Министерство образования Российской Федерации
Томский государственный педагогический университет

Н.Г. Филонов

СТАТИСТИКА

Учебное пособие

Томск 2004

ББК 60.6 я 73
Ф 55

Печатается по решению
редакционно-издательского совета
Томского государственного
педагогического университета

Ф 55 Филонов Н.Г. Статистика: Учебное пособие. Томск: Центр учебно-методической литературы Томского государственного педагогического университета, 2004. 208 с.

Учебное пособие содержит материалы для студентов по курсу «Статистика». Разделы пособия снабжены большим числом примеров и контрольных вопросов для самопроверки.

Содержание курса рассматривается в теоретико-методологическом аспекте, в соответствии с Государственным образовательным стандартом профессионального высшего образования.

Материал апробирован в работе со студентами факультета психологии и управления ТГПУ. Представляет интерес для инженеров, аспирантов, преподавателей, ученых, занимающихся вопросами статистики. Может быть использовано менеджерами, преподавателями и студентами других экономических специальностей, слушателями школ бизнеса.

Рецензенты: Чирков С.Н., профессор кафедры менеджмента ФПУ ТГПУ
Коханенко А.П., профессор кафедры квантовой электроники и фотоники РФФ ТГУ

ББК 60.6 я 73

ОГЛАВЛЕНИЕ

ВВЕДЕНИЕ

В современном обществе статистика стала одним из важнейших инструментов управления национальной экономикой. Развитие рыночных отношений в стране поставило перед статистикой новую задачу – реформирование общеметодологических и организационных основ статистической теории и практики.

Главной задачей статистики является расчет и анализ статистических показателей, благодаря чему органы власти, управления и различные ведомственные организации получают всестороннюю характеристику управляемых объектов, будь то вся национальная экономика или отдельные ее отрасли, предприятия и их подразделения.

Общая теория статистики разрабатывает общие принципы и методы статистического исследования общественных явлений, наиболее общие категории (показатели) статистики. Она является учебной дисциплиной, формирующей необходимые профессиональные знания у экономистов, менеджеров, руководителей предприятий.

Данное учебное пособие составлено в соответствии с учебной программой курса и предназначено для студентов экономических специальностей высших учебных заведений, а также для практических работников, занимающихся вопросами планирования, учета и анализа производственно-хозяйственной деятельности. Учебным пособием могут воспользоваться лица, самостоятельно изучающие статистику.

Основной целью, которую ставил автор учебного пособия – это изложить учебный материал в доступной для студентов форме, сохраняя, безусловно, научную основу содержания, логику изложения. В учебном пособии практически по каждому вопросу дано большое число примеров и задач, помогающих глубже понять и усвоить материал.

Учебное пособие состоит из тринадцати глав, соответствующих темам курса общей теории статистики. В конце каждой главы приведены вопросы для самопроверки, а в конце пособия – список используемой литературы и приложение.

ГЛАВА 1. ПРЕДМЕТ И МЕТОД СТАТИСТИЧЕСКОЙ НАУКИ

1.1. Понятие о статистике

Статистика (нем. Statistak, итал. Stato, позднелат. Status – государство) – наука, изучающая положение дел в государстве. Иными словами – **общественная наука, изучающая количественную сторону качественно определенных массовых социально-экономических явлений и закономерностей их развития в конкретных условиях места и времени**.

При определении того, что является предметом статистики, подчеркивается несколько характерных особенностей статистики как науки.

Статистика изучает:

- **массовые общественные явления** при помощи **статистических показателей** (численность населения, количественно произведенной в стране конкретной промышленной, сельскохозяйственной, строительной и другой продукции за определенный период) **и их динамику** (изменение уровня жизни населения и т.д.);
- **количественную сторону общественных явлений** и дает количественное, цифровое освещение общественных явлений;
- **количественную сторону общественных явлений** в неразрывной связи с их качественным содержанием; наблюдает в обществе **процесс перехода количественных изменений в качественные** (так, количественные изменения структуры экспорта и импорта товаров свидетельствуют о качественных изменениях в экономике страны);
- **количественную сторону общественных явлений** в конкретных условиях **места и времени** (динамику численности населения и занятости его по секторам экономики, объема производства, распределения доходов, потребления и т.д.); **характеризует явления общественной жизни в конкретных пространственных и временных границах;**
- **количественные связи между общественными явлениями** с помощью специальной методологии; использует математические методы и методы теории вероятности при вычислении ряда статистических показателей (ошибок выборки, тесноты связи и т.д.), в свою очередь гуманитарные и естественные науки широко применяют в своих исследованиях методы статистики для сбора, обработки и анализа данных.

В настоящее время термин **«статистика» употребляется в трех значениях**:

- **отрасль практической деятельности по сбору, обработке, анализу и публикации массовых цифровых данных** о самых различных явлениях и процессах общественной жизни; эту деятельность на профессиональном уровне осуществляет государственная статистика – Государственный комитет по статистике Российской Федерации и система его учреждений, организованных по административно-территориальному признаку, а также ведомственная статистика (на предприятиях, в объединениях, ведомствах, министерствах);
- **совокупность цифровых сведений, статистические данные**, представляемые в отчетности предприятий, организаций, отраслей экономики, а также публикуемые в сборниках, справочниках, периодической прессе, которые являются результатом статистической работы;
- **отрасль общественных наук, специальная научная дисциплина**, изучаемая в высших и средних специальных учебных заведениях.

Статистика как наука представляет собой целостную систему научных дисциплин: **теория статистики, экономическая статистика, социальная статистика**.

1.2. Возникновение и развитие статистики

Статистика имеет многовековую историю. Ее возникновение и развитие обусловлены общественными потребностями: подсчет населения, скота, учета земельных угодий, имущества и т.д. Наиболее ранние сведения о таких работах в Китае относятся к XIII в. до нашей эры. В Древнем Риме проводились учеты свободных граждан и их имущества.

Считается, что основы статистической науки заложены английским экономистом У. Петти (1623–1687). Он рассматривал статистику как науку об управлении. В 1746 г. немецкий профессор философии и права Ахенваль впервые в Марбургском университете начал читать новую дисциплину, названную им **статистикой**.

Во второй половине XIX века и начале XX века происходит интенсивное развитие статистики. Этому способствовало проведение различного рода периодических переписей и обследований. В это же время совершенствуются органы государственной статистики. Формируется специальная научная дисциплина – математическая статистика, являющаяся частью математики.

В развитии статистики видное место принадлежит представителям отечественной науки и практики. В эпоху Петра I статистика трактовалась преимущественно как **описательная наука**. Но уже со второй половины XIX в. и начале XX века выдвигается познавательное значение статистики и выходят в свет научные сочинения, посвященные теоретическим вопросам статистики. Профессор петербургского университета Ю.Э. Янсон (1835–1893) (автор университетского курса «Теория статистики») назвал статистику **общественной наукой**. А.А. Кауфман разрабатывает фундаментальные учебники «Теория и методы статистики», которые служили основным пособием при изучении статистики в вузах. Видный экономист А.И. Чупров (1842–1908) отмечал необходимость массового статистического исследования при помощи метода количественного наблюдения большого числа факторов для того, чтобы описать общественные явления, подметить законы и определить причины, их вызвавшие. Большой вклад в развитие математической статистики был внесен П.Л. Чебышевым и его учениками А.А. Марковым и А.М. Ляпуновым. Развитие статистики в России тесным образом связано с созданной после отмены крепостного права земской статистикой, которая пользовалась заслуженным авторитетом за объективность и профессионализм.

История развития статистики показывает, что **статистическая наука сложилась** в результате теоретического обобщения накопленного человечеством передового опыта учетно-статистических работ, **обусловленных, прежде всего, потребностями управления жизни общества.**

1.3. Органы государственной статистики Российской Федерации

В соответствии **со ст. 71 Конституции Российской Федерации** руководство статистикой в стране осуществляет **Госкомстат** как федеральный орган исполнительной власти.

Госкомстат Российской Федерации, его органы в республиках, краях, областях, автономных областях и округах, в городах Москве и Санкт-Петербурге, других городах и районах, а также подведомственные им организации, учреждения и учебные заведения составляют **единую систему государственной статистики страны**.

Формы и методы сбора и обработки статистических данных, методология расчета статистических показателей, установленные Госкомстатом, являются статистическими стандартами Российской Федерации. В основе деятельности Госкомстата лежат **федеральные статистические программы**, которые **финансируются из госбюджета**. Все остальные исследования ведутся за счет средств заказчиков.

Положение о Государственном комитете Российской Федерации по статистике (Госкомстате России) было утверждено Постановлением №834 правительства от 9 июля 1994 г. Структура Госкомстата РФ представлена на рис. 1.1.

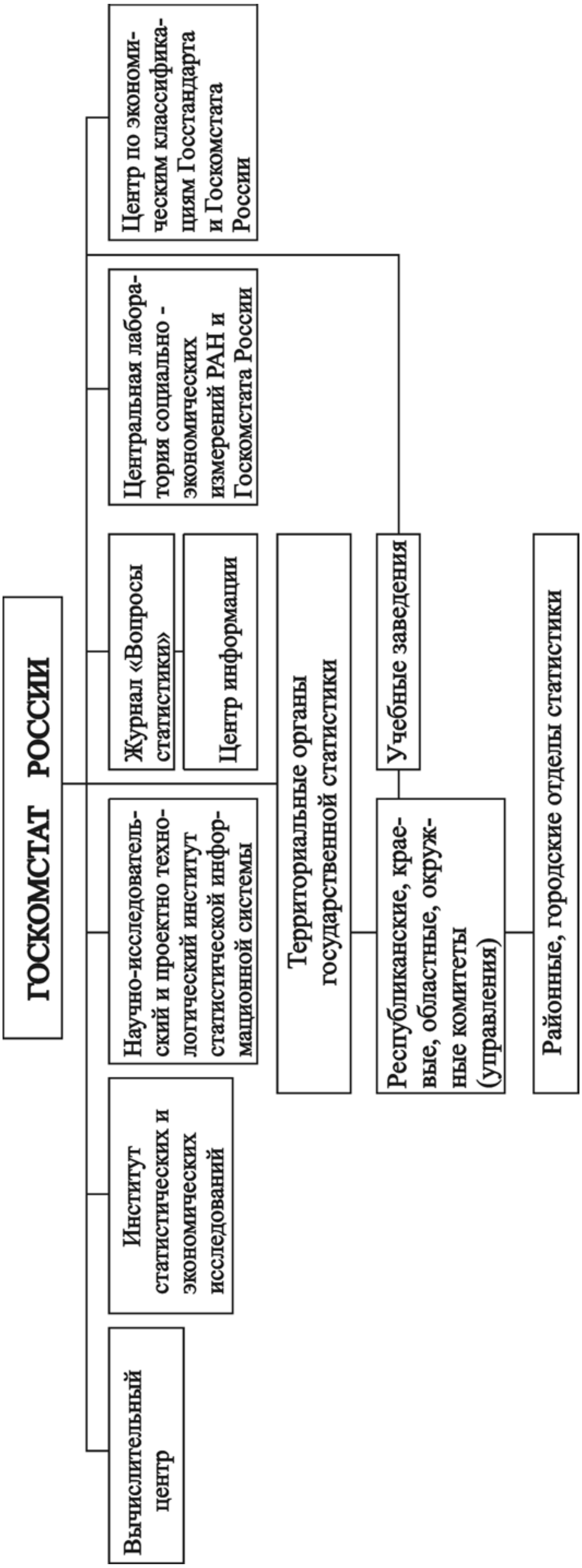

ГОСКОМСТАТ РОССИИ
Вычислительный центр
Институт статистических и экономических исследований
Научно-исследователь-ский и проектно техно-логический институт статистической инфор-мационной системы
Журнал «Вопросы статистики»
Центр информации
Центральная лабора-тория социально - экономических измерений РАН и Госкомстата России
Центр по экономи-ческим классифика-циям Госстандарта и Госкомстата России
Территориальные органы государственной статистики
Учебные заведения
Республиканские, крае-вые, областные, окруж-ные комитеты (управления)
Районные, городские отделы статистики

Рис. 1.1.

В соответствии с положением основными **задачами Госкомстата России** являются:

1) **предоставление официальной статистической информации** Президенту, правительству, Федеральному собранию Российской Федерации, федеральным органам исполнительной власти, общественности, международным организациям;

2) **разработка,** научно обоснованной **статистической методологии**, соответствующей потребностям общества на современном этапе и международным стандартам;

3**) координация статистической деятельности в государстве**;

4**) разработка экономико-статистической информации**, ее анализ, составление национальных счетов, проведение необходимых балансовых расчетов;

5) **гарантирование полноты** и научной обоснованности официальной статистической информации, обеспечение равного доступа к ее изучению всем пользователям.

