 — 0{,}8 = 5{,}7$ ч. , что можно записать неравенством:

$$5{,}7 \text{ ч.} \le \bar{x} \le 7{,}8 \text{ ч.}$$

2. Каков должен быть объем выборки *n*, чтобы с вероятностью 0,99 предельная ошибка выборки не превышала 0,5 ч.?

При $S^2 = 6{,}4; \quad t = 2{,}6; \quad \Delta_{\bar{x}} = 0{,}5$:

$$n=\frac{t^2S^2}{\Delta_{\bar{x}}^{2}}=\frac{6,76\cdot 6,4}{0,25}=173.$$

Вопросы для самоконтроля

1. Какое наблюдение называют выборочным?
2. Что означает понятие выборочной совокупности?
3. Почему при выборочном наблюдении неизбежны ошибки? Как они называются?
4. Как производятся собственно случайный, механический, типический и серийный отборы?
5. В чем различие повторной и безповторной выборки?
6. Как определяется средняя ошибка выборки?
7. Как определяется предельная ошибка выборки? Что она характеризует?
8. Что характеризует предельная ошибка выборки и как она рас-считывается?

9. Корреляционная связь и ее статистическое изучение

9.1. *Понятие корреляции*

В ряде наук существующие связи между различными явлениями познаются с помощью проведения соответствующих экспериментов.

Однако в общественных науках, в частности экономических, довольно часто заключение о связях вынуждены делать на основе статистических данных, поскольку постановка необходимых экспериментов является очень дорогой или практически нецелесообразной.

В таких случаях необходимо использовать корреляционный анализ, который позволяет прийти к тем же выводам, которые могли быть получены путем эксперимента.

При изучении взаимосвязей между различными признаками выделяют два вида признаков: факторные и результативные.

Факторные признаки (*x*) — это признаки, которые вызывают изменение других признаков.

Результативные признаки (*y*) — это признаки, которые испытывают на себе влияние факторных признаков, то есть меняются под их действием.

В различных ситуациях один и тот же признак может быть либо факторным, либо результативным. Например, при изучении зависимости между производительностью труда работников и уровнем их энерговооруженности, уровень производительности труда будет результативным признаком, а энерговооруженность — факторным признаком. В другом случае, при анализе производительности труда и себестоимости продукции, производительность труда будет уже факторным признаком, а себестоимость продукции — результативным.

Различают две категории связей между факторными и результативными признаками: 1) функциональную и 2) корреляционную.

При **функциональной связи** существует однозначная зависимость результативного признака y от факторного признака x, то есть каждому значению, которое может иметь переменная x, соответствует одно определенное значение переменной y. Такую зависимость обычно обозначают $y = f(x)$.

Например, площадь круга S является функцией его радиуса R:

$$S = \pi R^2.$$

Однако, функциональные связи имеют ограниченное распространение. Они обычно встречаются в точных науках — математике, физике, химии.

Что же касается экономических и общественных явлений, то здесь, в подавляющем большинстве случаев, каждому отдельному значению факторного признака x отвечает определенное множество значений результативного признака y. Например, производительность труда в большой степени зависит от энерговооруженности сотрудников и уровня организации производства, однако на результативный показатель значительное влияние оказывают также мотивация работников, уровень квалификации, спрос на продукцию и другие факторы, причем в каждом конкретном случае факторные признаки могут изменять силу и направленность своего воздействия. Поэтому такая связь не может быть функциональной.

Связь, при которой числовому значению факторного признака x соответствует не конкретная величина, а некоторый ряд вероятных значений результативного признака y, называется **корреляционной связью**. То есть, **при корреляционной связи устанавливается лишь общая тенденция** изменения результативного показателя в зависимости от изменения величины факторного признака. **Корреляционные связи проявляются в среднем, в массе случаев**.

По направлению различают связи **прямые**, когда результативный признак растет с увеличением факторного признака и **обратные**, при которых увеличению факторного признака соответствует снижение результативного признака.

Относительно **аналитической формы** связи бывают **линейными** и **нелинейными**.

С точки зрения **количества взаимодействующих факторов** различают **парную корреляцию**, при которой условно считают, что на изменение результативного признака y влияет, главным образом, только один факторный признак x, и **множественную корреляцию**, когда изучается влияние нескольких факторных признаков на результативный.

Основными задачами, которые решаются при корреляционном анализе, являются:

1) количественное измерение степени тесноты связи двух или более явлений и отбор тех факторов, которые оказывают наибольшее влияние на исследуемый показатель.

Этот этап называют корреляционным анализом;

2) определение математической модели (формы) изучаемой зависимости. Подобрав корреляционное уравнение, показывающее, как изменяется результативный показатель с изменением факторного признака (или признаков), можно использовать его для прогнозирования неизвестных значений.

Этот этап анализа часто называют регрессионным анализом, а построенное уравнение — уравнением регрессии.

9.2. *Парная корреляция*

9.2.1. Определение степени тесноты связи в случае линейной зависимости между признаками

Для определения степени тесноты связи при парной линейной зависимости между признаками используется **линейный коэффициент корреляции (*r*)**.

Наиболее удобной формулой для его расчета является следующая:

$$r = \frac{n\sum xy - \sum x \sum y}{\sqrt{\left[n\sum x^2 - \left(\sum x\right)^2\right] \times \left[n\sum y^2 - \left(\sum y\right)^2\right]}} \qquad (9.1)$$

Линейный коэффициент корреляции может принимать значения от –1 до +1. Чем ближе он по абсолютной величине к 1, тем теснее связь между признаками. При $r = |1|$ — связь функциональная. Если коэффициент корреляции равен нулю — связь отсутствует. Знак коэффициента указывает направление связи: знак «+» соответствует прямой зависимости; знак «–» — обратной зависимости.

В практике используют различные пороги значений коэффициента корреляции. Например, для качественной оценки силы связи можно использовать следующую шкалу (табл. 9.1):

Таблица 9.1

Теснота связи	Величина коэффициента корреляции при наличии	
	прямой связи	обратной связи
Слабая	0—0,3 (+)	0—(–0,3) +
Средняя	0,3—0,7	(–0,3)—(–0,7)
Сильная	0,7—1,0	(–0,7)—(–1,0)

Обычно считается, что при $|r| > 0{,}7$ установленную зависимость целесообразно использовать в анализе, планировании, прогнозировании и в решении других практических вопросов.

Пример. По данным табл. 9.2. вычислить линейный коэффициент корреляции, предполагая линейную зависимость между факторным признаком x и результативным признаком y.

Таблица 9.2

Исходные данные

№ п/п	Численность работающих, x_i, чел.	Объем выполненных работ, y_i, млн.руб.
1	20	9,61
2	79	50,11
3	39	17,32
4	62	28,77
5	47	19,34
6	20	15,12
7	31	19,34
8	39	16,05
9	39	18,34
10	64	31,22
11	36	19,0
12	51	29,12
Всего	**527**	**273,34**

Для расчета коэффициента корреляции построим вспомогательную таблицу (табл. 9.3).

Таблица 9.3

Вспомогательная таблица для расчета коэффициента корреляции

№ п/п	Численность работающих, x_i, чел.	Объем выполненных работ, y_i, млн.руб.	x_iy_i	x_i^2	y_i^2
1	20	9,61	192,2	400	92,35
2	79	50,11	3958,69	6241	2511,01
3	39	17,32	675,48	1521	299,98
4	62	28,77	1783,74	3844	827,71
5	47	19,34	908,98	2209	374,04
6	20	15,12	302,4	400	228,61
7	31	19,34	599,54	961	374,04
8	39	16,05	625,95	1521	257,60
9	39	18,34	815,26	1251	336,36
10	64	31,22	1998,08	4096	974,69
11	36	19	684	1296	361,00
12	51	29,12	1485,12	2601	847,97
Всего	**527**	**273,34**	**13929,44**	**26611**	**7485,37**

Рассчитаем линейный коэффициент корреляции:

$$r = \frac{12 \cdot 13929{,}44 - 527 \cdot 273{,}34}{\sqrt{\left(12 \cdot 26611 - 527^2\right) \times \left(12 \cdot 7485{,}37 - 273{,}34^2\right)}} = 0{,}92\,.$$

Значение коэффициента корреляции указывает на тесную связь объема выполненных работ от численности работающих.

9.2.2. Определение степени тесноты связи в случае нелинейной зависимости между признаками

В экономических исследованиях линейные однофакторные зависимости используются довольно часто. Однако в реальной действительности преобладают как раз не линейные, а криволинейные связи.

Для оценки тесноты связи в случае парной нелинейной зависимости используется **корреляционное отношение (η)**, в основе которого лежит правило сложения дисперсий (см. п. 7.4.2).

В зависимости от способа расчета различают эмпирическое и теоретическое корреляционное отношение.

Эмпирическое корреляционное отношение ($\eta_э$) (гл. 7, п. 7.4.2) можно использовать в тех случаях, когда исходные данные сгруппированы по факторному признаку и их количество в каждой группе достаточно велико:

$$\eta_э = \sqrt{\frac{\delta^2}{\sigma_o^{\ 2}}} \tag{9.2}$$

где δ^2 — межгрупповая дисперсия результативного признака;

$\sigma_о^2$ — общая дисперсия результативного признака.

Эмпирическое корреляционное отношение является достаточно субъективным показателем, поскольку его значение находится в зависимости от числа групп, на которые разбита изучаемая совокупность. Чем больше сформировано групп, тем больше будет межгрупповая (факторная) дисперсия и тем выше эмпирическое корреляционное отношение.

Более объективным и предпочтительным для оценки степени тесноты связи при нелинейной зависимости является теоретическое корреляционное отношение.

Теоретическое корреляционное отношение (η) целесообразно использовать, когда уже подобрано наиболее подходящее уравнение регрессии (п. 9.2.6) и рассчитаны его параметры. В этом случае числитель формулы представляет собой дисперсию выравненных значений результативного признака, то есть значений, рассчитанных по уравнению регрессии:

$$\eta = \sqrt{\frac{\sum(\hat{y}_i - \bar{y})^2}{\sum(y_i - \bar{y})^2}} \tag{9.3}$$

где $\hat{y}_i$ — теоретические уровни y, рассчитанные по уравнению регрессии;

$\bar{y}$ — среднее значение признака y;

y_i — индивидуальные значения результативного признака.

Значение корреляционного отношения варьируется от 0 до 1 и интерпретируется аналогично значению коэффициента корреляции (табл. 9.1). Направление связи можно установить по группировочной таблице.

Следует иметь в виду, что если для линейной парной зависимости рассчитать корреляционное отношение, то его значение совпадет по абсолютной величине со значением линейного коэффициента корреляции. Таким образом, корреляционное отношение является универсальным показателем тесноты связи.

Пример расчета теоретического корреляционного отношения (на базе уравнения регрессии) рассмотрен в п. 9.2.6.

9.2.3. Оценка существенности линейного коэффициента корреляции

После определения степени тесноты связи между признаками проводится проверка рассчитанного коэффициента корреляции на существенность (значимость).

Дело в том, что при расчете коэффициента изучается достаточно ограниченный объем информации, на основании которой делается вывод о существовании связи между фактическим и результативным признаками. Т.е. анализируемую совокупность данных можно считать выборочной.

Чисто теоретически можно предположить, что при наличии бóльшего объема данных, отклонения от нуля полученной величины коэффициента корреляции может быть целиком вызвано случайными колебаниями тех данных, на основе которых был вычислен этот коэффициент.

В связи с этим и возникает необходимость оценки существенности линейного коэффициента корреляции.

Особенно осторожно следует подходить к истолкованию полученных коэффициентов корреляции при незначительных объемах выборочной совокупности.

Погрешности, обусловленные выборочной информацией, называются **ошибками репрезентативности**, т.е. ошибками представительности исходной информации. Эти ошибки показывают, насколько достоверно данные выборки характеризуют всю генеральную совокупность.

Если количество наблюдений невелико ($n < 30$) ошибку репрезентативности коэффициента корреляции находят по формуле:

$$m_r = \sqrt{\frac{1-r^2}{n-2}}\,. \qquad (9.4)$$

Если количество наблюдений достаточно велико, то ошибка коэффициента парной линейной корреляции определяется так:

$$m_r = \frac{1-r^2}{\sqrt{n}}\,. \qquad (9.5)$$

Затем выдвигается нулевая гипотеза, т.е. гипотеза о том, что зависимость между признаками x и y отсутствует. Для проверки этой гипотезы рассчитывается t-критерий Стьюдента (В. Госсета) — t_r. Расчетный критерий t_r определяется как отношение выборочного коэффициента корреляции r к своей ошибке m_r;

– для небольшого объема выборки (n < 30):

$$t_r = \frac{|r|}{m_r} = |r| \cdot \sqrt{\frac{n-2}{1-r^2}} \qquad (9.6)$$

– для большого объема выборки:

$$t_r = |r| \cdot \frac{\sqrt{n}}{1-r^2}\,. \qquad (9.7)$$

Рассчитанное по формулам (8.6) и (8.7) значение t-критерия Стьюдента сравнивают с табличным значением t-критерия $t_{табл}$ при различных уровнях значимости α (10; 5; 2,5; 2; 1 и т.д.), соответствующих вероятности безошибочности выводов (соответственно 90%; 95%; 97,5%; 98%; 99% и т.д.) и числе степеней свободы k, равным n–2.

Табличное значение критерия Стьюдента выбирается из специальных таблиц (приложение). Если $t_r > t_{табл}$, то линейный коэффициент корреляции является значимым (существенным), т.е. с определенной вероятностью (90%; 95%; 97,5%; 98% или 99%) можно утверждать, что между выбранными признаками имеет место корреляция.