Основные **функции Госкомстата России** состоят в том, что он:

1) организует проведение государственных статистических наблюдений по разработанным им или согласованным с ним программам, формам и методикам;

2) обеспечивает функционирование ЕГРПО (Единого государственного регистра предприятий и организаций).

1.4. Предмет и методы статистики

Под **предметом** статистики понимается количественная сторона массовых общественных явлений в постоянной связи с их качественным содержанием, а также количественное выражение **закономерностей** общественного развития в конкретных условиях места и времени.

Предмет статистики исследуется при помощи определенных понятий, к которым относятся: статистическая совокупность, единица совокупности, признак, статистический показатель, система статистических показателей (рис. 1.2.).

1. Статистическая совокупность – это совокупность социально-экономических объектов или явлений общественной жизни, объединенных некоей **качественной основой**, общей связью, но отличающихся друг от друга отдельными признаками. Таковы, например, совокупность домохозяйств, совокупность семей, совокупность предприятий, фирм, объединений и т.д.

Статистическая совокупность – это множество объектов или явлений, изучаемых статистикой, которые имеют один или несколько общих признаков и различаются между собой по другим признакам.

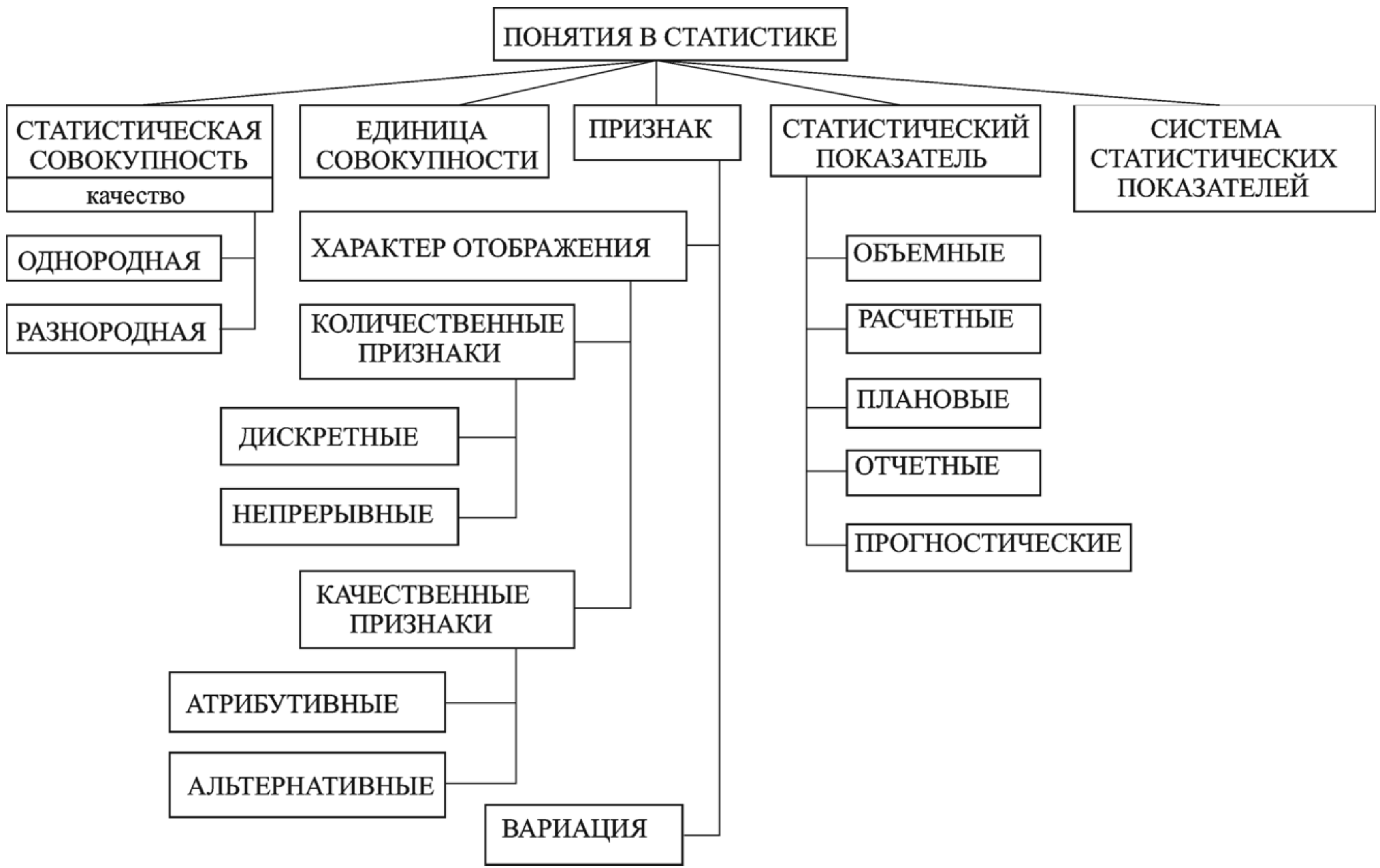

ПОНЯТИЯ В СТАТИСТИКЕ
СТАТИСТИЧЕСКАЯ СОВОКУПНОСТЬ
качество
ОДНОРОДНАЯ
РАЗНОРОДНАЯ
ЕДИНИЦА СОВОКУПНОСТИ
ПРИЗНАК
ХАРАКТЕР ОТОБРАЖЕНИЯ
КОЛИЧЕСТВЕННЫЕ ПРИЗНАКИ
ДИСКРЕТНЫЕ
НЕПРЕРЫВНЫЕ
КАЧЕСТВЕННЫЕ ПРИЗНАКИ
АТРИБУТИВНЫЕ
АЛЬТЕРНАТИВНЫЕ
ВАРИАЦИЯ
СТАТИСТИЧЕСКИЙ ПОКАЗАТЕЛЬ
ОБЪЕМНЫЕ
РАСЧЕТНЫЕ
ПЛАНОВЫЕ
ОТЧЕТНЫЕ
ПРОГНОСТИЧЕСКИЕ
СИСТЕМА СТАТИСТИЧЕСКИХ ПОКАЗАТЕЛЕЙ

Рис. 1.2.

Совокупности могут быть **однородными и разнородными.**

Совокупность называется однородной, если один или несколько изучаемых существенных признаков ее объектов являются общими для всех единиц. Совокупность оказывается однородной именно с точки зрения этих признаков.

Совокупность, в которую входят явления разного типа, считается разнородной. Совокупность может быть однородна в одном отношении и разнородна в другом. В каждом отдельном случае однородность совокупности устанавливается путем проведения качественного анализа, выяснения содержания изучаемого общественного явления.

2. Единица совокупности – это первичный элемент статистической совокупности, являющийся носителем признаков, подлежащих регистрации, и основой ведущегося при обследовании счета. Например, группа в университете обследуется по успеваемости, студенты – единицы совокупности.

3. Признак – это характерное свойство изучаемого явления, отличающее его от других явлений.

В разных отраслях статистики изучаются разные признаки. Так, например, объектом изучения является предприятие, а его признаками – вид продукции, объем выпуска, численность работающих и т.д. Или объект – отдельный человек, а признаки – пол, возраст, национальность, рост, вес и т.д.

По **характеру отображения** свойств единиц изучаемой совокупности признаки делятся на две основные группы: **признаки качества и признаки количества**.

Качественный признак может быть **атрибутивным и альтернативным.**

Атрибутивный признак (в философии «атрибут» – неотъемлемое свойство предмета) **–** признак, отдельные значения которого выражаются в виде понятий, наименований.

Профессия – токарь, слесарь, технолог, учитель, врач и т.д.

Цвет глаз, кожи и т.д.

Альтернативный признак – признак имеет противоположные по значению варианты: **да, нет**. Например, продукция может быть годной или бракованной (не годной); для представителей отдельных возрастных групп существует вероятность дожить или не дожить до следующей возрастной группы; каждое лицо может состоять в браке или нет и т.д.

Количественный признак – признак, определенные значения которого имеют количественные выражения:

Рост – 185, 172, 164, 158;

Вес – 105, 72, 54, 48.

Каждый объект изучения может обладать целым рядом статистических признаков, но от объекта к объекту одни **признаки меняются**, другие остаются **неизменными. Меняющиеся признаки** от одного объекта к другому принято называть **варьирующими**. Именно эти признаки изучаются в статистике, поскольку не изменяющийся признак изучать неинтересно. Предположим, что в вашей группе только мужчины, у всех один признак (пол – мужской) и по этому признаку больше сказать нечего. А если есть и женщины, то уже можно посчитать их процент в группе, динамику изменения численности женщин по месяцам учебного года и др.

Вариация – это многообразие, изменяемость величины признака у отдельных единиц совокупности наблюдения.

Вариация качественного признака: пол – мужской, женский.

Вариация количественного признака: з/п – 1000, 5 000, 10 000.

Отдельные **значения признака** называются **вариантами** этого признака.

Если построить ряд из меняющихся значений показателя, то такой ряд называется **вариационным** рядом показателя.

4. Явления и процессы в жизни общества изучаются статистикой посредством **статистических показателей.**

Статистический показатель – это количественная оценка свойства изучаемого явления.

Статистические показатели могут быть **объемными** (численность населения, трудовых ресурсов) и **расчетными** (средние величины). Они могут быть **плановыми, отчетными и прогностическими** (т.е. выступать в качестве прогнозных оценок).

Статистические показатели называются **натуральными**, когда они выражены в единицах счета или различных физических единицах измерения.

Статистические показатели следует отличать от **статистических данных. Статистические данные** – это конкретные численные значения статистических показателей. Они всегда определены не только качественно, но и количественно и зависят от конкретных условии места и времени.

5. Система статистических показателей – это совокупность статистических показателей, отражающая взаимосвязи, которые объективно существуют между явлениями. Система статистических показателей охватывает все стороны жизни общества на различных уровнях: страны, региона – макроуровень; предприятий, фирм, объединений, семей – микроуровень.

Для каждой общественно-экономической формации характерна определенная система статистических показателей.

В целях изучения и познания сущности явлений в статистике применяют различные взаимосвязанные между собой специфические приемы (методы) исследования, совокупность которых образует **статистическую методологию.** В основе статистической методологии лежат методы из математической статистики, теории вероятности и других разделов математики, которые воплощаются в **статистических методах**.

Статистические методы – это совокупность приемов, применяемых в процессе **статистического исследования.** Обычно статистические исследования делятся на взаимосвязанные и в большей мере самостоятельные **этапы**, как правило, обособленные друг от друга во времени. Выделяют три основных этапа: статистическое наблюдение, сводку и обработку материалов, анализ данных.

Контрольные вопросы

1. Что изучает статистика?
2. Что означает термин «статистика»?
3. Какова структура Госкомстата РФ?
4. Какие основные понятия в статистике?
5. Сущность статистической совокупности?
6. Признак и его характеристики?
7. Сущность количественных и качественных признаков?
8. Какие существуют виды статистических показателей?

ГЛАВА 2. СТАТИСТИЧЕСКОЕ НАБЛЮДЕНИЕ

2.1. Понятие о статистическом наблюдении

Статистическое наблюдение – это первая стадия всякого статистического исследования, представляющего собой **научно-организованный по единой программе учет фактов, характеризующих явления и процессы общественной жизни.**

Любое статистическое наблюдение осуществляется с помощью оценки и регистрации признаков единиц совокупности в соответствующих учетных документах.

Наблюдение является **фундаментом** статистического исследования. В процессе наблюдения формируются **данные**, которые на последующих этапах исследования **подвергаются обработке и анализу**. От качества наблюдения зависят окончательные данные, которыми статистика характеризует общественные явления. При неправильно организованном наблюдении все дальнейшее исследование будет обречено на неудачу. В свою очередь качество материалов, собранных в процессе статистического наблюдения, зависит от научной организации и надлежащего выполнения наблюдения.

Научная организация статистического наблюдения предполагает:

- определение объекта и единицы наблюдения;
- разработку программы наблюдения;
- разработку организационного плана проведения наблюдения.

Объект статистического наблюдения – это совокупность общественных явлений и процессов, которые подлежат данному статистическому наблюдению. Для объекта статистического наблюдения характерно то, что его нельзя изучать непосредственно в целом. Изучение объекта предполагает выделение в его составе отдельных единиц и их наблюдение.

Единицей статистического наблюдения является составной элемент объекта наблюдения, являющийся носителем признаков, подлежащих регистрации в процессе данного наблюдения. Так, например, объектом переписи населения является совокупность всех жителей страны, а единицей наблюдения – каждый человек.

Правильное определение единицы наблюдения имеет очень важное значение. Неточность в решении вопроса о единице наблюдения может привести к путанице, а иногда и к преднамеренному искажению действительности.

Вопрос о том, что брать за единицу наблюдения, в каждом конкретном случае должен быть решен с учетом цели наблюдения.

От единицы наблюдения следует отличать **отчетную единицу**. **Отчетная единица –** это источник сведений, та первичная ячейка, от которой должны поступать сведения о единицах наблюдения. При наблюдении в форме отчетности такими единицами являются объединения, предприятия, колхозы, учреждения; при обследовании бюджетов семей рабочих, служащих и колхозников – отдельная семья. Определение объекта, единицы наблюдения и отчетной единицы должно быть строго научным. Эти вопросы должны быть предельно ясны каждому участнику исследования.

Программа статистического наблюдения представляет собой **перечень вопросов**, по которым нужно получить в процессе наблюдения сведения в отношении каждой обследуемой единицы. Один и тот же объект может быть исследован с различных сторон.

При разработке программы наблюдения, прежде всего, необходимо сформулировать **цель и задачи всего исследования**.

Одним из основных требований, предъявляемых к **программе наблюдения**, а следовательно, и к документам, является **краткость и точность вопросов**. Важное значение имеет **последовательность вопросов**, так как ответы в известной степени должны контролировать друг друга.

Организационный план статистического наблюдения представляет собой перечень мероприятий, необходимых для успешного выполнения работы по сбору и обработке материалов, с указанием сроков и исполнителей.

В **плане наблюдения** прежде всего определяются **права и обязанности каждого соисполнителя.**

В **плане** указывается **срок проведения наблюдения**, т.е. время начала и окончания сбора сведений. Это время нельзя отождествлять с временем наблюдения, т.е. с тем временем, к которому относятся сведения. Статистические показатели характеризуют исследуемое явление либо за определенный момент времени.

В плане наблюдения точно определяют **территорию,** на которой производится наблюдение, а также **лиц и организаций**, ответственных за проведение **подготовительных работ**, **сбор, проверку и обработку** материалов на отдельных участках территории. Закрепленный за счетчиком участок называется счетным участком. Размер его зависит от сложности программы и срока проведения наблюдения, характера изучаемого объекта, расположения единиц наблюдения по территории и других факторов.

Большое место в плане отводится **проведению подготовительных работ**. При всем разнообразии конкретного содержания проводимых подготовительных мероприятий в различных наблюдениях можно выделить некоторые подготовительные работы **общего характера**. Одной из основных таких работ является **составление списка отчетных единиц.** Этот список (промышленных объединений и предприятий, строек, колхозов, совхозов и т.п.) необходим как для проверки полноты и своевременности поступивших сведений, так и для определения объема работ и расчета необходимого количества работников для проведения наблюдения.

Важным подготовительным мероприятием является **инструктаж** аппарата учетно-экономических служб предприятий и привлеченных для сбора данных лиц о правильном составлении документов (статистических отчетов, переписных листов и т. п.). Необходимо заблаговременно и в достаточном количестве отпечатать и разослать бланки документов и инструкции по их заполнению.

К числу важнейших **подготовительных мероприятий**, особенно при проведении переписей населения, относится **пропаганда среди населения средствами печати, радио, телевидения задач и целей обследования** для успешного проведения подобного рода статистического наблюдения.

2.2. Формы, способы и виды статистического наблюдения

С точки зрения **организации статистического наблюдения** различают **две основные формы: отчетность и специально организованное статистическое наблюдение.**

1. Отчетность как форма статистического наблюдения характеризуется тем, что статистические органы систематически получают от предприятий, учреждений и организаций в установленные сроки сведения об условиях и результатах работы за прошедший период, объем и содержание которых определены утвержденными формами отчетности. При этом источником сведений являются первичные учетные записи в документах бухгалтерского и оперативного учета. Учетно-статистический аппарат предприятий обрабатывает первичные записи в документах и результаты заносит в формы отчетов.

В настоящее время в России **отчетность является основной формой статистического наблюдения**. Основную массу сведений, необходимых для повседневного контроля за выполнением государственных планов, для планирования и управления народным хозяйством, а также для научных исследований статистические органы получают в форме отчетности.

2. Специально организованное статистическое наблюдение представляет собой сбор сведений в форме переписей, единовременных учетов и обследований. Их организуют для изучения тех явлений, которые не могут быть охвачены обязательной отчетностью. Это в первую очередь относится к сведениям о численности и составе населения, об источниках его доходов, видах расходов, неудовлетворенном спросе на предметы потребления и т.п. Сбор таких сведений может быть осуществлен только в форме специальных статистических наблюдений.

Виды статистического наблюдения различают **по времени регистрации данных** и **по степени охвата единиц изучаемой совокупности**.

По характеру регистрации данных во времени статистические наблюдения различают:

- непрерывные (текущие);
- прерывные: периодические и единовременные.

Текущие наблюдения ведут непрерывно по мере возникновения изучаемого факта. Практически в основе всей статистической отчетности лежит непрерывная первичная регистрация фактов на предприятиях и в учреждениях: учет произведенной продукции, учет поступления и отпуска материалов, учет выручки в магазинах, учет количества родившихся и т.д.

При текущем наблюдении нельзя допускать значительного разрыва между моментом возникновения факта и моментом его регистрации. Так, например, данные о производстве продукции нужно фиксировать ежедневно, а показатели за месяц получать путем суммирования данных ежедневного учета. Если же записи производить раз в месяц, то, естественно, показатели о производстве продукции будут неточны.

Прерывное наблюдение отражает состояние изучаемого явления на **определенный момент.** Оно может быть **периодическим,** если проводится через равные промежутки времени, или **единовременным,** если проводится нерегулярно, по мере необходимости. Примером периодического наблюдения являются ежегодные переписи не установленного оборудования на 1 января или регистрация цен колхозной торговли на сельскохозяйственные продукты 25-го числа каждого месяца. Типичным примером единовременного (прерывного) наблюдения являются переписи населения в России, проведенные в 1920, 1926, 1939, 1959, 1970, 1979 и 1989 гг.

Выбор того или другого вида статистического наблюдения зависит в первую очередь от сущности изучаемых явлений. Так, например, в производстве ежедневно расходуются сырье и материалы и, разумеется, учет их расхода и запасов должен быть непрерывным, ежедневным. По-другому обстоит дело с посевными площадями. Для учета посевных площадей под различными сельскохозяйственными культурами достаточны весенние и осенние учеты. Некоторые явления изменяются еще медленнее и для их изучения достаточно проводить наблюдения раз в 5 лет или реже.

По степени охвата единиц изучаемой совокупности статистические наблюдения различают:

- сплошные;
- несплошные.

При сплошном статистическом наблюдении обследованию подлежат все единицы, входящие в состав изучаемой совокупности. Однако не следует понимать так, что сплошное наблюдение обязательно должно во всех случаях охватить изучаемое явление по всей стране. Изучаемая совокупность может быть ограничена территориальными, ведомственными, отраслевыми и другими рамками. Важно то, что в пределах этой совокупности сплошное наблюдение предполагает обязательную регистрацию сведений о каждой единице объекта.

При несплошном наблюдении обследуется часть единиц изучаемой совокупности. При организации несплошного наблюдения, как правило, ставится задача распространить результаты наблюдения на всю совокупность. Например, из 20 тыс. электроламп 100 шт. подвергают проверке на продолжительность горения в часах. По результатам качества обследованной части делают вывод о качестве всей партии электроламп. При правильной организации несплошного наблюдения результаты его достаточно достоверны.

Несплошное наблюдение проводят в тех случаях, когда наблюдение сопряжено с уничтожением или порчей обследуемых единиц либо с целью экономии материальных и трудовых затрат, либо с целью контроля данных сплошного наблюдения. Кроме того, исследования социологические или такие, например, как изучение спроса населения на товары народного потребления, практически не могут быть сплошными.

В статистической практике несплошное наблюдение проводят в трех видах:

- обследование основного массива;
- выборочное наблюдение;
- монографическое описание.

Обследование основного массива представляет собой такое несплошное наблюдение, при котором из всей совокупности единиц для наблюдения отбирается такая часть их, у которой объем изучаемого признака составляет подавляющую долю в объеме признака по всей совокупности.

Выборочное наблюдение является наиболее совершенным видом несплошного наблюдения. В отличие от обследования основного массива при выборочном наблюдении, объем изучаемого признака отобранной для обследования части единиц может составлять незначительную долю. Характерной особенностью выборочного наблюдения является строго научный отбор единиц для обследования, обеспечивающий близость состава отобранной части единиц к составу единиц всей совокупности.

Для монографического описания характерно детальное изучение и описание отдельных единиц совокупности или небольших групп явлений, характерных в каком-либо отношении. Монографическое описание не ставит своей задачей давать характеристику всей совокупности. Непременным условием научной организации монографического наблюдения является его качественная представительность. Единицы или группы явлений, выбранные для монографического описания, должны быть типичными, чтобы на основе их исследования можно было судить о характере этих явлений.

Основными способами **получения статистической информации** являются **непосредственное наблюдение, документальный способ и опрос.**

Способ непосредственного наблюдения характеризуется тем, что представители органов государственной статистики или других организаций записывают данные в статистические документы после личного осмотра, пересчета, измерения или взвешивания единиц наблюдения.

При документальном способе наблюдения источником сведений служат различные документы. Этот способ наблюдения используется при составлении предприятиями и учреждениями статистической отчетности на основе документов первичного учета. Поскольку источником сведений при составлении первичных документов является непосредственное наблюдение, то при надлежащей организации первичного учета и правильной разработке на их основе форм статистической отчетности документальный способ наблюдения обеспечивает большую точность сведений.

Непосредственное наблюдение и документальный способ дают наиболее достоверные данные, но в ряде случаев они неприменимы. В частности, при переписях населения и бюджетных обследованиях семей рабочих, служащих и колхозников для получения сведений практикуют **способ опроса**.

При способе опроса источником сведений являются ответы опрашиваемых лиц, Опрос может быть организован по-разному: **экспедиционным способом, саморегистрацией, корреспондентским способом и анкетным способом.**

При **экспедиционном способе (устном опросе)** представители статистических органов спрашивают обследуемое лица (или его представителя) и, с его слов записывают сведения в бланк наблюдения. Этот способ опроса наиболее дорогой, трудоемкий, требует участия в обследовании, большого числа специально подготовленных счетчиков-регистраторов, но он гарантирует получение материалов высокого качества. Важнейшие статистические обследования населения и переписи населения в нашей стране проводят экспедиционным способом.

При **способе саморегистрации (самоисчислении)** обследуемым единицам (предприятиям или гражданам) **вручают бланк обследования и дают указания по его заполнению**. Заполненные бланки в установленный срок пересылают по почте, или представители статистических органов непосредственно на месте проверяют правильность заполнения и собирают бланки. Этот **способ применяется** при проведении **многочисленных единовременных учетов и обследований** на предприятиях, при обследовании бюджетов семей.

При **корреспондентском способе** сведения статистическим органам **сообщают добровольные корреспонденты**. Этот способ не требует больших затрат, но он не обеспечивает высокого качества материалов, так как проверить точность сообщаемых корреспондентами сведений непосредственна на месте не представляется возможным.

Анкетный способ сбора данных основан **на принципе добровольного заполнения адресатами анкет (листов опроса)**. Как правило, заполненных анкет возвращается меньше, чем рассылается. Кроме того, проверить достоверность собранного материала практически невозможно. Поэтому анкетный способ может применяться в тех случаях, когда **не требуется высокая точность сведений**, а нужны **приближенные характеристики.**

Все формы, способы и виды статистического наблюдения можно свести в схему, показанной на рис. 2.1.