Если же t-критерий коэффициента корреляции меньше его табличного уровня $(t_r < t_{табл})$, то этот коэффициент не является достоверным. Поэтому на основе полученных результатов нельзя сделать каких-либо определенных выводов относительно корреляции изучаемых признаков. Для решения этого вопроса необходимо провести повторные испытания на более обширном материале.

Пример. Оценим существенность (достоверность) полученного в п. 9.2.1 коэффициента корреляции.

Коэффициент корреляции составляет 0,92. Но этот коэффициент содержит ошибку репрезентативности. Так как число наблюдений меньше 30, для расчета ошибки воспользуемся формулой (9.4):

$$m_r = \sqrt{\frac{1-0{,}92^2}{12-1}} = 0{,}12393.$$

Следовательно, ошибка коэффициента парной линейной корреляции составляет ± 0,12393.

Рассчитаем t-критерий Стьюдента парной линейной корреляции t_r. Поскольку число наблюдений невелико (n = 12) используем формулу (9.6):

$$t_r = 0{,}92 \cdot \sqrt{\frac{12-2}{1-0{,}92^2}} = 7{,}42322.$$

В приложении для числа степеней свободы $k = n–2 = 10$ и уровня значимости 1% находим, что $t_{табл}$ = 3,169.

Рассчитанный t-критерий (t_r) превышает порог вероятности безошибочных выводов 99% для которого стандартное значение указанного критерия, при числе степеней свободы равном 10, составляет 3,169.

Поэтому, полученный t-критерий позволяет утверждать с вероятностью 99%, что имеет место корреляционная зависимость объема выполненных работ от численности работающих.

9.2.4. Оценка существенности корреляционного отношения

Расчет оценки существенности корреляционного отношения проводится аналогично расчету этой величины для коэффициента корреляции (п. 9.2.3).

Ошибку выборочного корреляционного отношения m_η рассчитывают по формулам:

– в случае небольшого количества наблюдений выборки (n < 30):

$$m_\eta = \sqrt{\frac{1-\eta^2}{n-2}} \tag{9.8}$$

– при достаточно большом числе наблюдений:

$$m_\eta = \frac{1-\eta^2}{\sqrt{n}} \tag{9.9}$$

Критерий Стьюдента t_η будет определяться по формулам:

– при незначительном количестве наблюдений:

$$t_\eta = \frac{\eta}{m_\eta} = \eta \cdot \sqrt{\frac{n-2}{1-\eta^2}} \tag{9.10}$$

– при достаточно большом числе наблюдений:

$$t_\eta = \frac{\eta}{m_\eta} = \eta \cdot \frac{\sqrt{n}}{1-\eta^2} \tag{9.11}$$

Пример расчета оценки существенности корреляционного отношения приведен в п. 9.2.6.

9.2.5. Определение параметров линейной зависимости

Если установлена достаточная степень тесноты связи между факторным и результативным признаками, то строится уравнение регрессии.

Для определения численных параметров уравнения регрессии обычно используется метод наименьших квадратов и решается система нормальных уравнений.

При линейной зависимости между признаками $\widehat{y}_x = a_0 + a_1 x$ (например, при изучении зависимости уровня оплаты труда (y) от его производительности (x)) система нормальных уравнений выглядит следующим образом:

$$\begin{cases} a_0 n + a_1 \sum x = \sum y \\ a_0 \sum x + a_1 \sum x^2 = \sum xy \end{cases} \qquad (9.12)$$

После расчета параметров уравнения регрессии необходимо определить, насколько точно описывает подобранная функция исходные данные и можно ли использовать полученное уравнение для прогнозирования неизвестных значений *y*.

Для этого сначала рассчитывается **средняя квадратическая ошибка уравнения регрессии (S_e)**, которая показывает, *на сколько в среднем* отклоняются фактические значения результативного признака от соответствующих им значений, рассчитанных по уравнению регрессии:

$$S_e = \sqrt{\frac{\sum \left(y_i - \hat{y}_{x_i}\right)^2}{n-m}}, \qquad (9.13)$$

где y_i — фактические значения результативного признака;

$\hat{y}_{x_i}$ — значения результативного признака, рассчитанные по уравнению регрессии и полученные путем подстановки значений факторного признака (x) в уравнение регререссии;

n — число наблюдений;

m — число параметров в уравнении регрессии.

Чем меньше значение S_e, тем точнее подобранное равнение регрессии описывает исходные данные.

Затем, полученная средняя квадратическая ошибка (S_e) сравнивается со средним уровнем результативного признака $\bar{y}$:

$$\frac{S_e}{\bar{y}} \cdot 100\%. \qquad (9.14)$$

Если полученное отношение не превышает 15%, то считают, что уравнение регрессии достаточно хорошо отражает изучаемую взаимосвязь и может быть использовано в экономической практике.

Пример. Для исходных данных табл. 9.2, предполагая линейную зависимость между признаками, рассчитаем параметры линейного уравнения и среднюю квардратическую ошибку уравнения регрессии.

Для расчетов составим вспомогательную таблицу 9.4.

Таблица 9.4

Вспомогательная таблица для расчета параметров уравнения
и средней квадратической ошибки

№ п/п	Численность работающих, x_i, чел.	Объем выполненных работ, y_i, млн.руб.	$x_i y_i$	x_i^2	$\hat{y}_{x_i}$	$\left(y_i - \hat{y}_{x_i}\right)^2$
1	20	9,61	192,2	400	9,50	0,01
2	79	50,11	3958,69	6241	42,26	61,61
3	39	17,32	675,48	1521	20,05	7,44
4	62	28,77	1783,74	3844	32,82	16,41
5	47	19,34	908,98	2209	24,49	26,53
6	20	15,12	302,4	400	9,50	31,62
7	31	19,34	599,54	961	15,61	13,95
8	39	16,05	625,95	1521	20,05	15,98
9	39	18,34	815,26	1251	20,05	2,92
10	64	31,22	1998,08	4096	33,93	7,35
11	36	19	684	1296	18,38	0,38
12	51	29,12	1485,12	2601	26,71	5,80
Всего	**527**	**273,34**	**13929,44**	**26611**	**273,34**	**190,00**

Определим параметры линейного уравнения регрессии, решив систему нормальных уравнений:

$$\begin{cases} 12 \cdot a_0 + 527 \cdot a_1 = 273{,}34 \\ 527 \cdot a_0 + 26611 \cdot a_1 = 13929{,}44 \end{cases}.$$

$$a_0 = -1{,}61; \quad a_1 = 0{,}56.$$

Линейное уравнение регрессии выглядит следующим образом:

$$\hat{y}_x = -1{,}61 + 0{,}56x.$$

Чтобы оценить, насколько точно подобранная функция описывает реальные данные, рассчитаем среднюю квадратическую ошибку уравнения регрессии (s_e):

$$S_e = \sqrt{\frac{190,00}{12-2}} = 4,36 \text{ млн.руб.}$$

$$\bar{y} = \frac{\sum y}{n} = \frac{273,34}{12} = 22,78 \text{ млн.руб.}$$

Отсюда: $\frac{S_e}{\bar{y}} \cdot 100\% = \frac{4,36}{22,78} \cdot 100\% = 19,14\%$.

Так как полученное значение 19,14% больше 15%, значит построенное уравнение регрессии не достаточно точно отражает взаимосвязь между численностью сотрудников проектных организаций и объемом выполняемых ими работ.

9.2.6. Определение параметров криволинейных зависимостей

В настоящее время не разработано универсального способа, позволяющего предварительно определять вид математической модели исследуемой зависимости.

Для идентификации модели в некоторых случаях могут быть использованы: графический способ; экономический анализ; способ разностей и другие, однако эти способы являются частными, т.е. имеют ограниченное применение.

В последние годы все более широкое распространение находит способ «перебора» функций на ЭВМ. Машина рассматривает и определяет параметры для всех функций, которые включены в программу, а затем, с помощью ряда статистических характеристик отбирается наиболее приемлемая модель.

При парной корреляции наиболее распространенными функциями нелинейной зависимости являются следующие:

1. Параболическая: $\hat{y}_x = a_0 + a_1 x + a_2 x^2$ (например, при изучении зависимости урожайности сельскохозяйственных культур от внесения высоких доз удобрения).

2. Гиперболическая: $\widehat{y}_x = a_0 + a_1 \frac{1}{x}$ (например, анализ зависимости издержек производства на единицу продукции от производительности оборудования, или производительности труда от трудоемкости производства продукции).

3. Показательная: $\widehat{y}_x = a_0 a_1^x$ (например, зависимость производства валовой продукции от фондовооруженности).

Параметры уравнения определяются с помощью системы нормальных уравнений:

– **для параболы** $\left(\widehat{y}_x = a_0 + a_1 x + a_2 x^2\right)$:

$$\begin{cases} a_0 + a_1 x + a_2 x^2 = \sum y \\ a_0 \sum x + a_1 \sum x^2 + a_2 \sum x^3 = \sum xy \\ a_0 \sum x^2 + a_1 \sum x^3 + a_2 \sum x^4 = \sum x^2 y \end{cases} \tag{9.15}$$

– **для гиперболы** $\left(\widehat{y}_x = a_0 + a_1 \frac{1}{x}\right)$:

$$\begin{cases} a_0 + a_1 \frac{1}{x} = \sum y \\ a_0 \sum \frac{1}{x} + a_1 \sum \left(\frac{1}{x}\right)^2 = \sum \frac{1}{x} y \end{cases} \tag{9.16}$$

– **для показательной функции** $\widehat{y}_x = a_0 a_1^x$:

$$\begin{cases} n \lg a_0 + \lg a_1 \sum x = \sum \lg y \\ \lg a_0 \sum x + \lg a_1 \sum x^2 = \sum x \lg y \end{cases} \tag{9.17}$$

Чтобы оценить, насколько точно описывает подобранная функция исходные данные и может ли она быть использована в практической деятельности, также как и в случае линейной зависимости рассчитывается средняя квадратическая ошибка уравнения регрессии по формуле (9.13) и полученное значение сравнивается со средним уровнем результативного признака y по формуле (9.14).

Пример. По исходным данным табл. (9.2) выполним следующее:

1. для построения уравнения регрессии выберем уравнение параболы второго порядка и рассчитаем ее параметры;
2. определим среднюю квадратическую ошибку уравнения регрессии;
3. оценим степень тесноты связи между признаками;
4. оценим существенность полученного показателя тесноты связи.

Решение.

1. Построим вспомогательную таблицу для расчета параметров параболы, средней квадратической ошибки уравнения регрессии и корреляционного отношения (табл. 9.5).

Рассчитаем параметры уравнения параболы второго порядка:

$$\begin{cases} 12a_0 + 527a_1 + 26611a_2 = 273{,}34 \\ 527a_0 + 26611a_1 + 1500389a_2 = 13929{,}44 \\ 26611a_0 + 1500389a_1 + 92011975a_2 = 801421{,}30 \end{cases}$$

$$a_0 = 15{,}80; \quad a_1 = -0{,}28; \quad a_2 = 0{,}01$$

Уравнение параболы будет иметь вид: $\hat{y}_x = 15{,}80 - 0{,}28x + 0{,}01x^2$.

2. Рассчитаем среднюю квадратическую ошибку уравнения регрессии, используя формулы (9.13 и 9.14):

$$S_e = \sqrt{\frac{93{,}01}{12-3}} = 3{,}21 \text{ млн. руб.}$$

Ошибка уравнения регрессии показывает, что теоретические (т.е. рассчитанные по уравнению параболы) значения объема выполненных работ отклоняются от исходных данных *в среднем* на ±3,21 млн. руб.

Для оценки возможности практического использования полученной модели рассчитаем среднее значение зависимого (результативного) признака и сравним с ним среднюю ошибку уравнения регрессии:

$$\bar{y} = \frac{\sum y}{n} = \frac{273{,}34}{12} = 22{,}78 \text{ млн.руб.}$$

$$\frac{3,21}{22,78}\cdot 100\% = 14,11\%.$$

Полученное значение свидетельствует о том, что парабола достаточно точно описывает фактические данные (14,11% < 15%) и может быть использована при дальнейшем анализе и прогнозировании неизвестных значений.

3. Оценим степень тесноты связи между признаками, используя теоретическое корреляционное отношение (формула 9.3).

$$\eta = \sqrt{\frac{1166,13}{1259,14}} = 0,96.$$

Величина корреляционного отношения свидетельствует о наличии тесной зависимости между численностью работающих и объемом выполненных работ.

4. Определим существенность полученного корреляционного отношения.

Так как объем данных небольшой, т.е. меньше 30, то для определения ошибки m_η воспользуемся формулой (9.8):

$$m_\eta = \sqrt{\frac{1 - 0,96^2}{12 - 2}} = 0,0885.$$

Критерий Стьюдента будет равен (формула 9.10):

$$t_\eta = \frac{0,96}{0,0885} = 10,8474.$$

В приложении для числа степеней свободы $k = n–2 = 10$ и уровня значимости 1% находим табличное значение критерия Стьюдента $t_{табл}$= 3,169.

Рассчитанный t-критерий (t_η) больше табличного t-критерия (10,8474 > 3,169). Таким образом, лишь с вероятностью меньшей 1% можно утверждать, что величина t_η = 10,847 могла появиться в силу случайностей. Такое событие является маловероятным, а поэтому можно считать с вероятностью 99%, что действительно существует зависимость между изучаемыми признаками.

9.2.7. Экономическая интерпретация результатов построения однофакторной экономической модели

Экономическая интерпретация результатов построения функции является завершающим этапом получения статистико-экономических зависимостей.

На основе полученных экономических моделей выполняется расчет их аналитических характеристик.

Для однофакторных зависимостей рассчитывают:

– коэффициент детерменации;

– квадрат алиенации;

– коэффициент эластичности.