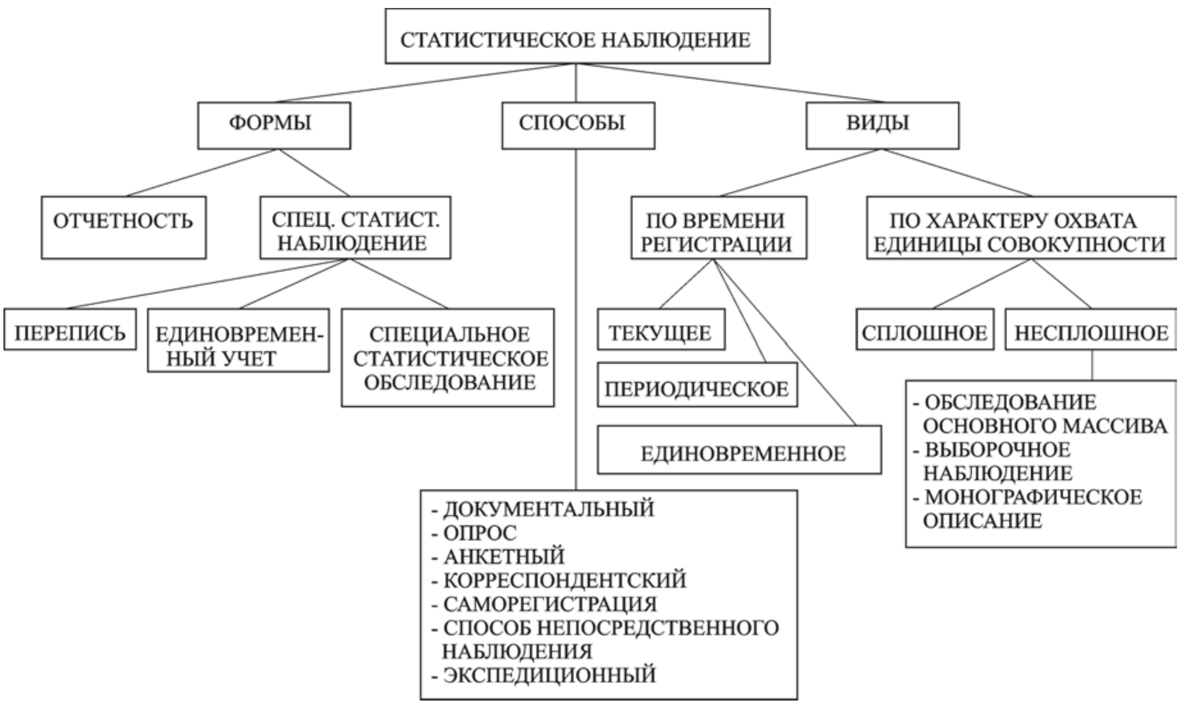


Рис. 2.1.

2.3. Контроль материалов наблюдения

Материалы, собранные в результате статистического наблюдения, подвергаются всесторонней проверке и контролю. Они должны быть проверены с точки зрения **полноты** охвата всех единиц **совокупности наблюдения** и **правильности заполнения документов**. Одновременно с этим статистические данные проверяются в порядке **логического и арифметического контроля**.

Логический контроль состоит в **сопоставлении ответов на взаимосвязанные между собой вопросы программы наблюдения.** Например, при переписи населения вопросы о возрасте, грамотности, семейном положении контролируются взаимосвязано. Если, например, окажется, что десятилетний ребенок женат или двухлетний грамотен, то ясно, что при заполнении бланка допущены ошибки в записи возраста или семейного положения.

Счетный, или арифметический контроль сводится к проверке общих и групповых цифровых итогов и их сопоставлению; задача его – **обнаружить и исправить неверные итоги числовых показателей.**

Пример 2.1. В документах на предприятии обнаружены арифметическая и логическая рассогласованность:

Таблица 2.1

Ошибки регистрации

№п/п	Заработная плата	Возраст	Общий стаж	Непрерывный стаж
1	800	22	10	8
2	1500	42	20	2
3	1500	25	45	55
4	2000	35	15	10
Итого	5200			

В 1 и 3 строке логические ошибки, в 2 столбце арифметическая ошибка – сумма равна 5800.

При обнаружении ошибок, которые нельзя объяснить или исправить, проводится **дообследование**, что, естественно, усложняет и задерживает статистическое исследование. Поэтому необходимо особое внимание уделять рациональной организации наблюдения, предусматривающей появление ошибок и обеспечивающей получение доброкачественных статистических данных. Для этого специально изучают природу ошибок статистического наблюдения.

Ошибки статистического наблюдения могут быть разбиты на две группы: **ошибки репрезентативности и ошибки регистрации.**

Ошибки репрезентативности свойственны только выборочному наблюдению. Они показывают, в какой степени выборочная совокупность представляет (репрезентует) генеральную совокупность. Эти ошибки возникают вследствие того, что наблюдению подвергается лишь часть единиц изучаемой совокупности, и сведения, полученные в результате этого, не могут абсолютно точно отобразить свойства всей массы явлений совокупности.

Ошибки регистрации, возникающие в результате неправильного установления фактов или неправильной их записи, подразделяются на **случайные и систематические**.

Случайные ошибки могут дать искажение как в одну, так и в другую сторону. Они могут быть совершены по невниманию, из-за низкой квалификации работника и т.д. Например, записывается цифра не в ту графу или вместо возраста 53 года записывается 63.

Систематические ошибки более опасны, так как они в значительной степени влияют на итоговые показатели. Систематические ошибки могут быть **преднамеренными и непреднамеренными.**

Непреднамеренные ошибки совершаются неумышленно. Так, например, тяготение к круглым цифрам вносит большие осложнения при сборе сведений о возрасте населения. Многие отвечают на вопрос о числе прожитых лет круглыми цифрами. Если человеку, например, 39 или 41 год, он отвечает 40. В результате получается очень много 40-летних и очень мало 39-ти и 41-летних.

Преднамеренные ошибки возникают в результате умышленного искажения фактов. Все преднамеренные ошибки относятся к **разряду систематических**.

2.4. Статистическая отчетность

Статистическая отчетность – это официальный документ, в котором содержатся сведения о работе подотчетного объекта, занесенные на специальную форму. Статистическая отчетность чаще всего базируется на **данных бухгалтерского учета**.

Первичный учет представляет собой регистрацию различных фактов (событий, процессов и т.д.), производимых по мере их свершения и, как правило, на первичном учетном документе. **Примером** может служить свидетельство о рождении ребенка. В торговле к первичным учетным документам относятся наряды на отпуск товаров, счета-фактуры, накладные и т.д. В функции первичного учета входят операции наблюдения, т.е. регистрация данных и подсчет итогов.

Каждое предприятие или учреждение представляет установленные формы статистической отчетности, характеризующие различные стороны их деятельности. Все **формы статистической отчетности** утверждают органы государственной статистики.

По своему **содержанию формы отчетности** бывают:

- типовыми (общими);
- специализированными.

Общая отчетность – это отчетность, содержащая одни и те же данные для определенной отрасли народного хозяйства и для предприятий (учреждений) всего народного хозяйства.

В **специализированной отчетности** содержатся специфические показатели отдельных отраслей промышленности, сельского хозяйства и т.п.

По периоду времени, за который предоставляется отчетность, по его длительности различают **отчетность текущую и годовую**. Если сведения представляются за год, то такую отчетность называют годовой. Отчетность за все другие периоды в пределах менее года, соответственно **квартальная, месячная, недельная** и т.п. называется **текущей**.

Контрольные вопросы

1. Какова сущность статистического наблюдения?
2. Какие бывают формы статистического наблюдения?
3. Способы статистического наблюдения?
4. Каковы виды статистического наблюдения?
5. Какие бывают типы ошибок наблюдения?
6. Сущность статистической отчетности?

ГЛАВА 3. СВОДКА И ГРУППИРОВКА СТАТИСТИЧЕСКИХ ДАННЫХ

3.1. Понятие о сводке и группировке статистического материала

В результате **первой стадии (этапа)** статистического исследования – статистического наблюдения – получают сведения о каждой единице совокупности. Задача **второй стадии** статистического исследования состоит в том, чтобы упорядочить и обобщить первичный материал, свести его в группы и на этой основе дать обобщающую характеристику совокупности. Этот этап в статистике называется **сводкой**.

Статистическая сводка – это научно организованная обработка материалов наблюдения, включающая в себя систематизацию, группировку данных, составление таблиц, подсчет групповых и общих итогов, расчет производных показателей (средних, относительных величин). Она позволяет перейти к обобщающим показателям совокупности в целом и отдельных ее частей, осуществлять анализ и прогнозирование изучаемых процессов.

Различают **простую сводку** (подсчет только общих итогов) и **статистическую группировку**, которая сводится к расчленению совокупности на группы по существенному, для единиц совокупности, признаку. **Группировка позволяет** получить такие результаты, по которым можно выявить состав совокупности, характерные черты и свойства типичных явлений, обнаружить закономерности и взаимосвязи.

Результаты сводки могут быть представлены в виде **статистических рядов распределения**.

Статистическим рядом распределения называют упорядоченное распределение единиц совокупности на группы по изучаемому признаку. В зависимости от признака ряды могут быть **вариационными (количественными)** и **атрибутивными (качественными).**

Количественные признаки – это признаки, имеющие количественное выражение у отдельных единиц совокупности, например, заработная плата рабочих, стоимость продукции промышленных предприятий, возраст людей, урожайность отдельных участков посевной площади и т.д.

Атрибутивные признаки – это признаки, не имеющие количественной меры. Например, пол (мужской, женский), отрасль народного хозяйства, вид продукции, профессия рабочего и т.д.

Вариационные ряды могут быть **дискретными** или **интервальными**.

Дискретный ряд распределения – это ряд, в котором варианты выражены целым числом.

Пример 3.1. Имеется следующее распределение рабочих по тарифным разрядам:

Таблица 3.1

Распределение рабочих по тарифным разрядам

Тарифный разряд	Число рабочих, чел.
1-й	10
2-й	20
3-й	40
4-й	60
5-й	50
6-й	20
Итого	200

Интервальный ряд распределения – это ряд, в котором значения признака заданы в виде интервала. На примере 3.1, распределение рабочих по разрядам можно представить в виде интервального ряда.

Таблица 3.2

Распределение рабочих по разрядам, представленное в виде интервального ряда

Тарифный разряд	Число рабочих, чел.
1–2-й	30
3–4-й	100
5–6-й	70
Итого	200

Статистические ряды распределения позволяют систематизировать и обобщать статистический материал. Однако они не дают всесторонней характеристики выделенных групп. Чтобы решить ряд конкретных задач, выявить особенности в развитии явления, обнаружить тенденции, установить зависимости, необходимо произвести **группировку статистических данных**.

Статистическая группировка – это процесс образования однородных групп на основе расчленения статистической совокупности на части или объединения изучаемых единиц в частные совокупности **по существенным** для них признакам, каждая из них характеризуется системой статистических показателей. Например, группировка промышленных предприятий по формам собственности, группировка населения по размеру среднедушевого дохода, группировка коммерческих банков по сумме активов баланса и т.д.

3.2. Группировочные таблицы

Результаты группировки оформляются в виде **группированных таблиц**, делающих информацию обозримой.

Таблица содержит сводную числовую характеристику исследуемой совокупности по одному или нескольким существенным признакам, взаимосвязанным логикой анализа.

Группировочная таблица содержит **три вида заголовков: общий, верхний и боковые**.

Заголовки таблиц должны быть краткими и раскрывать содержание показателей.

Таблица 3.3

Общий заголовок

Содержание строк	Наименование граф (верхние заголовки)					
А	1	2	3	4	5	...
Наименования строк						
(боковые заголовки)						
Итоговая строка						Итоговая графа

Общий заголовок отражает содержание всей таблицы с указанием, к какому месту и времени она относится. Он располагается над макетом по центру и является внешним (общим) заголовком.

Верхние заголовки характеризуют содержание граф (заголовки **сказуемого**), а боковые (заголовки **подлежащего**) – строк.

Подлежащее статистической таблицы – объект, характеризующийся цифрами.

Сказуемое – система показателей, которыми характеризуется объект изучения, т.е. подлежащее.

В зависимости от строения подлежащего все статистические таблицы можно разделить на **три группы**:

1. простые или перечневые;
2. групповые (сложные);
3. комбинационные.

Пример 3.2. Группировочная таблица.

Отношение студентов одного из факультетов к понижению размера стипендии (по результатам исследования в январе 1999 г).

Таблица 3.4

Отношение студентов к понижению размера стипендии

	Поддерживаю	Не поддерживаю	Безразлично
Студенты 1 курса	2	20	3
Студенты 2 курса	2	25	3
Студенты 3 курса	1	30	2
Студенты 4 курса	–	35	–
Студенты 5 курса	–	25	–
Итого	5	135	8

При проведении группировки статистических данных для решения ряд конкретных задач, выявить особенности в развитии явлений, обнаружить тенденции, установить зависимости, необходимо выбирать **группировочный признак** и разработать **систему показателей**, которыми будут характеризоваться выделенные группы. Определение и обоснование показателей целиком зависят от цели исследования и поставленной задачи. В зависимости от цели и задач исследования различают следующие виды группировок: **типологические, структурные, аналитические**.

К **типологическим группировкам** относят все группировки, которые характеризуют качественные особенности и различия между типами явлений.

Типологические группировки широко применяются в экономических, социальных и других исследованиях. Приведем пример **типологической** группировки (табл. 3.5).

Таблица 3.5

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

Группы предприятий по формам хозяйствования	Объем промышленной продукции, млрд. руб.	В % к итогу
Государственные	405,0	89,20
Арендные	19,0	4,19
Кооперативные	30,0	6,61
Всего	454,0	100,0

Происходят изменения в социальной занятости работников в народном хозяйстве: увеличилось число работников в кооперативном и индивидуальном секторах экономики.

Структурная группировка – это группировка, выявляющая состав (строение, структуру) однородной в качественном отношении совокупности по какому-либо признаку. Примером могут служить группировки предприятий по проценту выполнения плана, по числу рабочих и т.д. Состав населения может быть сгруппирован по полу, по возрасту, по уровню образования, по роду занятий и т.д.

Таблица 3.6

Распределение заводов по стоимости ОПФ и численности рабочих

Группы заводов по среднегодовой стоимости ОПФ, млн. руб.	Численность рабочих	
	человек	в % к итогу
1,0–2,2	820	13,86
2,2–3,4	3150	53,25
3,4–4,6	1945	32,89
Итого	5915	100,0

Наибольшая численность рабочих приходится на группу заводов со среднегодовой стоимостью ОПФ от 2,2 до 3,4 млн. руб.(ОПФ – основные производственные фонды).

Аналитическая группировка – это группировка, которая применяется для исследования взаимосвязи между явлениями. Используя аналитические группировки, определяют факторные и результативные признаки изучаемых явлений.

Факторные – это признаки, оказывающие влияние на другие, связанные с ними признаки.

Результативные – это признаки, которые изменяются под влиянием факторных. Пример аналитической группировки (табл. 3.7).

Таблица 3.7

Зависимость товарооборота магазинов от размеров торговых площадей

Группы магазинов по объему товарооборота, тыс. руб.	Торговая площадь
1700–2000	18,5
2000–3000	22,5
3000–4200	59,0
Всего	100,0

Чем больше торговая площадь (факторный признак), тем выше объем товарооборота (результативный признак).

Комбинированные группировки. Образование групп по двум и более признакам, взятым в определенном сочетании, называется комбинированной группировкой. При этом группировочные признаки принято располагать, начиная с атрибутивного, в определенной последовательности, исходя из логики взаимосвязи показателей.

Применение комбинированных группировок обусловлено многообразием экономических явлений, а также необходимостью их всестороннего изучения. Но увеличение числа группировочных признаков ограничивается уменьшением наглядности, что снижает эффективность использования статистической информации. Примером комбинированной группировки может служить разделение образованных групп по формам хозяйствования на подгруппы по уровню рентабельности (доходности) или по другим признакам (производительность труда, фондоотдача и т.д.).

В заключении можно отметить, что при составлении таблицы надо соблюдать ряд правил:

- четко формулировать наименование, которое должно точно отражать цель составления таблицы;
- ясно и кратко формулировать название строк и граф таблицы;
- соблюдать последовательность расположения показателей сказуемого;
- указывать единицы измерения; если они одинаковые, то ед. измерения выносятся в заголовок и указываются в скобках;
- нумеровать графы;
- иметь итоговые показатели;
- если в таблице производится сопоставление с каким-либо годом, то в заголовке, в скобках, отражается год сопоставления;
- территориальные, административные образования перечисляются по алфавиту;
- данные за многие годы располагаются в хронологическом порядке;
- если в таблице абсолютные и относительные показатели за ряд лет, то сначала приводятся абсолютные, затем относительные показатели за один год, затем так же за следующий год;
- если значение признака в какой-либо клетке неизвестно, ставится знак «Х», или «...», или «н.с.» (нет сведений);
- нулевые значения признака – знак «–».

3.3. Группировка по количественному признаку

При составлении структурных группировок на основе варьирующих количественных признаков необходимо определить **количество групп и интервалы группировки.**

Интервал – количественное значение, отделяющее одну единицу (группу) от другой, т.е. он очерчивает **количественные границы групп**.

Как правило, **величина интервала** представляет собой разность между максимальным и минимальным значениями признака в каждой группе.

Интервалы группировки могут **равные и неравные**.

Равные интервалы используются, когда изменение признака внутри совокупности происходит равномерно, либо если далее планируется последующая математическая обработка сгруппированных данных.

Неравные интервалы обычно используются как прогрессивно увеличивающиеся. В экономической статистике чаще всего устанавливаются границы интервалов, основанные именно на таком принципе – прогрессивно увеличивающиеся.

Вопрос о **числе групп и величине интервала** следует решать с учетом множества обстоятельств, прежде всего исходя из целей исследования, значения изучаемого признака и т.д.

Количество групп и величина интервала связаны между собой: **чем больше образовано групп, тем меньше интервал, и наоборот**. Количество групп зависит от числа единиц исследуемого объекта и степени колеблемости группировочного признака. При небольшом объеме совокупности нельзя образовывать большое число групп, так как группы будут малочисленными.

При определении **количества групп** необходимо стремиться к тому, чтобы были учтены особенности изучаемого явления. Поэтому **количество групп** должно быть **оптимальным**, в каждую группу должно входить достаточно большое число единиц совокупности, что отвечает требованию закона больших чисел.

В общем случае число групп в группировке выбирается из таких предпосылок: изменчивость признака, число наблюдений, однородность групп.

Ориентировочно определить оптимальное количество групп с равными интервалами: можно, по формуле американского ученого **Стерджесса:**

$$\mathbf{n = 1 + 3.322\ lgN}$$

где N – число единиц совокупности (число наблюдений). В таблице 3.8 приведено значение оптимального количества групп в зависимости от число единиц совокупности.

Таблица 3.8

Оптимальное количество групп с постоянным интервалом, рассчитанное по формуле Стерджесса

N	15–24	25–44	45–89	90–170	180–359	360–719
n	5	6	7	8	9	10

Если, например, требуется произвести группировку с равными интервалами по данным об уровне месячной заработной платы рабочих, которая в 1995 г. колебалась в пределах от 3000 до 7500 руб., и необходимо при этом выделить 5 групп, то величина интервала, руб.:

$$i = (7500 - 3000)/5 = 900$$

Если в результате деления получается не целое число и возникнет необходимость округления, то округлять нужно, как правило, в большую сторону, а не в меньшую.

Прибавляя к минимальному значению признака (в данном случае 3000 руб.) найденное значение интервала, получаем верхнюю границу первой группы, руб.:

$$3000 + 900 = 3900$$

Прибавляя далее значение интервала к верхней границе первой группы, получаем верхнюю границу второй группы:

$$3900 + 900 = 4800, \text{ и т.д.}$$

В результате получим такие группы рабочих по размеру заработной платы, руб.: 3000–3900; 3900–4800; 4800–5700; 5700–6600; 6600–7500. В этом распределении имеет место неопределенность: к какой группе, например, отнести рабочего с заработком в 4800 руб., ко второй или третьей? Для устранения неопределенности есть правило левой границы – левое число включает в себя обозначенное значение, а правое – не включает. Значит рабочий, получающий 480 руб., должен быть отнесен к третьей группе. Аналогично нужно поступать в отношении всех остальных групп.

Интервалы групп могут быть **закрытыми**, когда указаны нижняя и верхняя границы (как в приведенном примере), и **открытыми**, когда указана лишь одна из границ (первый или последний интервалы, величина которых принимается равной величине смежных с ними интервалов). Во втором случае, чтобы доказать, что рабочий с заработной платой, равной, например, верхней границе интервала, включается в последнюю группу, ее следует обозначить «750 и выше». И наоборот, чтобы показать, что значение, равное верхней границе интервала, не входит в данную группу, последнюю группу нужно обозначить «свыше 750». Подобные функции выполняют слова «до», «менее» и «более».

Пример 3.3.

Имеются данные о работе 24 заводов одной из отраслей промышленности (табл. 3.9).

Изучая на эту таблицу трудно судить о характере распределения заводов, например, по проценту выполнения плана, по числу работающих, по стоимости основных фондов. Трудно сказать, какие показатели наиболее характерны для заводов данной отрасли промышленности. Поэтому имеющиеся данные надо привести в систему по интересующему нас признаку.

В качестве **изучаемого признака** возьмем стоимость основных производственных фондов и построим к нему ряд распределения с равными закрытыми интервалами. Величина интервала определяется по формуле

$$i = \frac{\chi_{\max} - \chi_{\min}}{n}$$

где χ_{max} и χ_{min} – максимальное и минимальное значения стоимости основных фондов, n – число групп.

Таблица 3.9

Статистические данные о работе заводов одной из отраслей промышленности

№ п/п	Среднегодовая стоимость ОПФ, млн. руб.	Среднесписоч. число работающих за отчетный период, чел.	Производство продукции за отчетный период, млн. руб.
A	B	C	D
1	3,0	360	3,2
2	7,0	380	9,6
3	2,0	220	1,5
4	3,9	460	4,2
5	3,3	395	6,4
6	2,8	280	2,8
7	6,5	580	9,4
8	6,6	200	11,9
9	2,0	270	2,5
10	4,7	340	3,5
11	2,7	200	2,3
12	3,3	250	1,3
13	3,0	310	1,4
14	3,1	410	3,0

Продолжение табл. 3.9

№ п/п	Среднегодовая стоимость ОПФ, млн. руб.	Среднесписоч. число работающих за отчетный период, чел.	Производство продукции за отчетный период, млн. руб.
A	B	C	D
15	3,1	635	2,5
16	3,5	400	7,9
17	3,1	310	3,6
18	5,6	450	8,0
19	3,5	300	2,5
20	4,0	350	2,8
21	1,0	330	1,6
22	7,0	260	12,9
23	4,5	435	5,6
24	4,9	505	4,4
Итог	94,1	8630	114,8

Согласно таблице 3.8 (формулы Стерджесса) образуем пять групп заводов (т.к. $N = 24$ и $n = 5$). Тогда величина интервала равна

$$i = (7{,}0 - 1{,}0) / 5 = 1{,}2$$

Теперь образуем группы заводов, которые отличаются друг от друга по среднегодовой стоимости основных фондов на эту величину (по табл. 3.9).

1,0 – 2,2 – 3 завода (1,0 + 1,2 = 2,2)

2,2 – 3,4 – 9 заводов (2,2 + 1,2 = 3,4) и т.д.

На основании этого составляем таблицу, в которой показываем распределение заводов по размеру основных фондов (и удельный вес заводов группы в % к итогу) (табл. 3.10). По этим данным хорошо видно изменение стоимости основных фондов и легко обозначить границы групп. Видно, что для данной отрасли характерной является группа заводов с основными фондами от 2,2 до 3,4 млн. руб., которая составляет 37,5% всех заводов, и что более половины заводов (58,3%) имеют стоимость основных фондов в размере от 2,2 до 4,6 млн. руб.

Таблица 3.10

Распределение заводов по размеру основных фондов

Группы заводов по стоимости ОПФ, млн. руб.	Число заводов	Удельный вес заводов группы в % к итогу
1,0–2,2	3	12,5
2,2–3,4	9	37,5
3,4–4,6	5	20,8
4,6–5,8	3	12,5
5,8–7,0	4	16,7
Итого	24	100

Теперь перейдем непосредственно к методу группировки. Для этого необходимо выбрать группировочный признак. Выявим данной отрасли промышленности распределение предприятий по мощности, а также влияние этого признака на объем производства. Мощность предприятия в значительной степени определяется размером основных фондов (здания, сооружение, машины, оборудование).

Чтобы выявить распределение предприятий по мощности, необходимо разбить совокупность заводов отрасли на группы по размеру стоимости основных фондов. Выше мы рассматривали построения рядов распределения, были выявлены пять групп.

Составим таблицу с системой показателей, куда занесем результаты группировки заводов по среднегодовой стоимости основных производственных фондов (табл. 3.11).