Коэффициент детерминации K_D показывает удельный вес (долю) факторного признака в общей вариации зависимого показателя. Он рассчитывается как квадрат соответствующего показателя тесноты связи (коэффициента парной корреляции или корреляционного отношения):

$$K_D = r^2; \quad K_D = \eta^2.$$

Коэффициенты детерминации всегда находятся в пределах 0—1. Нередко детерминацию выражают в процентах. При этом ее обычно обозначают через D и называют просто «детерминация»:

$$D = 100\% \cdot r^2; \quad D = 100\% \cdot \eta^2 \tag{8.19}$$

Если детерминация меньше 10%, она считается слабой; когда она находится в пределах 10—50%, ее относят к средней. Детерминация с уровнем свыше 50% — сильная; а при 100% — полная.

Квадрат алиенации *A* характеризует в процентах долю неучтенных факторов в изменениях зависимого признака.

$$A = 100\% - D \tag{9.20}$$

$$A = 100\% \cdot (1 - K_D) \tag{9.21}$$

Под эластичностью подразумевается упругость или гибкость. Применительно к экономическим моделям **коэффициент эластичности** $\bar{\varepsilon}_x$ отражает соотношение темпов прироста зависимого y и факторного x признаков. Он показывает, *на сколько процентов в среднем* прирастает зависимый признак y с приростом факторного признака x на 1%:

– для прямолинейной зависимости $y = a_0 + a_1 x$:

$$\bar{\varepsilon}_x = \frac{a_1 \bar{x}}{a_0 + a_1 \bar{x}} \tag{9.22}$$

– для параболы второго порядка $y_x = a_0 + a_1 x + a_2 x^2$:

$$\bar{\varepsilon}_x = \frac{a_1 \bar{x} + 2 a_2 \bar{x}^2}{a_0 + a_1 \bar{x} + a_2 \bar{x}^2} \tag{9.23}$$

– для гиперболы $y_x = a_0 + a_1 \frac{1}{x}$:

$$\bar{\varepsilon}_x = -\frac{a_1}{a_0 \bar{x} + a_1} \tag{9.24}$$

– для показательной функции $y_x = a_0 a_1^x$:

$$\bar{\varepsilon}_x = \bar{x} \cdot \ln a_1 \tag{9.25}$$

Пример. Подбирая уравнения регрессии к исходным данным табл. 9.2 мы определили, что наиболее подходящей функцией для моделирования зависимости между численностью работников и объемом выполняемых работ является нелинейная функция — парабола второго порядка (п. 9.2.6), ей же соответствует более высокая степень тесноты связи.

Поэтому аналитические характеристики будем рассчитывать для полученной параболической зависимости:

$$\hat{y}_x = 15{,}80 - 0{,}28x + 0{,}01x^2.$$

Корреляционное отношение здесь равно 0,96. Поэтому детерминация составляет:

$$D = 100\% \cdot 0{,}96^2 = 92{,}16\%.$$

Таким образом, удельный вес основного фактора (т.е. численности работающих) в вариации объема выполненных работ составляет 92,16%.

Квадрат алиенации составляет:

$$A = 100\% - 92{,}16\% = 7{,}84\%.$$

Значит, доля неучтенных факторов в вариации объема выполненных работ составляет 7,84%.

Для расчета коэффициента эластичности сначала рассчитаем среднее значение *x*.

$$\bar{x} = \frac{\sum x}{n} = \frac{527}{12} = 43{,}92$$

$$\bar{\varepsilon}_x = \frac{-0{,}28 \cdot 43{,}92 + 2 \cdot 0{,}01 \cdot 43{,}92^2}{15{,}80 - 0{,}28 \cdot 43{,}92 + 0{,}01 \cdot 43{,}92^2} = 1{,}153.$$

Следовательно, увеличение численности работающих на 1% ведет к росту объема выполненных работ в среднем на 1,153%.

9.3. *Непараметрические методы оценки связи*

Методы корреляционного анализа не универсальны: их можно применять, если все изучаемые признаки являются количественными. При использовании этих методов нельзя обойтись без вычисления основных параметров распределения (средних величин, дисперсий), поэтому они получили название **параметрических методов**.

Между тем в статистической практике приходится сталкиваться с задачами измерения связи между качественными признаками, к которым параметрические методы анализа неприменимы. Статистической наукой разработаны методы, с помощью которых можно измерить связь между явлениями, не используя при этом количественные значения признака, а значит, и параметры распределения. Такие методы получили название **непараметрических**.

Наиболее распространенными показателями оценки тесноты связи являются коэффициент ассоциации Юла, коэффициент контингенции, коэффициент корреляции рангов Спирмэна, коэффициент конкордации и др.

Коэффициент ассоциации Юла используется в тех случаях, когда изучаются взаимосвязи между двумя альтернативными признаками (то есть признаками, имеющими только два возможных значения: да — нет; экзамен сдал — не сдал, состоит в браке — холост и т.д.).

Расчет коэффициента ассоциации проводится с помощью «таблицы четырех полей» (табл. 9.6).

Таблица 9.6

Факторный признак	Результативный признак		Всего
	первое значение альтернативного признака	второе значение альтернативного признака	
Первое значение альтернативного факторного признака	a	b	$a + b$
Второе значение альтернативного факторного признака	c	d	$c + d$
Всего	$a + c$	$b + d$	n

Коэффициент ассоциации Юла рассчитывается по формуле:

$$K_a = \frac{ad - bc}{ad + bc}.$$

Пример. Известны следующие данные о распределении рабочих предприятия по стажу и разряду работы (табл. 9.7), чел. Необходимо установить тесноту связи между стажем работы и разрядом рабочих, рассчитав коэффициент ассоциации.

Таблица 9.7

Разряд рабочих	Стаж работы		Всего
	До 5 лет	5 лет и более	
1—3	40 (a)	20 (b)	60
4—6	10 (c)	50 (d)	60
Всего	50	70	120

$$K_a = \frac{40 \cdot 50 - 20 \cdot 10}{40 \cdot 50 + 20 \cdot 10} = 0{,}82.$$

Величина коэффициента ассоциации говорит о наличии тесной связи между анализируемыми альтернативными признаками (табл.9.1).

Если хотя бы один из четырех показателей отсутствует, величина коэффициента ассоциации будет равна единице, что дает преувеличенную оценку степени тесноты связи между признаками. В этом случае следует отдать предпочтение **коэффициенту контингенции K_k**:

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

Коэффициент контингенции всегда меньше коэффициента ассоциации. Связь между признаками считается подтвержденной, если $K_a > 0{,}5$, а $K_k > 0{,}3$.

В социально-экономических исследованиях нередко встречаются ситуации, когда признак является качественным, однако единицы совокупности можно упорядочить. Такое упорядочивание единиц совокупности называется **ранжированием**. Например, ранжирование любой совокупности людей по уровню образования, профессии и т.д.

Для оценки степени тесноты связи между признаками проводится ранжирование единиц совокупности по результативному и факторному (или факторным) признакам. Ранжирование всех признаков проводится в одном и том же направлении: либо по возрастанию, либо по убыванию.

При ранжировании каждой единице совокупности присваивается ранг, т.е. порядковый номер. При совпадении значений признака у отдельных единиц им присваивается объединенный средний порядковый номер. Например, если у 4-й и 5-й единиц совокупности значения признаков одинаковы, оба получат ранг, равный (4 + 5) : 2 = 4,5.

Если связь между x и y прямая, то с увеличением ранга признака x ранг признака y также будет возрастать, при обратной связи возрастанию рангов признака x будет, как правило, соответствовать убывание рангов признака y.

Измерение связи между ранжированными факторным и результативным признаками проводится с помощью **коэффициента корреляции рангов Спирмэна** ρ:

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

где d — разность рангов признаков x и y;

n — число единиц совокупности.

Пример. Экспертами оценивались вкусовые качества различных сортов сыра. Суммарные оценки получены следующие (табл. 9.8). Оцените, насколько согласуется оценка сыра с его ценой. Проверьте эту гипотезу с помощью коэффициента корреляции рангов Спирмэна.

Таблица 9.8

Сорт сыра	Оценка в баллах, y	Цена, x руб./кг
1	11	157
2	14	160
3	17	200
4	15	210
5	13	170
6	13	185
7	18	180
8	10	115
9	19	230
10	25	240

Проведем ранжирование факторного (цена за 1 кг) и результативного (оценка экспертов) признаков. Результаты ранжирования представлены в табл. 9.9.

Таблица 9.9

Сорт сыра	Ранг		Разность рангов	d^2
	Оценки в баллах, R_y	Цены, R_x	$d = R_y - R_x$	
1	2	2	0	0
2	5	3	2	4
3	7	7	0	0
4	6	8	-2	4
5	3,5	4	-0,5	0,25
6	3,5	6	-2,5	6,25
7	8	5	3	9
8	1	1	0	0
9	9	9	0	0
10	10	10	0	0
Итого	—	—	—	**23,5**

Значение коэффициента корреляции рангов Спирмэна составит:

$$\rho = 1 - \frac{6 \cdot 23{,}5}{10\left(10^2 - 1\right)} = 0{,}858.$$

Это свидетельствует о прямой и довольно значительной силе связи между качеством сыра и его ценой.

Если необходимо оценить взаимосвязь *нескольких* признаков, поддающихся ранжированию, то можно использовать **коэффициент конкордации** ω:

$$\omega = \frac{12S}{m^2\left(n^3 - n\right)},$$

где m — число факторов;

n — число ранжируемых единиц;

S — отклонение суммы квадратов рангов от средней квадратов рангов.

Пример: десять предприятий, выпускающих однотипную продукцию были ранжированы экспертами по уровню качества выпускаемой продукции; затратам на ее продвижение и спросу на продукцию (табл. 9.10).

Необходимо оценить степень тесноты связи между признаками с помощью коэффициента конкордации.

Таблица 9.10

Предприятие	Ранг по показателям			Сумма строк	Квадраты сумм
	качество выпускаемой продукции	затраты на продвижение продукции	уровень спроса на продукцию		
1	4	4	3	11,0	121,00
2	3	1	1	5,0	25,00
3	1	3	2	6,0	36,00
4	6	7	5	18,0	324,00
5	5	5	7	17,0	289,00
6	8	6	6	20,0	400,00
7	2	2	4	8,0	64,00
8	7	8	8	23,0	529,00
9	10	9	9,5	28,5	812,25
10	9	10	9,5	28,5	812,25
Всего	—	—	—	**165,0**	**3412,50**

$$S = 3412{,}5 - \frac{(165{,}0)^2}{10} = 690$$

$$\omega = \frac{12 \cdot 690}{3^2\left(10^3 - 10\right)} = 0{,}929.$$

Коэффициент конкордации свидетельствует о возможном наличии достаточно тесной зависимости между изучаемыми признаками.

Вопросы для самоконтроля

1. Какие признаки называются факторными? результативными?
2. В чем состоит отличие между корреляционной и функциональной связью?
3. Какие задачи решаются при корреляционном анализе?
4. Назовите показатели тесноты связи между двумя признаками.
5. Как оценить существенность линейного коэффициента корреляции?
6. Какие виды уравнений парной регрессии наиболее распространены при характеристике социально-экономических явлений?
7. Что характеризует и как рассчитывается средняя квадратическая ошибка уравнения регрессии?
8. Для чего рассчитывается коэффициент детерминации? В каких пределах он находится?
9. Что показывает квадрат алиенации? Как он рассчитывается?
10. Что такое коэффициент эластичности? В чем его назначение?
11. Что такое множественная корреляция?
12. Что такое непараметрические показатели тесноты связи?
13. В каких случаях рассчитывается коэффициент Юла?
14. Как определяется коэффициент контингенции?
15. Коэффициент корреляции рангов Спирмена и его назначение.
16. В каких случаях используется и как рассчитывается коэффициент конкордации?

10. Ряды динамики

10.1. *Основные понятия*

Одной из важнейших задач статистики является изучение явлений во времени. Эта задача решается при помощи составления и анализа рядов динамики.

Ряд динамики (временной ряд) представляет собой числовые значения определенного статистического показателя (y_i) в последовательные моменты или периоды времени (t_i).

Числовые значения того или иного статистического показателя, составляющие динамический ряд, называются **уровнями ряда**.

Ряды динамики, как правило, представляют в виде таблиц или графически.

В зависимости от характера уровней ряда различают два вида динамических рядов: интервальные и моментные.

Интервальным называется такой ряд динамики, уровни которого характеризуют размер явления за определенный период времени (месяц, квартал, год и т.п.) — табл. 10.1.

Таблица 10.1

	2005	2006	2007	2008
Объем перевозок грузов судоходной компанией, тыс. т	315	304	290	299

Примерами этого ряда могут служить данные о перевозках грузов по месяцам или годам; данные о выпуске книг, брошюр и журналов в РФ, объемах капитального строительства и т.п.

Моментным называется ряд динамики, уровни которого характеризуют состояние явления на определенные моменты времени. Например, численность сотрудников фирмы по состоянию на первое число каждого месяца; уровень запасов какого-либо материала на начало месяца; число дошкольных учреждений в России на конец года, курс валют и ценных бумаг; средний возраст населения и т.п. (табл. 10.2, 10.3).

Таблица 10.2

	2005	2006	2007
Среднегодовая численность занятых на транспорте, тыс. чел.	4429	4471	4477

Таблица 10.3

	2003	2004	2005	2006
Эксплуатационная длина автомобильных дорог РФ с твердым покрытием (на конец годы), тыс. км	745	738	725	755

Важное аналитическое отличие моментных рядов от интервальных состоит в том, что сумма уровней моментных рядов не имеет реального экономического содержания, хотя иногда и подсчитывается в промежуточных расчетах. Сумма же уровней в интервальных рядах динамики дает вполне реальный показатель — общий объем перевозок за год, или несколько лет; общий итог выпуска книжной продукции; общие объемы капитального строительства за несколько лет и т.п.

При изучении явлений в статистике приходится иметь дело с различными видами динамических рядов. При этом основным требованием предъявляемым к анализируемым рядам, является сопоставимость их уровней.

Несопоставимость уровней может возникнуть по различным причинам, основными из которых можно назвать следующие:

– изменение территории, к которой отнесены те или иные показатели.
– изменение даты учета.
– изменение методологии учета или расчета показателей.
– изменение цен (для стоимостных показателей).
– различная продолжительность периодов, к которым относятся уровни.

Могут быть и другие причины несопоставимости. Следовательно, прежде чем анализировать динамический ряд, надо убедиться в сопоставимости уровней ряда и если последняя отсутствует, добиться ее дополнительными расчетами.

10.2. *Показатели рядов динамики*

Каждый динамический ряд состоит из n-го числа варьирующих во времени показателей. Обычно первый уровень ряда называют начальным уровнем (y_1), а последний — конечным (y_n).

Анализ направления и размера изменений уровней динамического ряда во времени проводится с помощью следующих показателей:

1. абсолютных приростов;
2. коэффициентов роста или темпов роста;
3. темпов прироста;
4. абсолютного значения одного процента прироста.

В случае, когда сравнение проводится каким-то одним уровнем, обычно начальным, получают **базисные показатели динамики**.

Если сравнение производится с предшествующим (то есть предыдущим) моментом или периодом времени, то говорят о **цепных показателях динамики**.

Абсолютный прирост (Δ) показывает, на сколько в абсолютном выражении уровень текущего периода больше (или меньше) базисного:

– базисные абсолютные приросты (постоянная база сравнения)

$$\Delta' = y_i - y_k$$

– цепные абсолютные приросты (переменная база сравнения)

$$\Delta = y_i - y_{i-1},$$

где y_k — уровень, принятый за постоянную базу сравнения (часто начальный уровень);

y_i — уровень любого периода (кроме первого), называемый уровнем текущего периода;

y_{i-1} — уровень периода, предшествующего текущему.

Между базисными и цепными абсолютными приростами существует взаимосвязь: **сумма цепных абсолютных приростов равна базисному абсолютному приросту последнего периода ряда динамики**. Эту взаимосвязь можно выразить следующей формулой:

$$\sum \Delta = y_n - y_1 .$$

Пример. Рассчитайте абсолютные приросты для следующего временного ряда.

Таблица 10.4

	2006	2007	2008	2009	2010
Объем перевозок грузов транспортным предприятием, тыс. т	3,2	4,0	3,8	4,3	4,9

Базисные абсолютные приросты:

Δ_1' : $4{,}0 - 3{,}2 = 0{,}8$ тыс.т

Δ_2' : $3{,}8 - 3{,}2 = 0{,}6$ тыс.т

Δ_3' : $4{,}3 - 3{,}2 = 1{,}1$ тыс.т

Δ_4' : $4{,}9 - 3{,}2 = 1{,}7$ тыс.т

Цепные абсолютные приросты:

Δ_1 : $4{,}0 - 3{,}2 = 0{,}8$ тыс.т

Δ_2 : $3{,}8 - 4{,}0 = -0{,}2$ тыс.т

Δ_3 : $4{,}3 - 3{,}8 = 0{,}5$ тыс.т

Δ_4 : $4{,}9 - 4{,}3 = 0{,}6$ тыс.т

Проверка:

$$0{,}8 - 0{,}2 + 0{,}5 + 0{,}6 = 1{,}7 \text{ тыс.т.}$$

Коэффициент роста (K_p) показывает, во сколько раз один уровень больше или меньше уровня, принятого за базу сравнения:

– базисные коэффициенты роста (постоянная база сравнения):

$$K_p' = \frac{y_i}{y_k}$$

– цепные коэффициенты роста (переменная база сравнения):

$$K_p = \frac{y_i}{y_{i-1}} .$$

Между базисными и цепными коэффициентами роста существует взаимосвязь: **произведение последовательных цепных коэффициентов роста равно базисному коэффициенту роста за весь период**.

$$K_{p1} \cdot K_{p2} \cdot K_{p3} \cdot \ldots \cdot K_{pn-1} = \frac{y_n}{y_1}.$$

Если коэффициенты роста выражают в процентах, то их называют **темпами роста (T_p)**:

– базисные темпы роста: $T'_p = \frac{y_i}{y_k} \cdot 100\%;$

– цепные темпы роста: $T_p = \frac{y_{i-1}}{yi} \cdot 100\%.$

Пример.

базисные коэффициенты роста:

$$K'_{p1} = \frac{4{,}0}{3{,}2} = 1{,}25$$

$$K'_{p2} = \frac{3{,}8}{3{,}2} = 1{,}18$$

$$K'_{p3} = \frac{4{,}3}{3{,}2} = 1{,}34$$

$$K'_{p4} = \frac{4{,}9}{3{,}2} = 1{,}53$$

цепные коэффициенты роста:

$$K_{p1} = \frac{4{,}0}{3{,}2} = 1{,}25$$

$$K_{p2} = \frac{3{,}8}{4{,}0} = 0{,}95$$

$$K_{p3} = \frac{4{,}3}{3{,}8} = 1{,}13$$

$$K_{p4} = \frac{4{,}9}{4{,}3} = 1{,}14$$

Проверка: 1,25 · 0,95 · 1,13 · 1,14 = 1,53.

Темп прироста (T_n) показывает, *на сколько процентов* один уровень больше или меньше другого, принятого за базу сравнения:

базисный темп прироста:

$$T'_n = T'_p - 100\%, \text{ или}$$

$$T'_n = (K'_p - 1) 100\%, \text{ или}$$

$$T'_n = \frac{\Delta'}{y_k} \cdot 100\%$$

цепной темп прироста:

$$T_n = T_p - 100\%, \text{ или}$$

$$T_n = (K_p - 1) 100\%, \text{ или}$$

$$T_n = \frac{\Delta}{y_{i-1}} \cdot 100\%$$

Абсолютное значение 1% прироста (*A*) показывает, какая абсолютная величина скрывается за относительным показателем — одним процентом прироста:

– базисное абсолютное значение 1% прироста: $A' = \dfrac{y_k}{100\%}$

– цепное абсолютное значение 1% прироста: $A = \dfrac{y_{i-1}}{100\%} = \dfrac{\Delta}{T_n}$.

Как видно из формул, расчет этого показателя имеет экономический смысл только на цепной основе, поскольку на базисной основе для всех уровней будет получено одно и то же значение показателя — сотая часть базисного (первого) уровня.

Для нашего примера:

$$A_1 = \frac{3{,}2}{100} = 0{,}032 \text{ тыс.т.}$$

$$A_2 = \frac{4{,}0}{100} = 0{,}040 \text{ тыс.т.}$$

$$A_3 = \frac{3{,}8}{100} = 0{,}038 \text{ тыс.т.}$$

$$A_4 = \frac{4{,}3}{100} = 0{,}043 \text{ тыс.т.}$$

Этот показатель имеет важное практическое значение в экономическом анализе. Так, в динамических рядах, уровни которых постоянно растут, темпы роста могут замедляться или оставаться на одном уровне, а значение одного процента прироста расти.

Приведенные выше показатели являются основными характеристиками, используемыми при анализе рядов динамики. Они позволяют судить об изменении уровней в абсолютном и относительном выражении.

Для характеристики динамики явлений в ряде случаев используются **пункты роста (%)**, когда сравнение производится с отдаленным периодом.

Пункты роста представляют собой разность темпов прироста с постоянной базой двух смежных периодов. Пункты роста можно складывать, в результате получают темп прироста соответствующего периода по сравнению с базисным (табл. 10.5).

Таблица 10.5

	2006	2007	2008	200	2010
Уровень ряда, тыс.т	3,2	4,0	3,8	4,3	4,9
Темп роста с постоянной базой, %	—	125	118,7	134,4	153,1
Темп прироста с постоянной базой, %	—	25	18,7	34,4	53,1
Пункты роста, %	—	25	-6,3	15,7	18,7

При сопоставлении динамики развития двух явлений можно сравнивать коэффициенты роста или темпы прироста двух динамических рядов за одинаковые промежутки времени. Эти показатели называются **коэффициентами опережения (**$К_{оп}$**)**:

$$K_{оп} = \frac{K_p(\geq)}{K_p(\leq)}, \qquad K_{оп} = \frac{T_n(\geq)}{T_n(\leq)}.$$

где $K_p(\geq)$ - больший коэффициент роста;

$K_p(\leq)$ - меньший коэффициент роста;

$T_n(\geq)$ - больший темп прироста;

$T_n(\leq)$ - меньший темп прироста.

С помощью коэффициентов опережения могут сопоставляться ряды динамики одинакового содержания, но имеющие отношение к разным территориям (районам, областям), разным предприятиям, а также ряды динамики разного содержания, которые характеризуют один и тот же объект.

10.3. *Средние показатели рядов динамики*

Для обобщенной характеристики изменений временного ряда используют **средние величины**:

– средний уровень ряда;
– средний абсолютный прирост;
– средний коэффициент роста;
– средний темп роста;
– средний темп прироста;
– средняя величина абсолютного значения 1% прироста.

Средний уровень ряда $(\bar{y})$ аналогичен средней арифметической в рядах распределения. Методы его расчета различны для интервальных и моментных рядов.

В **интервальном ряду с равными интервалами** средний уровень рассчитывается по простой средней арифметической:

$$\bar{y} = \frac{\sum y_i}{n}.$$

В **интервальном ряду с неравными временными промежутками** средний уровень рассчитывается по формуле взвешенной средней арифметической:

$$\bar{y} = \frac{\sum y_i t_i}{\sum t_i}.$$

В **моментных рядах динамики с равными интервалами** средний уровень определяется по формуле средней хронологической:

$$\bar{y} = \frac{0{,}5 \cdot y_1 + y_2 + y_3 + \ldots + y_{n-1} + 0{,}5 y_n}{n-1} = \frac{0{,}5(y_1 + y_n) + \sum_{i=2}^{n-1} y_i}{n-1}.$$

Если для **моментного ряда** динамики есть данные только на начало и конец периода, то средний уровень может быть рассчитан по формуле:

$$\bar{y} = \frac{y_1 + y_n}{2},$$

где y_1, y_n — уровни ряда, соответственно, на начало и конец периода.

Средний абсолютный прирост $(\overline{\Delta})$ показывает, *на сколько в среднем* увеличивается или уменьшается уровень ряда по сравнению с предыдущим уровнем:

$$\overline{\Delta} = \frac{\sum \Delta}{n-1} = \frac{y_n - y_1}{n-1}.$$

Средний коэффициент роста $(\overline{K}_p)$ показывает, *во сколько раз в среднем* один уровень больше или меньше предыдущего уровня.

При расчете среднего коэффициента роста используется формула средней геометрической:

$$\overline{K}_p = \sqrt[n-1]{K_{p1} \cdot K_{p2} \cdot K_{p3} \ldots \cdot K_{pn-1}},$$

где K_{p1}, K_{p2}, K_{p3}, ...K_{pn-1} - цепные коэффициенты роста.

n – число уровней ряда.

Средний коэффициент роста можно определить также по данным последнего и первого уровней ряда:

$$\overline{K} = \sqrt[n-1]{\frac{y_2}{y_1} \frac{y_3}{y_2} \frac{y_4}{y_3} \ldots \frac{y_n}{y_{n-1}}} = \sqrt[n-1]{\frac{y_n}{y_1}}.$$

Средний темп роста $(\overline{T}_p)$ показывает, *сколько процентов в среднем* составляет один уровень ряда относительно предыдущего уровня:

$$\overline{T}_p = \overline{K}_p \cdot 100\%.$$

Средний темп прироста $(\overline{T}_n)$ рассчитывается на основе среднего темпа роста и показывает, *на сколько процентов в среднем* один уровень больше или меньше предыдущего уровня.

$$\overline{T}_n = \overline{T}_p - 100\%.$$

Средняя величина абсолютного значения 1% прироста $(\overline{A})$:

$$\overline{A} = \frac{\overline{\Delta}}{\overline{T}_n}.$$

10.4. *Выявление и характеристика основной тенденции развития*

Закономерность в изменении уровней ряда в одних случаях проявляется довольно наглядно, в других она может затушевываться колебаниями, вызываемыми случайными или другими причинами. В этих случаях одной из первых задач исследования является выявление **основной тенденции в изменении уровней**. Основную тенденцию, отражающую общее направления развития, называют **трендом**.

Выявление общей тенденции динамического ряда (тренда) обеспечивается при помощи особых приемов, наиболее распространенными из которых являются:

– метод укрепления интервалов;
– метод скользящей средней;
– метод аналитического выравнивания.

Метод укрупнения интервалов, заключается в определении итоговых или средних показателей для укрупненных интервалов времени (табл. 10.6).

Таблица 10.6

Объемы перевозок грузов транспортным предприятием, тыс.т

I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII
276,8	251,0	289,7	275,6	280,7	270,2	274,4	279,3	273,2	290,71	287,1	303,8

Для выявления общей тенденции роста выпуска продукции произведем укрупнение интервалов путем объединения исходных (месячных) данных в квартальные. Получаем показатели грузооборота транспорта по кварталам:

I квартал	II квартал	III квартал	IV квартал
817,5	826,5	826,9	881,6

В результате укрупнения интервалов общая тенденция роста грузооборота выступает более отчетливо.

Довольно часто при обработке динамического ряда с целью определения тенденции развития применяют сглаживание **методом скользящей средней**. Суть этого метода заключается в замене фактических уровней рядом подвижных (скользящих) средних, которые рассчитываются для определения последовательно подвижных (скользящих) интервалов и относятся к середине каждого из них. Сглаживание указанным способом можно производить по любому числу членов ряда.