Таблица 3.11

Группировочная таблица

Табл. 3.6. №	Группы заводов по среднегод. стоимости ОПФ млн. руб.	Заводы		Основные произв. фонды		Численность рабочих		Валовая продукция	
		число зав.	в % к итогу	млн. руб.	в % к итогу	чел.	в % к итогу	млн. руб.	в % к итогу
I	1,0–2,2	3	12,5	5,0	5,3	820	9,5	5,6	4,8
II	2,2–3,4	9	37,5	27,4	29,1	3150	35,5	26,5	23,1
III	3,4–4,6	5	20,8	19,4	20,6	1945	22,5	23,0	20,1
IV	4,6–5,8	3	12,5	15,2	16,2	1295	15,0	15,9	13,9
V	5,8–7,0	4	16,7	27,1	28,8	1420	16,5	43,8	38,1
Итого		24	100	94,1	100	8630	100	100	114,8

Таким образом, в отличие от ряда распределения (табл. 3.11) группировка позволяет сделать конкретные и содержательные выводы. Данная группировка показывает, что наиболее крупные предприятия имеют лучшие производственные показатели. Около 29% предприятий (группы 4 и 5) имеют 45% всех основных фондов и дают 52% всего объема промышленной продукции, имея лишь 31% общего числа рабочих.

3.4. Вторичные группировки

Перегруппировка ранее сгруппированных статистических данных называется вторичной группировкой. К этому методу прибегают в тех случаях, когда в результате первоначальной группировки нечетко проявился характер распределения изучаемой совокупности.

Получение новых групп на основе имеющихся возможно двумя способами перегруппировки: **объединением первоначальных интервалов** (путем их укрупнения) и **долевой перегруппировкой** (на основе закрепления за каждой группой определенной доли единиц совокупности).

Рассмотрим приемы вторичной группировки на примере.

Пример 3.4.

Произвести укрупнение интервалов на основе данных таблицы 3.12:

Таблица 3.12

Распределение магазинов по товарообороту

Группы магазинов по размеру товарооборота за IV квартал, тыс. руб.	Число магазинов	Товарооборот за IV квартал, тыс. руб.
До 10		93
10–15	8	112
15–20	15	200
20–30	3	68
30–50	9	378
50–60	7	385
60–70	3	180
70–100	8	600
100–200	22	2400
Свыше 200	12	3744
Итого	100	8160

Приведенная группировка недостаточно наглядна, потому что не показывает четкой и строгой закономерности в изменении товарооборота по группам.

Уплотним ряды распределения и сформируем шесть групп. Новые группы образованы путем суммирования первоначальных групп.

Таблица 3.13

Распределение магазинов по товарообороту с укрупненными интервалами

Группы магазинов по размеру товарооборота за IV квартал, тыс. руб.	Число магазинов	Товарооборот за IV квартал, тыс. руб.	Товарооборот в среднем на 1 магазин, тыс. руб.
До 10	15	93	
10–20	21	312	6,2
20–50	12	446	37,1
50–100	18	1165	64,8
100–200	22	2400	109,0
Свыше 200	12	3744	312,0
Итого	100	8160	81,6

Совершенно четко видно, чем крупнее магазины, тем выше уровень товарооборота.

Пример 3.5.

Имеются следующие данные о распределении колхозов по числу дворов (табл. 3.14).

Таблица 3.14

Распределение колхозов по числу дворов для двух районов

№ п/п	Группы колхозов по числу дворов	Удельный вес колхозов группы в процентах к итогу	Группы колхозов по числу дворов	Удельный вес колхозов группы в % к итогу
1	До 100	4,3	До 50	1,0
2	100–200	18,4	50–70	1,0
3	200–300	19,5	70–100	2,0
4	300–500	28,1	100–150	10,0
5	Свыше 500	29,7	150–250	18
			250–400	21
			400–500	23
			Свыше 500	24
	Итого	100	Итого	100

Эти данные не позволяют провести сравнение распределения колхозов в 2-х районах по числу дворов, так как в этих районах имеется различное число групп колхозов. Необходимо ряды распределения привести к **сопоставимому виду**.

За основу сравнения необходимо взять распределение колхозов 1 района. Следовательно, по второму району надо произвести вторичную группировку, чтобы образовать такое же число групп и с теми же интервалами, как и в первом районе. Получим следующие данные (табл. 3.15).

Таблица 3.15

Результат вторичной группировки

Группы колхозов по числу дворов	Удельный вес колхозов группы в % к итогу		Расчеты
	I район	II район	
До 100	4,3	4,0	1+1+2=4
100–200	18,4	19,0	10+9=19
200–300	19,5	16,0	9+7=16
300–500	28,1	37,0	21–7=14, 14+23=37
Свыше 500	29,7	24,0	24
Итого	100,0	100,0	

Для определения числа колхозов, которые надо взять из пятой группы во вновь образованную, условно примем, что это число колхозов должно быть пропорционально удельному весу отобранных дворов в группе.

Определяем удельный вес 50 дворов в пятой группе:

$(50\times18) / (250 - 150) = 9$

Определяем удельный вес 50 дворов в шестой группе:

$(50\times21) / (400 - 250) = 7$ и т.д.

Рассмотрим еще один пример приведения двух группировок с различными интервалами к единому виду для распределения акционеров двух районов области по размеру дивидендов на одну акцию в 1996 г. (по условным данным табл. 3.16).

Таблица 3.16

Распределение акционеров двух районов по размеру дивидендов на одну акцию

Первый район			Второй район		
№ группы	Группы акционеров по размеру дивидентов, тыс. руб.	Удельный вес акционеров группы, % (к итогу)	№ группы	Группы акционеров по размеру дивидентов, тыс. руб.	Удельный вес акционеров группы, % (к итогу)
1	10–40	18	1	10–60	10
2	40–80	12	2	60–120	20
3	80–120	40	3	120–200	40
4	120–160	25	4	200–300	30
5	160–200	5	–	–	–
	Итого	100		Итого	100

Приведенные данные не позволяют сравнить распределение акционеров двух районов по размеру дивидендов на одну акцию, так как в этих районах имеется различное число групп акционеров, и, кроме того, различны величины интервалов.

Необходимо ряды распределения привести к сопоставимому виду. За основу сравнения возьмем структуру распределения акционеров второго района (как наиболее крупную). Следовательно, по первому району нужно произвести вторичную группировку или перегруппировку акционеров, образовав такое же число групп и с теми же интервалами, как во втором районе

В результате перегруппировки получаем следующие сопоставимые данные, характеризующие распределение акционеров двух районов по размеру дивидендов на одну акцию (табл. 3.17 на стр. 42).

Анализ сопоставимых данных вторичной группировки позволяет сделать вывод о том, что акционеры второго района имеют более высокие размеры дивидендов (120 тыс. руб. и более на одну акцию выплачивают 70% акционеров этого района, а в первом районе – только 30%).

Таблица 3.17

Вторичная группировка акционеров по размеру дивидендов на одну акцию (группировка единая)

№ группы	Группы акционеров по размеру дивидендов на акцию, тыс. руб.	Удельный вес акционеров группы, % к итогу		Расчет
		Второй район	Первый район	
1	10–60	10	24	18+0,5×12=24
2	60–120	20	46	0,5×12+40=46
3	120–200	40	30	25+5=30
4	200–300	30	–	–
	Итого	100	100	100

Контрольные вопросы

1. Понятие метода группировок при анализе статистических данных?
2. Что понимается под группировочными таблицами?
3. Какие бывают виды группировок?
4. Сущность группировки по количественному признаку?
5. Как определяется оптимальное число групп и величина интервала группировки?
6. Какие бывают вторичные группировки?

ГЛАВА 4. ГРАФИЧЕСКОЕ ИЗОБРАЖЕНИЕ СТАТИСТИЧЕСКИХ ДАННЫХ

4.1. Понятие о графике. Основные элементы графика

Графический метод – это метод условных изображений статистических данных при помощи геометрических фигур, линий, точек и разнообразных символических образов.

Главное достоинство графиков – **наглядность**. При правильном построении графика статистические показатели привлекают к себе внимание, становятся выразительными, лаконичными и запоминающимися.

Для построения графика необходимо знать, для каких **целей составляется график**, изучить исходный материал и владеть методикой графических изображений.

Современную науку невозможно представить себе без применения графических методов, настолько прочно вошли они в арсенал средств научного общения и в методику научного исследования.

Особое место графические методы занимают в статистике и экономике, имеющих дело с большими комплексами цифр, сведенных в громоздкие таблицы. Здесь графические методы помогают, прежде всего, описанию, а затем и анализу этих данных. С помощью графиков легко выявить и наглядно представить закономерности, которые часто трудно бывает уловить в сложных статистических таблицах. При этом используются различные графики, многообразие видов которых обусловлено различиями в их статистическом содержании, способах построения и широтой круга изображаемых ими общественных явлений и процессов.

Графиками в статистике называются условные изображения числовых величин и их соотношений в виде различных **геометрических образов – точек, линий, плоских фигур и т.п.** Использование графиков для изложения статистических показателей позволяет придать последним наглядность и выразительность, облегчить их восприятие, а во многих случаях помогает уяснить сущность изучаемого явления, его закономерности и особенности, увидеть тенденции его развития, взаимосвязь характеризующих его показателей.

Каждый **график состоит** из **графического образа** и **вспомогательных элементов**.

Графический образ – это совокупность точек, линий и фигур, с помощью которых изображаются статистические данные. Эти знаки образуют собственно языковую ткань графика, его основу.

Вспомогательными элементами графика являются:

Поле графика – это пространство, на котором размещаются образующие график геометрические фигуры. Размер графика зависит от его назначения. Предназначенные аудитории (лекционного зала, цеха) графики изготавливаются в формате чертежного листа, т.е. размером 814х1152 мм, небольшие графики – 203х288 мм и др. Соотношение сторон поля в большинстве графиков берется от 1:1,33 до 1:1,50.

Размер поля графика и пропорции его сторон в каждом случае определяются исполнителем, возможны отступления от приведенных соотношений. Однако не следует строить графики, сильно удлиненные в горизонтальном или вертикальном направлении. Такие графики **эстетически невыразительны.**

Геометрические знаки или образы – это многообразные знаки, с помощью которых изображают статистические величины. В статистических графиках в качестве геометрических знаков используются **точки, отрезки прямой линии, квадраты, прямоугольники, а также фигуры в виде рисунков или силуэтов изображаемых предметов**.

Знак составляет основу графика, его язык. Одни и те же данные графически должны быть изображены различными знаками в зависимости от того, какой аспект явления должен подчеркнуть график, на что нацелить внимание его читателя. Например, чтобы показать тенденцию развития промышленного производства за ряд лет, целесообразно использовать отрезки прямых линий, соединяющих точки на поле графика. Направление прямых линий вверх (вниз) будет свидетельствовать о росте (снижении) промышленного производства, а угол наклона будет характеризовать интенсивность этого непрерывного процесса за отдельные периоды. Если же ставится задача – показать изменение объемов промышленного производства, тогда эти же данные следует изобразить в виде прямоугольных столбиков различной высоты.

Масштабные ориентиры статистических графиков **–** это масштаб, масштабные шкалы и масштабные знаки.

Масштаб – это условная мера перевода числовой величины в графическую и обратно. При построении графика масштаб должен быть таким, чтобы подлежащие нанесению на график данные поместились на поле графика. На вертикальной шкале графика должна быть нулевая точка. В тех случаях, когда минимальное значение признака намного выше нуля, нецелесообразно вести отсчет от нулевой точки, так как поле графика будет заполнено неравномерно. В таких случаях рекомендуется делать разрыв вертикальной шкалы.

Масштабная шкала – это линия, разделенная на отрезки точками. Наиболее часто в статистических графиках используются располагающиеся по осям координат равномерные **прямолинейные** масштабные шкалы, в которых отрезки между двумя соседними точками (графические интервалы) строго пропорциональны размерам и периодам времени изображаемых на графике данных. В секторных диаграммах используются **криволинейные** масштабные шкалы. Площадь круга делится на сектора пропорционально изображаемым на графике числам. Различают также шкалы **равномерные** и **неравномерные.** Прямолинейные шкалы в основном равномерные. Неравномерным шкалам соответствуют неравные числовые значения. Так, в логарифмической шкале графические отрезки пропорциональны не числам, а их логарифмам.

Масштабные знаки – это эталоны величин, изображаемых на графике в виде отдельных графических знаков: квадратов, кругов, рисунков, силуэтов и др. Ими пользуются для сравнения графических знаков со знаком-эталоном.

Экспликация графика – это пояснения, раскрывающие содержание графика: заголовок графика, единицы измерения, условные обозначения.

Пояснительные надписи к отдельным элементам графика могут быть помещены либо на поле графика, либо в форме условных обозначений за пределами поля графика.

4.2. Основные виды графиков

Графики, применяемые для изображения статистических данных, чрезвычайно разнообразны. В данной главе будут рассмотрены наиболее часто применяемые в статистической практике графики.