Пример. Выявить тренд по данным табл. 10.6, используя метод скользящих средних.

Для расчета скользящих средних построим вспомогательную таблицу 10.7.

Таблица 10.7

Вспомогательная таблица

Месяцы	Объем перевозок грузов транспортным предприятием, тыс.т	Скользящая сумма из 3 членов ряда	Скользящая средняя из 3 членов ряда
I	276,8	—	—
II	251,0	817,5	272,5
III	289,7	816,3	272,1
IV	275,6	846,0	282,0
V	280,7	826,5	275,5
VI	270,2	825,3	275,1
VII	274,4	823,9	274,6
VIII	279,3	826,9	275,6
IX	273,2	843,2	281,1
X	290,7	851,0	283,7
XI	287,1	881,6	293,9
XII	303,8	—	—

Для сглаживания ряда, необходимо суммировать последовательно по 3 члена ряда и результаты делить на 3.

Первая средняя:

$$\bar{y}_1 = \frac{y_1 + y_2 + y_3}{3} = \frac{276{,}8 + 251{,}0 + 289{,}7}{3} = \frac{817{,}5}{3} = 272{,}5 \text{ (февраль)}$$

Вторая средняя:

$$\bar{y}_2 = \frac{y_2 + y_3 + y_4}{3} = \frac{251{,}0 + 289{,}7 + 275{,}6}{3} = \frac{816{,}3}{3} = 272{,}1 \text{ (март)}$$

Третья средняя:

$$\bar{y}_3 = \frac{y_3 + y_4 + y_5}{3} = \frac{289{,}7 + 275{,}6 + 280{,}7}{3} = \frac{846{,}0}{3} = 282{,}0 \text{ (апрель) и т.д.}$$

Недостатком этого метода является то, что сглаженный ряд укорачивается по сравнению с фактическим на $\frac{n-1}{2}$ члена ряда с одного и другого конца (n — число уровней ряда).

Более совершенным методом обработки динамических рядов с целью установления основной тенденции развития является подбор наиболее подходящей математической функции — **метод аналитического выравнивания**.

Простейшими трендовыми моделями, используемыми для аналитического выравнивания являются:

прямая вида $\hat{y}_t = a_0 + a_1 t$;

парабола 2-го порядка $\hat{y}_t = a_0 + a_1 t + a_2 t^2$;

показательная функция $\hat{y}_t = a_0 a_1^t$

и другие.

Параметры уравнений находятся путем решения системы нормальных уравнений, полученных способом наименьших квадратов.

Вычисление упрощается, если (при равных интервалах) отсчет времени t вести от середины ряда.

При нечетном числе уровней ряда серединная точка принимается за 0; тогда предшествующие периоды обозначаются, соответственно, через –1, –2, –3 и т.д., а последующие за средним периоды, соответственно, через +1, +2, +3 и т.д.

При четном числе уровней ряда два серединных момента времени принимаются за –1 и +1, а величина интервала принимается за 2 (так что, считая от середины ряда, $t = \pm 1, \pm 3, \pm 5$ и т.д.).

В этом случае $\sum t = 0$ и системы уравнений будут выглядеть следующим образом:

– для линейной функции $\widehat{y}_t = a_0 + a_1 t$

$$\begin{cases} a_0 n = \sum y \\ a_1 \sum t^2 = \sum yt \end{cases}$$

– для параболы второго порядка $\widehat{y}_t = a_0 + a_1 t + a_2 t^2$

$$\begin{cases} a_0 n + a_1^2 \sum t^2 = \sum y \\ a_1 \sum t^2 = \sum yt \\ a_0 \sum t^2 + a_2 \sum t^4 = \sum yt^2 \end{cases}$$

– для показательной функции $\widehat{y}_t = a_0 a_1^t$:

$$\begin{aligned} &\sum \lg y = n \lg a_0 \\ &\sum t \lg y = \lg a_1 \sum t^2 \end{aligned}.$$

Пример: Провести аналитическое выравнивание данных, представленных в табл. 10.8 по прямой:

Таблица 10.8

Объем перевозок грузов транспортным предприятием

Год	Перевозки грузов, (y) тыс.т	Условное обозначение времени (t)	t^2	y_t	$y_t = 7{,}78 + 1{,}75t$
2006	4,8	-2	4	-9,6	4,28
2007	5,9	-1	1	-5,9	6,03
2008	6,7	0	0	0	7,78
2009	10,0	1	1	+10,0	9,53
2010	11,5	2	4	+23,0	11,28
$\boldsymbol{n = 5}$	$\boldsymbol{\Sigma y = 38{,}9}$	$\boldsymbol{\Sigma t = 0}$	$\boldsymbol{\Sigma t^2 = 10}$	$\boldsymbol{\Sigma yt = 17{,}5}$	$\boldsymbol{\Sigma y_t = 38{,}9}$

Используя рассчитанные итоги, определяем:

$$a_0 = \frac{\sum y}{n} = \frac{38{,}9}{5} = 7{,}78$$

$$a_1 = \frac{\sum yt}{\sum t^2} = \frac{17{,}5}{10} = 1{,}75.$$

Отсюда $\widehat{y}_t = 7{,}78 + 1{,}75t$.

Подставляя последовательно в данную формулу значения $t = -2;-1;0;1;2$ находим выравненные уровни:

для 2006 г. $\hat{y}_t = 7{,}78 + 1{,}75 \times (-2) = 4{,}28$ тыс.т

для 2007 г. $\hat{y}_t = 7{,}78 + 1{,}75 \times (-1) = 6{,}03$ тыс.т и т.д.

Выравненные уровни показаны в последней графе таблицы.

Полученные величины теоретических уровней ряда $\hat{y}_t$, нанесем пунктирной линией на график с фактическими данными. Эта линия и есть графический образ основной тенденции (или тренда) перевозок грузов (рис. 10.1).

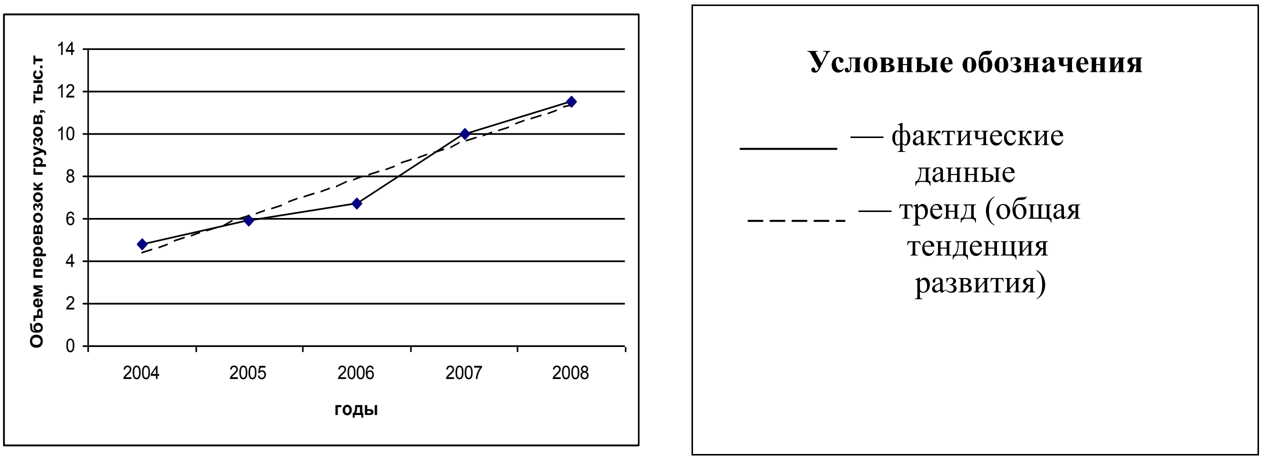


Рис. 10.1. Динамика перевозок грузов транспортным предприятием в 2006—2010 гг.

По полученным моделям (прямой, параболы, гиперболы и т.д.) рассчитывается **стандартная ошибка аппроксимации (среднее квадратическое отклонение тренда)**.

Среднее квадратическое отклонение ($S_{\hat{y}}$) показывает, *на сколько в среднем* отклоняются значения, рассчитанные по уравнению тренда от фактических уровней динамического ряда:

$$S_{\hat{y}} = \sqrt{\frac{\sum \left(y_{t_i} - \hat{y}_{t_i}\right)^2}{n - m}}$$

где y_{t_i} — фактические значения уровней динамического ряда;

$\hat{y}_{t_i}$ — соответствующие значения, рассчитанные по уравнению тренда;

n — число уровней ряда;

m — число параметров в уравнении тренда.

Чем меньше $S_{\hat{y}}$, тем точнее описывает тренд фактические данные.

Величина **относительной ошибки** будет определяться следующим образом:

$$\frac{S_{\hat{y}}}{\bar{y}} \cdot 100\% .$$

10.5. *Анализ сезонных колебаний*

При анализе рядов динамики важное значение имеет выявление сезонных колебаний.

Сезонными колебаниями называются **устойчивые**, повторяющиеся из года в год **внутригодовые** изменения в ряду динамики. Сезонные колебания возникают, например, при переработке сельскохозяйственного сырья в промышленности; при перевозке сельскохозяйственных грузов речным транспортом; при использовании электроэнергии в зависимости от времени года; перевозку пассажиров и т.д.

Знание особенностей сезонных колебаний может быть использовано при решении практических задач: расчете запасов сырья и материалов, определении потребности в рабочей силе, потребности во флоте для перевозок и т.д.

Измерение сезонных колебаний в статистике производится путем вычисления **индексов сезонности**.

Индексы сезонности (i_c) показывают, сколько процентов составляет уровень какого-либо месяца по отношению к среднегодовому уровню, либо по отношению к уровню, вычисляемому по уравнению общей тенденции развития.

Если из года в год уровни ряда по абсолютным значениям остаются примерно постоянными, то сезонные колебания рассчитываются относительно среднегодового уровня.

Если временной ряд содержит определенную тенденцию в развитии, то для расчета индексов сезонности исходные данные сначала выравнивают, чтобы выявить эту тенденцию, а затем сравнивают колебания уровней ряда с соответствующими точками тренда.

Для большей надежности индексы сезонности рассчитываются по данным не менее, чем за 3 года.

Ниже будет рассмотрен первый метод расчета индексов сезонности для случая, когда ряд динамики не содержит ярко выраженной тенденции в развитии.

Если годовой уровень явления из года в год остается относительно неизменным, то индексы сезонности рассчитываются по формуле:

$$i_{c_i} = \frac{\bar{y}_i}{\bar{y}} \cdot 100\%,$$

где $\bar{y}_i$ — средняя из фактических уровней одноименных месяцев;

$\bar{y}$ — общая средняя за исследуемый период.

Для сопоставления величины сезонных колебаний по нескольким объектам или периодам может быть использовано **среднее квадратическое отклонение**, рассчитываемое по формуле:

$$\sigma = \sqrt{\frac{\sum \left(i_{c_i} - 100\right)^2}{n}},$$

где i_{c_i} — индекс сезонности для каждого месяца;

n — число месяцев (12).

Пример. Имеются данные о вводе в действие жилых домов предприятиями города (тыс.кв.м общей площади) за 3 года (табл. 10.9). Данные условные.

На основе приведенных данных требуется:

1) выявить наличие сезонной неравномерности;

2) определить величину сезонной волны, используя индексы сезонности.

Таблица 10.9

Динамика ввода в действие жилых домов предприятиями города
(тыс.кв.м общей площади) в 2006—2008 гг. и расчет индексов сезонности

Месяцы	Ввод жилых домов, тыс.кв.м общей площади				Индексы сезонности, %
	2006	2007	2008	средняя за 3 одноименных месяца, $\bar{y}_i$	
Январь	0,4	0,5	0,7	0,5	18,5
Февраль	0,6	0,6	0,9	0,7	25,9
Март	3,0	3,1	2,5	2,9	107,4
Апрель	1,1	0,8	1,0	1,0	37,0
Май	1,2	1,2	1,1	1,2	44,4
Июнь	3,9	3,4	3,4	3,6	133,3
Июль	1,5	1,2	1,2	1,3	48,1
Август	1,4	1,4	1,3	1,4	51,8
Сентябрь	3,2	3,1	3,4	3,2	118,5
Октябрь	1,5	1,2	1,6	1,4	51,8
Ноябрь	1,9	1,9	2,2	2,0	74,1
Декабрь	13,0	12,3	12,7	12,6	466,7
Итого	**32,7**	**30,7**	**32,0**		—
Общая средняя за исследуемый период, $\bar{y}$	—	—	—	2,7	—

Для выявления наличия сезонной неравномерности используем графический метод (рис. 10.2).

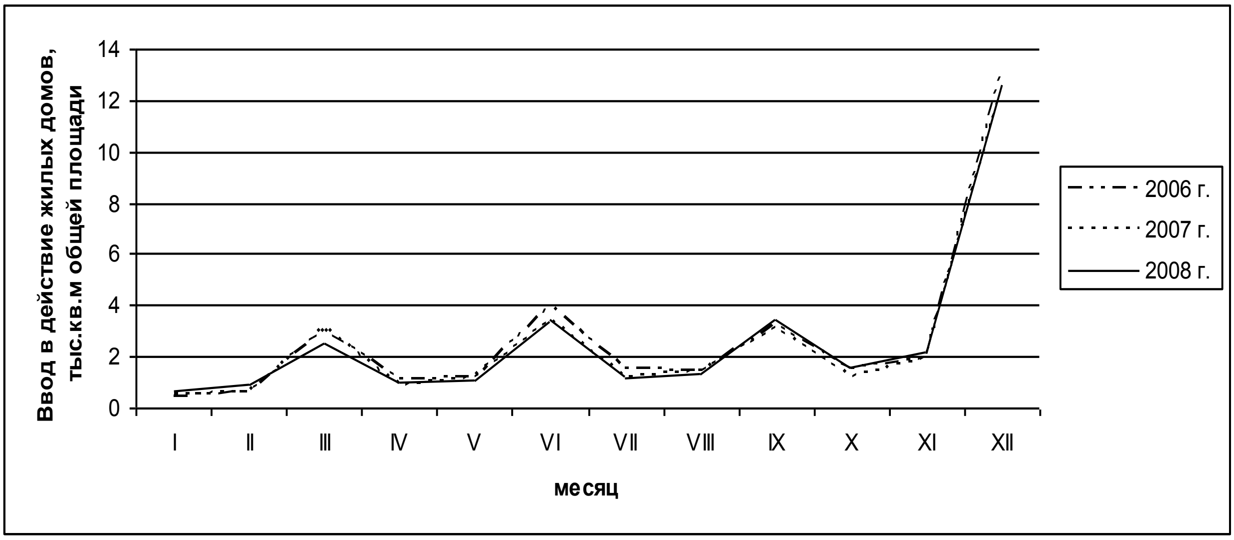


Рис. 10.2. Динамика ввода жилья предприятиями города

Как видно из графика, максимальные и минимальные объемы строительства практически приходятся на одинаковые месяцы.

Поскольку данные от года к году существенно не меняются, индексы сезонности определяются по формуле:

$$i_{c_i} = \frac{\bar{y}_i}{\bar{y}} \cdot 100\% .$$

Сначала определяем средние уровни для каждого месяца:

$$\bar{y}_{янв} = \frac{0{,}4 + 0{,}5 + 0{,}7}{3} = 0{,}5 \text{ тыс.кв.м}$$

$$\bar{y}_{янв} = \frac{0{,}6 + 0{,}6 + 0{,}9}{3} = 0{,}7 \text{ тыс.кв.м}$$

и т.д.

Затем определяется общая средняя за весь период:

$$\bar{y} = \frac{32{,}7 + 30{,}7 + 32{,}0}{36} = 2{,}7 \text{ тыс.кв.м}$$

И, наконец, рассчитываются индексы сезонности:

для января: $\frac{0{,}5}{2{,}7} \cdot 100\% = 18{,}5\%;$

для февраля: $\frac{0{,}7}{2{,}7} \cdot 100\% = 25{,}9\%$ и т.д.

Графическое изображение сезонной волны показано на рис. 10.3.

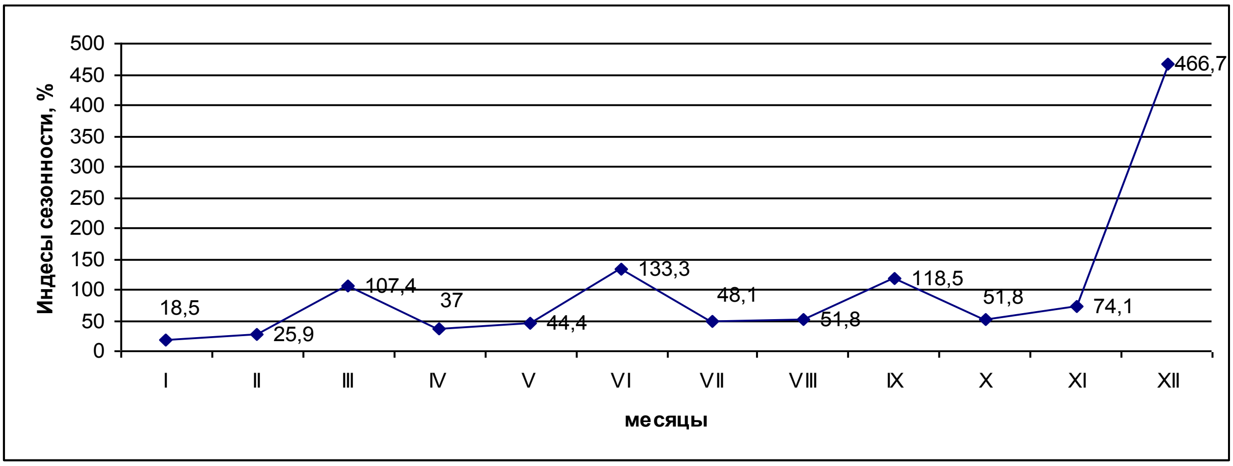


Рис. 10.3. Сезонность ввода в действие жилых домов

Сезонную волну также можно изобразить с помощью радиальной замкнутой диаграммы (гл. 4).

Вопросы для самоконтроля

1. Что такое ряд динамики, уровни ряда?
2. Какие динамические ряды называют моментными и почему нельзя суммировать их уровни? Приведите примеры.
3. Какие ряды динамики называют интервальными? Почему их уровни можно суммировать? Приведите примеры.
4. Назовите основные причины несопоставимости временных рядов.
5. Назовите показатели рядов динамики, которые применяются для оценки свойств динамических рядов?
6. Какие показатели называют базисными и цепными?
7. Что характеризует показатель абсолютного прироста и как он рассчитывается?
8. Что представляет собой темп роста? Как он рассчитывается?
9. Что характеризует темп прироста и как он рассчитывается?
10. Как вычисляется средняя величина уровней в интервальных рядах?
11. Как рассчитывается средняя хронологическая для моментных рядов динамики?
12. Как вычисляется средний абсолютный прирост? Средний темп роста?
13. Что характеризует средний темп прироста и как он рассчитывается?
14. Для чего определяют общую тенденция развития (тренд)?
15. Какие существуют способы и методы выравнивания рядов динамики?
16. Как проводится выравнивание рядов динамики способом скользящей средней?
17. В чем заключается суть метода аналитического выравнивания временного ряда?
18. Охарактеризуйте технику выравнивания ряда динамики по аналитическому методу.
19. Что представляют собой сезонные колебания, в чем практическое значение их изучения?
20. Как рассчитываются индексы сезонности?

11. Индексы и их использование в экономических исследованиях

10.1. *Основные понятия*

Индекс — это относительная величина, которая характеризует изменение уровней социально-экономического явления во времени (динамический индекс) в пространстве (территориальный индекс) или по сравнению с планом (индексы плана и перевыполнения плана).

Индекс является результатом сравнения двух одноименных показателей, поэтому индексы исчисляются только в процентах или коэффициентах.

Показатель, с которым производится сравнение (знаменатель индекса) называется **базисным**, а показатель, который сравнивают — **текущим** или **отчетным**.

Данные базисного периода помечают подстрочным знаком «0», а текущего — «1».

Индексный метод имеет свою терминологию и символику. Обычно для обозначения индексируемых величин используют следующие условные обозначения:

q — количество продукции одного вида в натуральном выражении;

p — цена за единицу продукции;

z — себестоимость единицы продукции;

t — затраты труда (рабочего времени) на производство единицы продукции;

qp — общая стоимость проданного товара (товарооборот);

qz — общая себестоимость (затраты на производство) продукции;

qt — общие расходы рабочего времени на выпуск продукции.

Классификация индексов

1. **По степени охвата единиц совокупности** различают **индивидуальные** и **общие (сводные)** индексы.

Индивидуальные индексы (*i*) характеризуют изменение одного элемента совокупности.

Общие (сводные) индексы (*I*) характеризуют изменение сложного явления в целом. Если индексы охватывают не все явления, а какую-то часть, то они называются **групповыми**.

Общие индексы в зависимости от способа расчета могут быть **агрегатными** или **средними взвешенными** (арифметическими и гармоническими).

2. **В зависимости от содержания и характера изучаемых показателей** различают индексы **количественных (объемных)** и **качественных** показателей.

К индексам **объемных** показателей относятся индексы, показатели которых выражены абсолютными величинами: индексы физического объема производства продукции; индексы товарных запасов; индексы физического объема потребления продукции; национального дохода и др.

К индексам **качественных** показателей относятся индексы цен, себестоимости, средней заработной платы, производительности труда. Качественный показатель характеризует уровень изучаемого явления в расчете на количественную единицу (например, средняя заработная плата определяется делением фонда заработной платы на численность работников, средняя цена — делением выручки от продажи товара на количество проданного товара, себестоимость — делением общих затрат на производство продукции на объем выпущенной продукции и т.д.).

11.2. *Индивидуальные индексы*

Индивидуальные индексы характеризуют изменение отдельных единиц (элементов) совокупности.

Так, индивидуальный индекс физического объема продукции выразится как $i_{q1/0}=\dfrac{q_1}{q_0}$, индекс цен — $i_{p1/0}=\dfrac{p_1}{p_0}$, индекс себестоимости — $i_{z1/0}=\dfrac{z_1}{z_0}$, индекс трудоемкости — $i_{t1/0}=\dfrac{t_1}{t_0}$, индекс товарооборота — $i_{qp1/0}=\dfrac{q_1p_1}{q_0p_0}$ и т.д.

Пример. Рассчитайте индивидуальные индексы цен, физического объема продукции и товарооборота для данных, представленных в табл. 11.1.

Таблица 11.1

Товар	Цена за единицу товара, руб.		Количество проданного товара		Товарооборот, тыс.руб.		Индивидуальные индексы, %		
	май p_0	июнь p_1	май q_0	июнь q_1	май p_0q_0	июнь p_1q_1	цены, i_p	объема, i_q	Товарооборота, i_{qp}
Товар А, ед.	100	105	300	450	30,00	47,25	105	150	157,5
Товар Б, кг	80	86	800	720	64,00	61,92	107,5	90	96,7

Индекс цены: для товара А: $i_p = \frac{p_1}{p_0} = \frac{105}{100} = 1{,}05$ (или 105%)

для товара Б: $i_p = \frac{86}{80} = 1{,}075$ (или 107,5%).

Таким образом, цены на товар А выросли в июне на 5%, на товар Б — на 7,5%.

Индекс физического объема: для товара А: $i_q = \frac{q_1}{q_0} = \frac{450}{300} = 1{,}5$ (или 150%)

для товара Б: $i_q = \frac{720}{800} = 0{,}9$ (или 90%).

Объем продаж товара А увеличился в 1,5 раза, а товара Б сократился и составил 90% от уровня предыдущего месяца.

Индекс товарооборота: для товара А: $i_{qp} = \frac{q_1 p_1}{q_0 p_0} = \frac{47{,}25}{30{,}00} = 1{,}575\,(157{,}5\%)$

для товара Б: $i_{qp} = \frac{61{,}92}{64{,}00} = 0{,}967 \quad (96{,}7\%)$.

Рост цены и объема продаж товара А привели к увеличению выручки на 57,5% по сравнению с маем. Сокращение объема продаж товара Б в отчетном периоде на 10% обусловило снижение выручки магазина на 3,3% (100,0% – 96,7%), несмотря на то, что цена на этот товар в июне выросла на 7,5%.

11.3. *Общие индексы*

Общие (сводные) индексы используют в тех случаях, когда необходимо оценить изменение сложного явления, элементы которого не поддаются непосредственному суммированию.

В общих индексах имеются две величины: одну, изменение которой изучают, называют **индексируемой** величиной; вторую, постоянную в общих индексах, которая приводит разнородные элементы совокупности к сопоставимому виду называют **соизмерителем** или **весом** индексируемой величины.

11.3.1. Общие индексы количественных показателей

Разберем построение общего агрегатного индекса на примере индекса физического объема.

Так как мы строим индекс объема, значит индексируемой, то есть изменяемой, величиной в формуле, будет объем (q).

При разнородной по составу совокупности, когда невозможно непосредственно сложить единицы, чтобы узнать общее количество (например, разный ассортимент продукции в магазине), или в случае, когда единицы совокупности имеют различный удельный вес, необходимо выразить объемные показатели в какой-либо единой форме для обеспечения их сопоставимости. Для этого чаще всего используют стоимостную форму, то есть объем (количество) умножают на цену единицы продукции. Таким образом, в качестве соизмерителя (веса) индексируемой величины выступает цена (p).

Для исключения влияния цен, которые тоже могут меняться, их фиксируют на уровне какого-либо периода: базисного или отчетного.

В зависимости от того, веса какого периода берутся за базу сравнения, различают **индексы Ласпейреса** (веса берутся по **базисному** периоду) и **индексы Пааше** (веса берут по **текущему** периоду).

Таким образом, **общий агрегатный индекс физического объема продукции** можно записать:

индекс Ласпейреса:

$$I_{q1/0} = \frac{\sum q_1 p_0}{\sum q_0 p_0} = \frac{q_1^А p_0^А + q_1^Б p_0^Б + q_1^В p_0^В}{q_0^А p_0^А + q_0^Б p_0^Б + q_0^В p_0^В};$$

или

индекс Пааше:

$$I_{q1/0} = \frac{\sum q_1 p_1}{\sum q_0 p_1} = \frac{q_1^А p_1^А + q_1^Б p_1^Б + q_1^В p_1^В}{q_0^А p_1^А + q_0^Б p_1^Б + q_0^В p_1^В},$$

где $q^А, q^Б, q^В$ — количество продукции каждого вида в натуральном выражении;

$p^А, p^Б, p^В$ — цена за единицу продукции каждого вида.

Аналогично строятся общие агрегатные индексы объема, когда в качестве соизмерителя (веса) индексируемой величины выступает себестоимость единицы продукции или затраты рабочего времени на единицу продукции:

Индексы Ласпейреса:

$$I_{q1/0} = \frac{\sum q_1 z_0}{\sum q_0 z_0}$$

$$I_{q1/0} = \frac{\sum q_1 t_0}{\sum q_0 t_0}$$

Индексы Пааше:

$$I_{q1/0} = \frac{\sum q_1 z_1}{\sum q_0 z_1}$$

$$I_{q1/0} = \frac{\sum q_1 t_1}{\sum q_0 t_1}$$

Общий агрегатный индекс стоимости продукции (товарооборота):

$$I_{q1/0} = \frac{\sum q_1 p_1}{\sum q_0 p_0}.$$

В этом случае индексируемой величиной является товарооборот qp.