Графические изображения используются чаще всего для сравнения между собой статистических величин, определения роли отдельных факторов во всей их совокупности, изучения структуры и структурных сдвигов, связи между признаками, изменения явлений во времени, определения степени распространения явления в пространстве, и т.д.

Основными элементами графиков, изображающих количественные соотношения, являются **шкала, масштаб, оси координат и числовая (координатная) сетка**.

График должен иметь **заглавие,** отражающее содержание изображаемого явления, **время и место**, к которому относятся данные, и **расшифровку** условных обозначений. Для большей наглядности графика применяют **различную штриховку, окраску** и т.д.

По способу построения статистические графики подразделяются на **диаграммы, картограммы и картодиаграммы** (рис. 4.1).

Столбиковые диаграммы являются наиболее простым видом диаграмм. При их построении данные изображаются в виде столбиков от числовых значений изображаемых величин по определенному масштабу. Примером применения столбиковой диаграммы могут служить данные о производстве автомобилей одной из фирм (рис. 4.2).

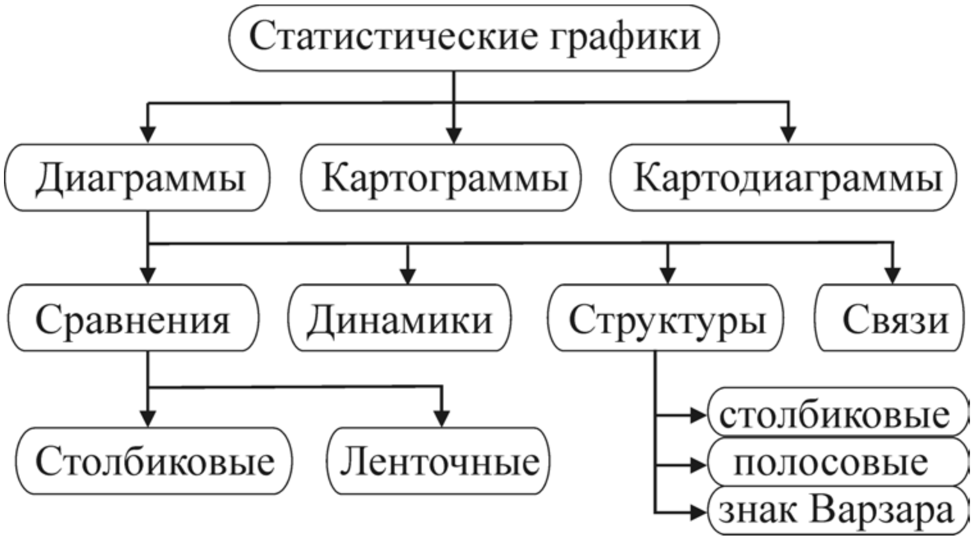


Рис. 4.1.

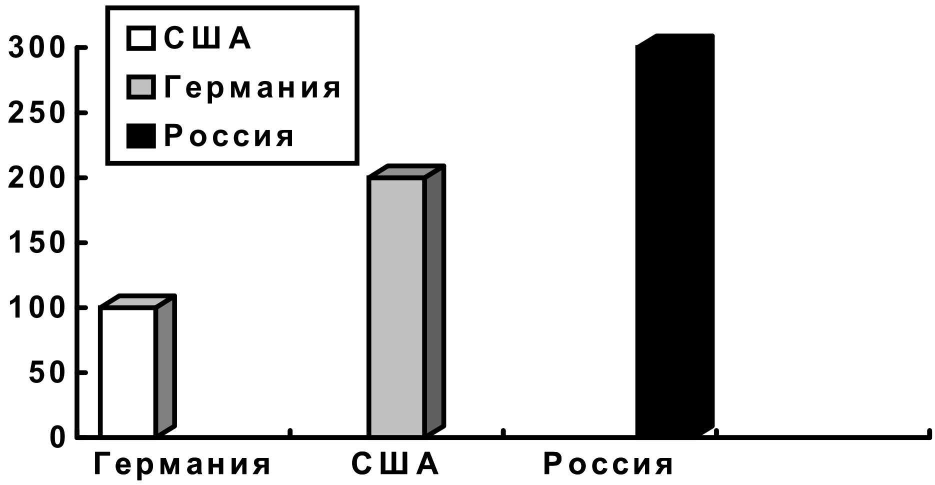


Рис. 4.2. Столбиковая диаграмма сравнения

На масштабной шкале проставляются круглые или округленные значения изображаемых величин. Такая диаграмма называется простой, так как столбики не имеют внутренних долей. Если же они делятся на части, то диаграмма называется **сложной** (рис. 4.3).

Разновидностью столбиковых диаграмм являются **ленточные диаграммы**. Они изображают размеры признака в виде расположенных по горизонтали прямоугольников одинаковой ширины, но различной длины, пропорционально изображаемым величинам. Начало полос должно находиться на одной и той же вертикальной линии (рис. 4.4.).

В ленточных диаграммах удобнее, чем в столбиковых, располагать надписи. Ленточную диаграмму используют также для характеристики отдельных частей (структуры) совокупности.

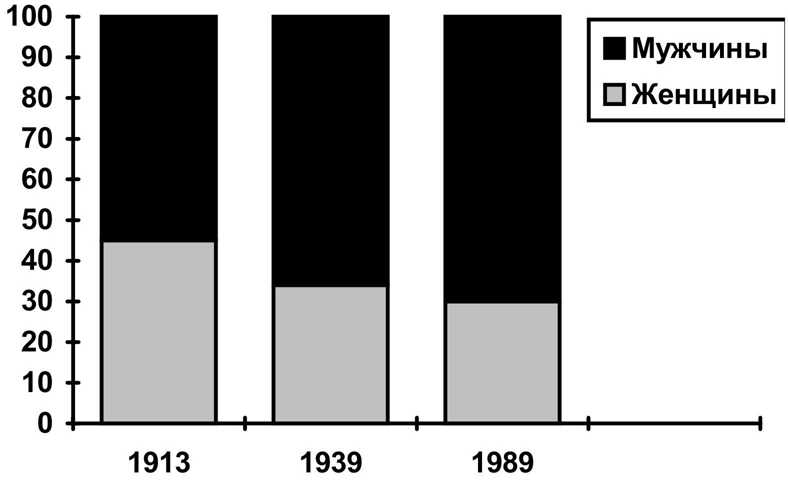


Рис. 4.3. Структурно-столбиковая (сложная) диаграмма

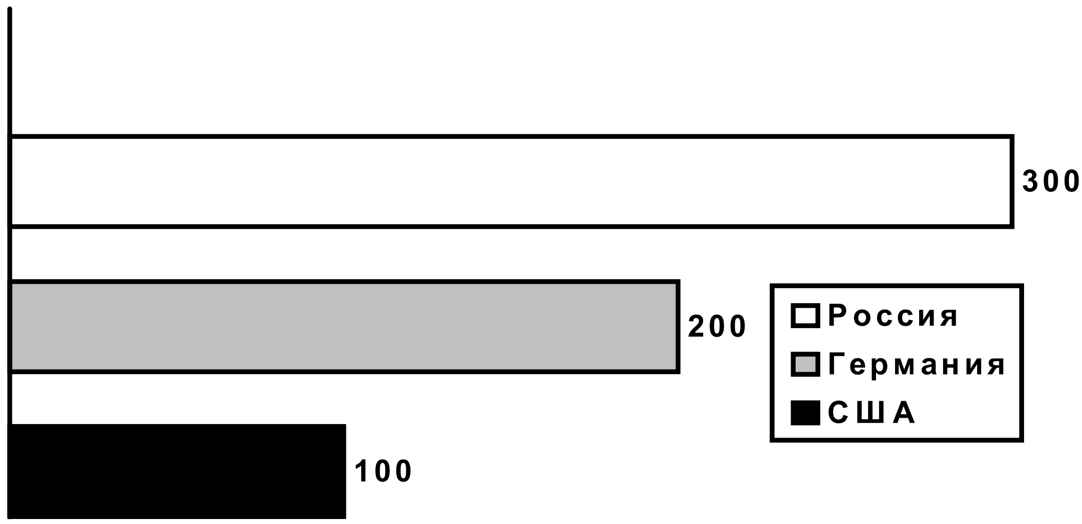


Рис. 4.4. Ленточная (полосовая) диаграмма сравнения

Квадратные и круговые диаграммы относятся к типу **плоскостных диаграмм**. Они представляют собой различные по размерам квадраты или круги, площади которых пропорциональны величине изображаемых статистических данных. Если числа обозначить буквой d, то стороны квадратов будут равны $\sqrt{d}$. Известно, что площадь круга $S = \pi R^2$. Поэтому радиусы отдельных кругов будут равны $\sqrt{S}$, т.е. корню квадратному из значений изображаемых величин.

Недостаток квадратных и круговых диаграмм заключается в том, что они менее наглядны, чем столбиковые диаграммы, так как сравниваются площади, а не высоты, и строить их несколько сложнее.

Нередко состав, структура того или иного явления изображается с помощью кругов, разделенных на **сектора**, пропорциональные долям частей явлений. Круг принимается за целое (100%) и разбивается на сектора, дуги которых пропорциональны значениям отдельных частей изображаемых величин. Дуга каждого сектора круга рассчитывается по формуле $\frac{360^{0} \times d}{100}$, где 360° – весь круг (100%), а d – величина изображаемого явления в процентах. Такие диаграммы называются **секторными** (рис. 4.5).

Рис. 4.5. Структурно-секторная диаграмма
(состав населения в РФ в 1998 г.)

Секторные диаграммы следует применять лишь в тех случаях, когда совокупность делится не более чем на 4–5 частей, а также при условии значительных различий сравниваемых структур, иначе они теряют свою выразительность. Так, диаграмма, представленная на рис. 4.5 наглядно показывает рост удельного веса городского населения и соответствующее сокращение удельного веса сельского населения.

Наиболее распространенным видом диаграмм являются **линейные.** Чаще всего они используются для изображения динамических рядов и при изучении связи между явлениями. При построении линейных диаграмм применяют координатную или числовую сетку. На оси абсцисс системы прямоугольных координат на равном расстоянии друг от друга наносятся точки, соответствующие числу членов динамического ряда, а на оси ординат – показатели по принятому масштабу. После этого наносят данные и, соединив концы перпендикуляров, получают ломаную линию, характеризующую изображаемый динамический ряд (рис. 4.6).

Общий вид графика зависит от правильного соотношения масштабов на осях абсцисс и ординат. В противном случае колебания будут либо малозаметными, либо слишком резкими. Если данные относятся к различным периодам времени, интервалы между ними при нанесении

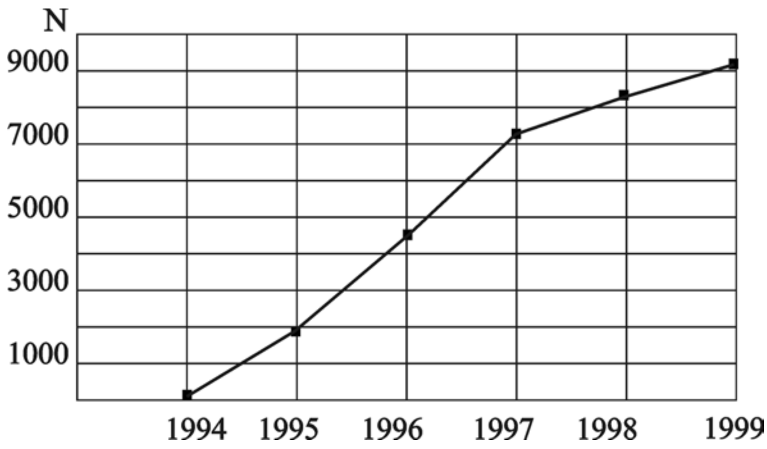


Рис. 4.6. Сборка компьютеров фирмой

на оси абсцисс должны быть пропорциональны длительности периодов. При помощи линейных диаграмм можно выражать одновременно ряд показателей, что дает возможность сравнивать их друг с другом.

Полигон – ломаная кривая, строящаяся на основе прямоугольной системы координат, когда по оси X откладываются значения признака, а по оси Y – частоты.

Гладкая кривая, соединяющая точки – эмпирическая плотность распределения.

Знаки Варзара, предложенные русским статистиком В.Б. Варзаром (1851–1940), применяют в тех случаях, когда нужно сравнить величины, представляющие собой произведение двух сомножителей, и показать роль каждого из них в формировании этой величины.

Пример 4.1.