Общий агрегатный индекс затрат на выпуск всей продукции:

$$I_{q1/0} = \frac{\sum q_1 z_1}{\sum q_0 z_0}.$$

В данном случае индексируемая величина — общая стоимость затрат на производство продукции — qz.

Средневзвешенные индексы количественных показателей

Средние взвешенные индексы физического объема продукции используются в тех случаях, когда известны индивидуальные индексы объема по отдельным видам продукции и стоимость отдельных видов продукции (или затраты на отдельные виды продукции) в базисном или отчетном периоде.

Средний взвешенный арифметический индекс физического объема продукции:

$$I_{q1/0} = \frac{\sum i_q q_0 p_0}{\sum q_0 p_0} = \frac{i_q^A q_0^A p_0^A + i_q^Б q_0^Б p_0^Б + i_q^В q_0^В p_0^В}{q_0^A p_0^A + q_0^Б p_0^Б + q_0^В p_0^В},$$

где i_q — индивидуальный индекс по каждому виду продукции.

$q_0 p_0$ — стоимость продукции каждого вида в базисном периоде.

Средний взвешенный гармонический индекс физического объема продукции.

$$I_{q1/0} = \frac{\sum i_q q_1 p_1}{\sum \frac{1}{i_q} q_1 p_1} = \frac{q_1^A p_1^A + q_1^Б p_1^Б + q_1^В p_1^В}{\frac{1}{i_q^A} q_1^A p_1^A + \frac{1}{i_q^Б} q_1^Б p_1^Б + \frac{1}{i_q^В} q_1^В p_1^В}$$

где $q_1 p_1$ — стоимость продукции каждого вида в текущем периоде.

Общий территориальный индекс физического объема продукции:

$$I_{q\,А/Б} = \frac{\sum q_А \bar{p}}{\sum q_Б \bar{p}},$$

где $q_А$, $q_Б$ — количество выпущенной (реализованной) продукции каждого вида в натуральном выражении соответственно территории А и Б;

$\bar{p}$ — средняя цена каждого вида продукции по сравниваемым территориям, определяемая как средняя взвешенная арифметическая.

11.3.2. Общие индексы качественных показателей

Агрегатные индексы качественных показателей строятся аналогично агрегатным индексам количественных показателей, только в данном случае в качестве соизмерителей или весов выступает объем продукции в натуральном выражении (q).

Общий агрегатный индекс цен:

Индекс Ласпейреса:

$$I_{p1/0} = \frac{\sum p_1 q_0}{\sum p_0 q_0}$$

Индекс Пааше:

$$I_{p1/0} = \frac{\sum p_1 q_1}{\sum p_0 q_1}$$

В экономической практике агрегатным индексам цены отводится особая роль. С их помощью отслеживается динамика инфляционных процессов в народном хозяйстве и осуществляется пересчет важнейших показателей национальных счетов из фактических цен в сопоставимые.

Эти индексы не идентичны и при одинаковых исходных данных не совпадают, так как имеют различное экономическое содержание.

Индекс цен Ласпейреса показывает, во сколько бы раз товары базисного периода подорожали (подешевели) из-за изменения цен на них в отчетный период.

На основе этого индекса целесообразно определять **индекс покупательной способности рубля**:

$$I_{покуп.спос.} = \frac{1}{I_p^Л}.$$

Индекс Ласпейреса используют при прогнозировании объема товарооборота в связи с вероятным изменением цен на товары в будущем периоде.

Индекс цен Пааше дает ответ на вопрос: на сколько товары в текущем периоде стали дороже (дешевле), чем в базисном.

Разница между числителем и знаменателем индекса Паше $\left(\Delta^{p} = \sum p_1 q_1 - \sum p_0 q_1\right)$ показывает абсолютное изменение товарооборота (стоимости продукции) за счет изменения цен.

Индекс Пааше применяют при изучении отчетных данных, когда целью является качественная оценка изменения товарооборота в результате изменения цен в отчетном периоде.

Средние взвешенные индексы цен применяются в том случае, если известны индивидуальные индексы цен по отдельным видам продукции, а также стоимость отдельных видов продукции.

Средний взвешенный арифметический индекс цен:

$$I_{p1/0} = \frac{\sum i_p p_0 q_0}{\sum p_0 q_0},$$

где $p_0 q_0$ — стоимость продукции каждого вида в базисном периоде.

Средний взвешенный гармонический индекс цен:

$$I_{p1/0} = \frac{\sum p_1 q_1}{\sum \frac{1}{i_p} p_1 q_1},$$

где i_p — индивидуальный индекс цен по каждому виду продукции;

$p_1 q_1$ — стоимость продукции каждого вида в текущем периоде.

Общий территориальный индекс цен

$$I_{p A/Б} = \frac{\sum p_A (q_A + q_Б)}{\sum p_Б (q_A + q_Б)},$$

где p_A, $p_Б$ — цена за единицу продукции каждого вида соответственно на территории А и Б;

$q_A + q_Б$ — суммарный объем продукции двух территорий.

Общие агрегатные индексы себестоимости и затрат рабочего времени строятся аналогично индексам цен Пааше:

$$I_{z1/0} = \frac{\sum z_1 q_1}{\sum z_0 q_1} \qquad I_{t1/0} = \frac{\sum t_1 q_1}{\sum t_0 q_1}$$

11.4. *Построение системы индексов*

В случаях, когда необходимо проследить изменение показателей не за два, а за более продолжительные промежутки времени, встает вопрос о выборе базы сравнения и весов индексов.

Системой индексов называется ряд последовательно построенных индексов.

В зависимости от базы сравнения системы индексов бывают **базисными** и **цепными**.

Система базисных индексов — это сравнение всех индексируемых величин с какой-то одной величиной, принятой за базу сравнения (аналогично расчету базисных коэффициентов роста в рядах динамики). Обычно за базу принимается начальный период — «0».

Пример: базисные индивидуальные индексы физического объема продукции:

$$i_{q\,1/0} = \frac{q_1}{q_0}; \quad i_{q\,2/0} = \frac{q_2}{q_0}; \quad i_{q\,3/0} = \frac{q_3}{q_0} \text{ и т.д.}$$

Система цепных индексов — это сравнение индексируемой величины с величиной предшествующего периода (аналогично расчету цепных коэффициентов роста в рядах динамики).

Пример: цепные индивидуальные индексы физического объема продукции:

$$i_{q\,1/0} = \frac{q_1}{q_0}; \quad i_{q\,2/0} = \frac{q_2}{q_1}; \quad i_{q\,3/0} = \frac{q_3}{q_2} \text{ и т.д.}$$

Системы общих индексов (базисных и цепных) можно строить с **постоянными** или **переменными весами**. В качестве постоянного веса обычно берется вес начального периода, а в качестве переменного — в каждом индексе берется вес текущего периода.

Пример: Системы базисных индексов физического объема с постоянными и переменными весами:

Постоянные веса:

$$I_{q1/0} = \frac{\sum q_1 p_0}{\sum q_0 p_0}$$

$$I_{q2/0} = \frac{\sum q_2 p_0}{\sum q_0 p_0}$$

$$I_{q3/0} = \frac{\sum q_3 p_0}{\sum q_0 p_0}$$

и т.д.

Переменные веса:

$$I_{q1/0} = \frac{\sum q_1 p_1}{\sum q_0 p_1}$$

$$I_{q2/0} = \frac{\sum q_2 p_2}{\sum q_0 p_2}$$

$$I_{q3/0} = \frac{\sum q_3 p_3}{\sum q_0 p_3}$$

и т.д.

Системы цепных индексов физического объема с постоянными и переменными весами:

Постоянные веса:

$$I_{q1/0} = \frac{\sum q_1 p_0}{\sum q_0 p_0}$$

$$I_{q2/1} = \frac{\sum q_2 p_0}{\sum q_1 p_0}$$

$$I_{q3/2} = \frac{\sum q_3 p_0}{\sum q_2 p_0}$$

и т.д.

Переменные веса:

$$I_{q1/0} = \frac{\sum q_1 p_1}{\sum q_0 p_1}$$

$$I_{q2/1} = \frac{\sum q_2 p_2}{\sum q_1 p_2}$$

$$I_{q3/2} = \frac{\sum q_3 p_3}{\sum q_2 p_3}$$

и т.д.

Системы других общих индексов (цены; себестоимости и пр.) строятся аналогично.

11.5. *Изучение динамики средних величин*

Часто при помощи индексов изучают динамику средних показателей: средней себестоимости, средней заработной платы, средней дальности перевозок, средней производительности труда и т.п.

Изменение средней величины того или иного показателя зависит от двух факторов:

– от изменения значения каждой отдельной единицы изучаемого явления;

– от изменения структуры явления, то есть доли отдельных групп единиц совокупности.

Для оценки силы влияния каждого из этих факторов используют систему взаимосвязанных индексов: индекса переменного состава, индекса постоянного (фиксированного) состава и индекса влияния структурных сдвигов.

Индекс переменного состава показывает относительное изменение средней величины в целом, за счет обоих факторов:

$$I_{пер} = \frac{\bar{x}_1}{x_0} = \frac{\sum x_1 d_1}{\sum d_1} : \frac{\sum x_0 d_0}{\sum d_0} = \frac{\sum x_1 d_1}{\sum x_0 d_0},$$

где $x_{0,}\, x_1$ — отдельные значения признака, соответственно в базисном и текущем периодах;

$d_{0,}\, d_1$ — доля (удельный вес) отдельных значений признака в общем объеме, соответственно в базисном и текущем периодах

$$\sum d_0 = 1; \quad \sum d_1 = 1.$$

Разница между числителем и знаменателем индекса показывает абсолютное изменение средней величины под действием обоих факторов:

$$\Delta_{пер} = \bar{x}_1 - \bar{x}_0 = \sum x_1 d_1 - \sum x_0 d_0 .$$

Индекс постоянного состава показывает, как изменилась средняя величина $\bar{x}$ только за счет изменения индексируемой величины, то есть изменения значений отдельных единиц совокупности:

$$I_{пост} = \frac{\sum x_1 d_1}{\sum x_0 d_1}.$$

Разница между числителем и знаменателем индекса показывает абсолютное изменение средней под действием этого фактора:

$$\Delta_{пост} = \sum x_1 d_1 - \sum x_0 d_1 .$$

Индекс влияния структурных сдвигов показывает, какое влияние на динамику средней $\bar{x}$ оказали изменения в структуре совокупности при неизменных значениях отдельных единиц совокупности:

$$I_{стр} = \frac{\sum x_0 d_1}{\sum x_0 d_0}.$$

Разница между числителем и знаменателем индекса показывает абсолютное изменение средней величины только за счет структурных сдвигов:

$$\Delta_{стр} = \sum x_0 d_1 - \sum x_0 d_0.$$

Взаимосвязь между перечисленными индексами можно представить следующими равенствами:

$$I_{пер} = I_{пост} \cdot I_{стр}$$
$$\Delta_{пер} = \Delta_{пост} + \Delta_{стр}.$$

Пример. Имеются следующие данные о выпуске и стоимости одного и того же продукта по трем предприятиям (табл. 11.2).

Необходимо определить изменение средней себестоимости единицы продукции в абсолютных и относительных величинах в целом по предприятиям и за счет отдельных факторов:

1) изменения себестоимости выпуска на отдельных предприятиях;
2) структурных изменений в производстве продукции.

Таблица 11.2

Количество выпущенной продукции и себестоимость единицы продукции одного вида по трем предприятиям отрасли

Предприятия	Базисный период		Отчетный период	
	Себестоимость единицы, руб. (z_0)	Выпуск продкции, тыс.ед. (q_0)	Себестоимость единицы, руб. (z_1)	Выпуск продкции, тыс.ед. (q_1)
1	15,0	10	14,2	10
2	13,0	6	12,5	7
3	10,0	4	9,5	8
Итого	**—**	**20**	**—**	**25**

Определим удельный вес продукции каждого предприятия в общем объеме выпуска и составим вспомогательную таблицу для расчета индексов (табл. 11.3).

Таблица 11.3

Вспомогательная таблица

Предприятие	Базисный период			Отчетный период			z_0d_1
	z_0	удельный вес предприятия в общем объеме выпуска продукции, d_0	z_0d_0	z_1	удельный вес предприятия в общем объеме выпуска продукции, d_1	z_1d_1	
1	15,0	0,50	7,50	14,2	0,40	5,68	6,00
2	13,0	0,30	3,90	12,5	0,28	3,50	3,64
3	10,0	0,20	2,00	9,5	0,32	3,04	3,20
Всего	—	**1,00**	**13,40**	—	**1,00**	**12,22**	**12,84**

Рассчитаем среднюю себестоимость единицы продукции по трем предприятиям в отчетном и базисном периодах:

$$\bar{z}_1 = \sum z_1 d_1 = 12{,}22 \text{ руб./ед.};$$

$$\bar{z}_0 = \sum z_0 d_0 = 13{,}40 \text{ руб./ед.}$$

Вычислим индекс переменного состава:

$$I_{пер} = \frac{\sum z_1 d_1}{\sum z_0 d_0} = \frac{12{,}22}{13{,}40} = 0{,}912 \text{ (или 91,2\%).}$$

Абсолютное изменение средней себестоимости:

$$\Delta_{пер} = 12{,}22 - 13{,}40 = -1{,}18 \text{ руб./ед.}$$

Таким образом, средняя себестоимость единицы изделия снизилась в отчетном периоде на 8,8% или на 1,18 руб.

На изменение средней себестоимости оказали влияние два фактора:

1) изменение себестоимости производства на каждом предприятии и
2) изменение удельного веса каждого предприятия в общем выпуске продукции.

Влияние первого фактора оценим с помощью индекса постоянного состава:

$$I_{пост} = \frac{\sum z_1 d_1}{\sum z_0 d_1} = \frac{12{,}22}{12{,}84} = 0{,}952 \text{ (или 95,2\%)}$$

$$\Delta_{пост} = 12{,}22 - 12{,}84 = -0{,}62 \text{ руб./ед.}$$

Следовательно, только за счет изменения себестоимости продукции на отдельных предприятиях средняя себестоимость снизилась на 0,62 руб./ед. или на 4,8% (100% – 95,2%).

Влияние второго фактора оценим с помощью индекса влияния структурных сдвигов:

$$I_{стр} = \frac{\sum z_0 d_1}{\sum z_0 d_0} = \frac{12{,}84}{13{,}40} = 0{,}958 \text{ (или 95,8\%)}$$

$$\Delta_{стр} = 12{,}84 - 13{,}40 = -0{,}56 \text{ руб./ед.}$$

Таким образом, за счет изменения удельного веса отдельных предприятий в общем выпуске продукции, средняя себестоимость снизилась на 0,56 руб./ед. или на 4,2% (100% – 95,8%). В частности, снижению средней себестоимости способствовало увеличение доли продукции 3-го предприятия, у которого самая низкая себестоимость (с 20% до 32%), что указывает на позитивные изменения в структуре производства.

Проверка: 0,912 = 0,952 · 0,958

– 1,18 = (–0,62) + (–0,56).

11.6. *Использование индексного метода в анализе взаимосвязей экономических явлений*

Индексный метод используется при изучении роли отдельных факторов в динамике какого-либо сложного явления, позволяя определить размер абсолютного изменения сложного явления за счет каждого фактора в отдельности.

Роль отдельных факторов в изменении результативного показателя оценивается путем построения системы взаимосвязанных индексов.

В экономической модели, где результат равен произведению составляющих показателей, абсолютное изменение итогового показателя за счет каждого фактора в отдельности оценивается путем построения системы взаимосвязанных индексов.

Предположим, что сложный результативный показатель F равен: $F = a \cdot b \cdot c$, где a, b, c — показатели-факторы.

Изменение сложного явления может быть представлено индексом:

$$I_F = \frac{F_1}{F_0} = \frac{a_1 \cdot b_1 \cdot c_1}{a_0 \cdot b_0 \cdot c_0} = I_a \cdot I_b \cdot I_c .$$

Абсолютное изменение явления F под влиянием всех факторов представляет собой разность между числителем и знаменателем индекса:

$$\Delta_F = F_1 - F_0 = a_1 \cdot b_1 \cdot c_1 - a_0 \cdot b_0 \cdot c_0 .$$

Для выявления влияния каждого фактора в отдельности на изменение сложного показателя, индекс сложного показателя разлагают на частные (факторные) индексы, характеризующие роль каждого из них.

Обычно применяются следующие методы разложения общего индекса на частные:

1) **метод обособленного изучения факторов;**
2) **последовательно-цепной метод.**

В основе этих методов лежит принцип элиминирования (т.е. исключения) изменений величины всех факторов, кроме изучаемого.

Сущность **метода обособленного изучения факторов** заключается в том, что при выявлении влияния отдельного фактора сложный показатель берется в том виде, который бы он имел, если бы изменился только один данный фактор, а все прочие остались бы неизменными на уровне *базисного* периода. Тогда изменение результирующего показателя F за счет каждого фактора в отдельности определяется по формулам:

фактора a: $$I_F^a = \frac{a_1 \cdot b_0 \cdot c_0}{a_0 \cdot b_0 \cdot c_0};$$

фактора b: $I_F^b = \dfrac{a_1 \cdot b_1 \cdot c_0}{a_1 \cdot b_0 \cdot c_0}$;

фактора c: $I_F^c = \dfrac{a_1 \cdot b_1 \cdot c_1}{a_1 \cdot b_1 \cdot c_0}$.

Абсолютное изменение результативного показателя за счет каждого фактора получается как разность между числителем и знаменателем индекса:

$$\Delta_F^a = a_1 \cdot b_0 \cdot c_0 - a_0 \cdot b_0 \cdot c_0$$

$$\Delta_F^b = a_1 \cdot b_1 \cdot c_0 - a_1 \cdot b_0 \cdot c_0$$

$$\Delta_F^c = a_1 \cdot b_1 \cdot c_1 - a_1 \cdot b_1 \cdot c_0.$$

Недостатком этого метода является то, что факторные индексы при данном методе не разлагают полностью величины абсолютного изменения результативного показателя. Получается некоторый неразложенный остаток, который следует рассматривать как результат совместного действия факторов:

$$F_1 - F_0 \neq \Delta_F^a + \Delta_F^b + \Delta_F^c$$

При **последовательно-цепном методе** необходимо соблюдать определенную последовательность в расположении факторов:

1) на первом месте в модели следует ставить качественный фактор;
2) увеличение цепи факторов на один фактор (например, $a \cdot b$) каждый раз должно давать показатель, имеющий реальный экономический смысл.

Сущность метода заключается в следующем: при определении влияния первого фактора все остальные факторы сохраняются в числителе и знаменателе на уровне текущего (отчетного) периода; при построении второго факторного индекса первый фактор берется на уровне базисного периода, а все последующие остаются на уровне текущего, при построении третьего факторного индекса первый и второй факторы берутся на уровне базисного периода, а четвертый и последующие — на уровне текущего и т.д.

Таким образом, частные индексы будут выглядеть так:

фактора a: $I_F^a = \dfrac{a_1 \cdot b_1 \cdot c_1}{a_0 \cdot b_1 \cdot c_1}$;

фактора b: $I_F^b = \dfrac{a_0 \cdot b_1 \cdot c_1}{a_0 \cdot b_0 \cdot c_1}$;

фактора c: $I_F^c = \dfrac{a_0 \cdot b_0 \cdot c_1}{a_0 \cdot b_0 \cdot c_0}$.

Абсолютное изменение F за счет каждого фактора:

фактора a: $\Delta_F^a = a_1 \cdot b_1 \cdot c_1 - a_0 \cdot b_1 \cdot c_1 = (a_1 - a_0) b_1 c_1$

фактора b: $\Delta_F^b = a_0 \cdot (b_1 - b_0) c_1$

фактора c: $\Delta_F^c = a_0 \cdot b_0 (c_1 - c_0)$.

Таким образом: $F_1 - F_0 = \Delta_F = \Delta_F^a + \Delta_F^b + \Delta_F^c$.

Результаты расчетов, выполненных различными методами будут несколько отличаться друг от друга, так как любой метод предполагает наличие определенных допущений. Поэтому, если необходимо провести анализ за несколько лет, то для получения сопоставимых результатов, необходимо использовать один и тот же метод.

Абсолютное изменение сложного экономического показателя за счет каждого фактора может быть определено и в том случае, если этот показатель представляет собой сумму произведений, определяющих его величину показателей. К числу таких показателей относятся: общая стоимость всей выработанной (или реализованной) продукции, общая сумма затрат на ее производство, общая сумма затрат труда на производство всей продукции.

Пример. Общий агрегатный индекс стоимости продукции (I_{qp}) можно представить следующей формулой:

$$I_{qp1/0} = \frac{\sum q_1 p_1}{\sum q_0 p_0}.$$

Тогда, общее абсолютное изменение стоимости продукции за счет двух факторов составляет:

$$\Delta^{qp}_{\sum qp} = \sum q_1 p_1 - \sum q_0 p_0 .$$

Абсолютное изменение общей стоимости продукции за счет отдельных факторов:

а) изменение физического объема продукции

$$\Delta^{q}_{\sum qp} = \sum q_1 p_0 - \sum q_0 p_0$$

б) среднего изменения цен на продукцию

$$\Delta^{p}_{\sum qp} = \sum q_1 p_1 - \sum q_1 p_0 .$$

В том случае, если влияние обоих факторов *однонаправлено*, можно оценить долю каждого фактора в общем абсолютном изменении результативного показателя следующим образом:

а) физического объема продукции:

$$\frac{\Delta^{q}_{\sum qp}}{\Delta^{qp}_{\sum qp}} (\cdot 100\%)$$

б) среднего изменения цен на продукцию:

$$\frac{\Delta^{p}_{\sum qp}}{\Delta^{qp}_{\sum qp}} (\cdot 100\%) .$$

Влияние отдельных факторов на абсолютное изменение общих затрат на выпуск продукции и на абсолютное изменение общего объема затрат рабочего времени выполняется аналогично.

Вопросы для самоконтроля

1. Что в статистике называется индексом?
2. Какие условные обозначения используют в индексном методе?
3. По каким признакам классифицируют индексы?
4. Что характеризуют индивидуальные индексы? Приведите примеры.

5. Что характеризуют общие индексы? Приведите примеры.
6. Какая величина в расчете индексов называется индексируемой величиной? Соизмерителем?
7. В чем заключается методологическая суть построения общих индексов агрегатной формы?
8. Как строятся индексы Ласпейреса и Пааше?
9. Приведите примеры общих агрегатных индексов количественных показателей.
10. В каких случаях используют средневзвешенные индексы? Какие формы средневзвешенных индексов используются?
11. Какой общий индекс называют среднеарифметическим и как он рассчитывается?
12. В каких случаях используется общий индекс цен Ласпейреса и как он рассчитывается?
13. В каких случаях используется общий индекс цен Пааше и как он рассчитывается?
14. Какой общий индекс называют среднегармоническим и как он рассчитывается?
15. Приведите примеры систем индексов.
16. В каких случаях используют индексы переменного состава, постоянного состава и структурных сдвигов?
17. Объясните суть индекса переменного состава на примере индекса себестоимости.
18. Объясните суть индекса постоянного состава на примере индекса себестоимости.
19. Объясните суть индекса структурных сдвигов на примере индекса себестоимости.
20. Какой зависимостью связаны индексы переменного, постоянного состава и структурных сдвигов?
21. В чем заключается суть последовательно-цепного индексного метода?

Приложение

Значения α-процентных пределов $t_{a,k}$ в зависимости

от k степеней свободы и заданного уровня значимости a

для распределения Стьюдента

k \ α	10,0	5,0	2,5	2,0	1,0	0,5	0,3	0,2	0,1
1	6,314	12,706	25,452	31,821	63,657	127,3	212,2	318,3	636,6
2	2,920	4,303	6,205	6,965	9,925	14,089	18,216	22,327	31,600
3	2,353	3,182	4,177	4,541	5,841	7,453	8,891	10,214	12,922
4	2,132	2,776	3,495	3,747	4,604	5,597	6,435	7,173	8,610
5	2,015	2,571	3,163	3,365	4,032	4,773	5,376	5,893	6,869
6	1,943	2,447	2,969	3,143	3,707	4,317	4,800	5,208	5,959
7	1,895	2,365	2,841	2,998	3,499	4,029	4,442	4,785	5,408
8	1,860	2,306	2,752	2,696	3,355	3,833	4,199	4,501	5,041
9	1,833	2,262	2,685	2,821	3,250	3,690	4,024	4,297	4,781
10	1,812	2,228	2,634	2,764	3,169	3,581	3,892	4,144	4,587
12	1,782	2,179	2,560	2,681	3,055	3,428	3,706	3,930	4,318
14	1,761	2,145	2,510	2,624	2,977	3,326	3,583	3,787	4,140
16	1,746	2,120	2,473	2,583	2,921	3,252	3,494	3,686	4,015
18	1,734	2,101	2,445	2,552	2,878	3,193	8,428	3,610	3,922
20	1,725	2,086	2,423	2,528	2,845	3,153	3,376	3,552	3,849
22	1,717	2,074	2,405	2,508	2,819	3,119	3,335	3,505	3,792
24	1,711	2,064	2,391	2,492	2,797	3,092	3,302	3,467	3,745
26	1,706	2,056	2,379	2,479	2,779	3,067	3,274	3,435	3,704
28	1,701	2,048	2,369	2,467	2,763	3,047	3,250	3,408	3,674
30	1,697	2,042	2,360	2,457	2,750	3,030	3,230	3,386	3,646
∞	1,645	1,960	2,241	2,326	2,576	2,807	2,968	3,090	3,291

РЕКОМЕНДУЕМАЯ ЛИТЕРАТУРА

1. Гришин А.Ф., Кочерова Е.В. Статистические модели: построение, оценка, анализ: Уч. пособие. — М.: Финансы и статистика, 2005.
2. Едронова В.Н., Малафеева М.В. Общая теория статистики: учебник. — 2-е изд., перераб. и доп. — М.: Магистр, 2007.
3. Ефимова М.Р. и др. Общая теория статистики: Учебник / М.Р. Ефимова, Е.В. Петрова, В.Н. Румянцев и др. — М.: Инфра-М, 2004.
4. Ефимова М.Р., Ганченко О.И., Петрова Е.В. Практикум по общей теории статистики: Учеб. пособие. — 2-е изд. перераб. и доп. — М.: Финансы и статистика, 2006.
5. Общая теория статистики: Учебник / Под ред. О.Э. Башиной, А.А. Спирина — 5-е изд., доп. и перераб. — М.: Финансы и статистика, 2004.
6. Статистика. Первая книга: Учеб.-метод. пособие: Пер. с англ.; Под ред. О.Э. Башиной. — М.: Финансы и статистика, 2004.
7. Теория статистики: Учебник / Р.А. Шмойлова, В.Г. Минашкин, Н.А. Садовникова, Е.Б. Шувалова; под ред. Р.А. Шмойловой. — 4-е изд., перераб. и доп. — М.: Финансы и статистика, 2006.

ДЛЯ ЗАМЕТОК

ДЛЯ ЗАМЕТОК

ДЛЯ ЗАМЕТОК