Требуется сравнить эффективность двух сельскохозяйственных предприятий. Посевная площадь пшеницы в сельхозпредприятии №1 составила 200 га, в сельхозпредприятии №2 – 800 га. Урожайность пшеницы в этих предприятиях равна соответственно 40 и 10 ц с 1 га. Оба предприятия собрали урожай зерновых 8 000 ц (рис. 4.7).

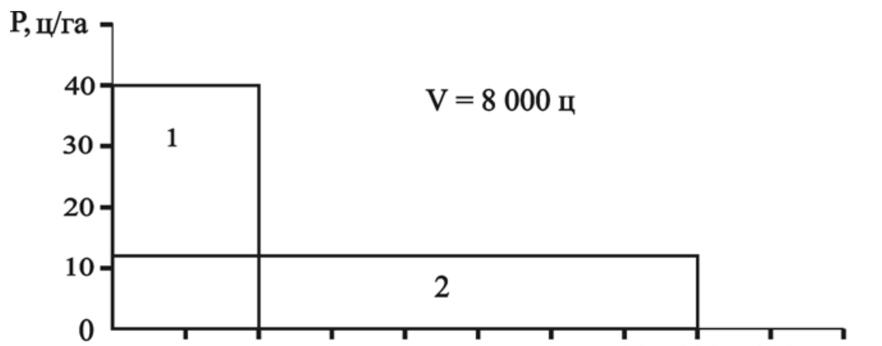


Рис. 4.7. Эффективность двух сельскохозяйственных предприятий

Можно видеть, что, несмотря на одинаковую величину урожая, себестоимость центнера зерновых в предприятии №1 будет ниже т.к. посевные площади меньше, техники меньше, горюче смазочных материалов меньше, рабочих меньше и т.д. Эффективность предприятия №1 выше.

Пример 4.2.

Валовой сбор с/х культуры равен произведению урожайности и посевной площади (рис. 4.8).

На этом графике можно сравнить между собой:

- урожайность (по длине основания);
- посевные площади (по длине боковой стороны);
- валовой сбор (по площади прямоугольника).

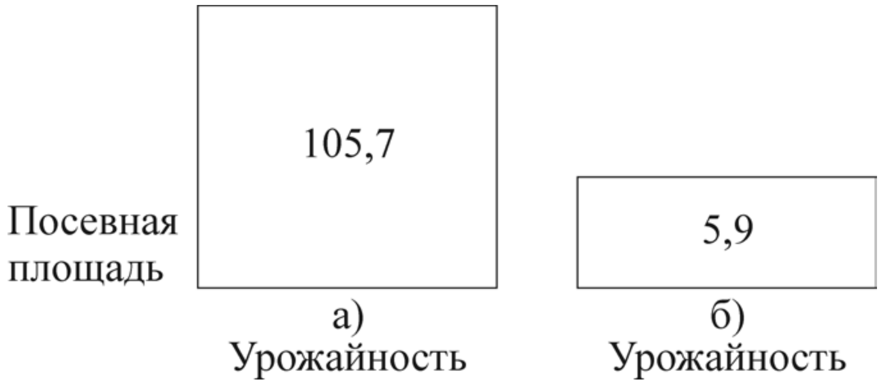


Рис. 4.8. Валовой сбор, урожайность и посевные площади в 1977г.:
а – зерновых культур, б – подсолнечника

Статистическая карта – вид графика, который иллюстрирует содержание статистических таблиц, где подлежащим является административное или географическое деление совокупности. На лист изображения наносится контурная географическая карта, отражающая деление совокупности на группы. Статистическая карта называется картограммой, вся информация на ней отображается в виде штриховки, линий, точек, окраски, отражающих изменение какого-либо показателя.

На картодиаграмме, на фоне карты, присутствуют элементы диаграммных фигур. Преимущество картодиаграммы перед диаграммой состоит в том, что она не только дает представление о величине изучаемого показателя на различных территориях, но и изображает пространственное размещение изучаемого показателя.

В зависимости от формы применяемых графических образов статистические графики могут быть **точечными, линейными, плоскостными и фигурными**.

В **точечных** графиках в качестве графических образов применяется совокупность точек.

В **линейных** графиках графическими образами являются линии.

Для **плоскостных** графиков графическими образами являются геометрические фигуры: прямоугольники, квадраты, окружности.

Гистограммы. При обработке и отображении экспериментальных данных, в которых изучаемый признак может принимать любое значение из некоторого интервала, используют следующие способы представления данных:

- гистограммы;
- полигон частот;
- полигон накопленных частот (кумулята).

Гистограмма состоит из примыкающих друг к другу прямоугольников, изображенных на координатной сетке.

Существует несколько случаев построения гистограмм.

Равные интервалы группировки данных.

Рассмотрим на примере.

Имеются данные о группировке рабочих по стажу лет, данные оформлены в виде таблицы 4.1.

Таблица 4.1

Распределение рабочих по стажу

Группы рабочих по стажу лет	Число рабочих	Накопительные частоты
1–3	4	4
3–5	12	16
5–7	15	31
7–9	10	41
9–11	9	50
Итого	50	

На рисунке откладываются прямоугольники с высотой, прямо пропорциональной частоте данного интервала (рис. 4.9).

Наибольшее число рабочих имеет стаж работы от 5 до 7 лет.

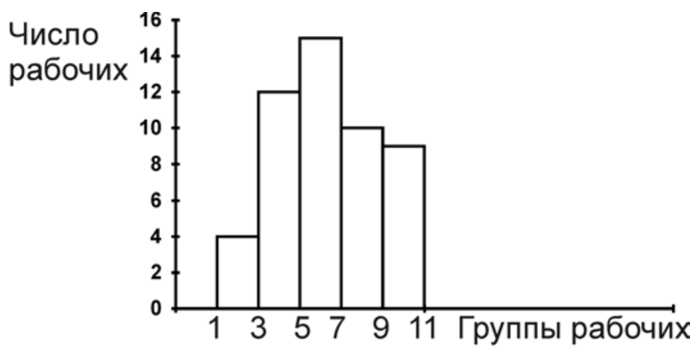


Рис. 4.9.

Открытые крайние интервалы группировки. Предположим, что первый и последний интервалы открытые. В этих случаях используется стандартный прием. Условно ширина первого открытого интервала принимается равной ширине следующего интервала. Ширина последнего принимается равной ширине предыдущего. В нашем примере гистограмма будет такой же, как на рис. 4.9.

Таблица 4.2

Распределение рабочих по стажу

Группы рабочих по стажу лет	Число рабочих	Накопительные частоты
До 3	4	4
3–5	12	16
5–7	15	31
7–9	10	41
9 и более	9	50
Итого	50	

Неравные интервалы группировки. Предположим, что вместо двух интервалов (3–5 и 5–7) стал один. Интервал стал шире в два раза, а высота стала не 27, а 13,5, с тем, чтобы площадь прямоугольника не менялась. Высоту прямоугольника можно определить по формуле – n/h, n – частоты попадания (27), а h – количество интервалов (2) (рис. 4.10).

Таблица 4.3

Распределение рабочих по стажу

Группы рабочих по стажу лет	Число рабочих	Накопительные частоты
1–3	4	4
3–7	27	31
7–9	10	41
9 и более	9	50
Итого	50	

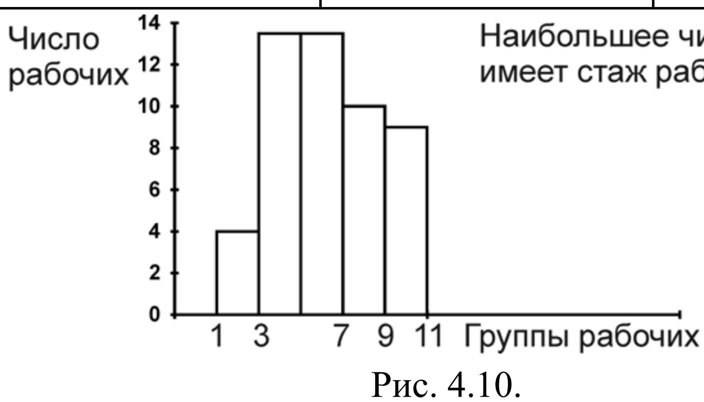


Рис. 4.10.

Полигон накопительных частот. В данном случае для построения используются накопительные частоты. Построим по данным таблицы 4.1.

Полигон частот – ломаная линия, соединяющая точки, соответствующие срединным значениям интервалов группировки и частотам интервалов.

Полигон частот получается из гистограммы, если соединить середины вершин прямоугольников ломаной линией (рис. 4.11).

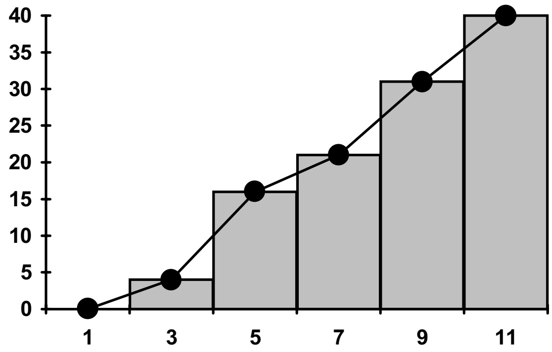


Рис. 4.11. Распределение рабочих по стажу

4.3. Статистические карты

В ряде случаев в статистических исследованиях применяют **картограммы и картодиаграммы.** Их применение связано с более наглядным изображением экономико-географической характеристики изучаемых явлений. Они показывают размещение изучаемого явления по определенной территории – району, муниципальному округу, области и т.п.

Картограмма – это такая статистическая карта, на которой распределение изучаемого признака по территории отображено условными знаками (точками, штриховкой и т.д.), соответствующими определенным интервалам значений величины этого признака. Эти знаки размещаются внутри контура каждого района соответственно среднему значению показателя по данному району. Картограмма применяется в тех случаях, когда возникает необходимость показать территориальное распределение какого-нибудь одного статистического признака между отдельными для выявления закономерности этого распределения. Картограммы бывают фоновыми и точечными. Наибольшее распространение имеют фоновые картограммы, которые делятся на **хорохроматические карты и карты хороплеты** (от греч. «хорос» – пространство).

Хорохроматические карты – это картограммы, которые относятся к наиболее легко воспринимаемым. Они лишь показывают, где распространено то или иное явление или объект и на какой площади, и не содержит иных количественных сведений. Такие картограммы применяются главным образом для показа использования земель. Каждый вид использования территории передается или определенной закраской, или определенной штриховкой. Поэтому хорохроматические карты иногда называют «картами многокрасочного фона». **Карты хороплет** – это своего рода распространение гистограмм в область пространственной организации, поскольку совокупность сгруппированных данных размещена здесь с помощью карты по ареалам.

При составлении карты хороплет выполняются три операции.

Во-первых, всю совокупность данных, исходя из ее содержательного смысла, расчленяют на группы.

Во-вторых, все группы привязывают к ячейкам территории, по которым проводился первоначальный сбор данных.

В-третьих, выделенные группы показывают на карте в картографической корректной форме.

Часто вместо раскраски применяется штриховка различной интенсивности. Такая картограмма наглядно показывает, например, урожайность зерновых культур по районам. При помощи фоновых диаграмм можно изображать размещение промышленности, отдельных видов производимой продукции, посевных площадей, численности скота и т.д. Чем больше групп, тем точнее изображение, но большое число групп создает пестроту и снижает наглядность. Поэтому практически лучше всего применять не более четырех-пяти тонов.

Сущность точечной картограммы заключается в том, что символами графического изображения статистических данных являются точки, размещенные в пределах определенных территориальных границ. Каждой точке, нанесенной на картограмму, условно придается конкретное числовое значение, что позволяет использовать ее как инструментарий прямого счета. Например, имеются четыре условных региона с добычей угля в 200, 500, 1000 и 1400 тыс. т в год. При составлении картограммы примем точку за 50 тыс. т и нанесем на каждый регион соответствующее количество точек.

Картодиаграмма – это сочетание диаграммы с географической картой. В качестве изобразительных знаков в картодиаграммах используются диаграммные фигуры, которые размещаются на контуре географической карты. **Картодиаграммы дают возможность графически отразить более сложные статистико-географические построения, чем картограммы, и изобразить их более точно.** Так, при помощи картодиаграммы можно выразить пространственное (географическое) распределение структур изучаемых статистических совокупностей, особенности каждого региона (в нашем примере условного) и т.д.

В качестве диаграммных знаков в картодиаграмме часто используются различные геометрические фигуры, особенно круги, которые наиболее просты и удобны для выражения сравниваемых количественных показателей на карте. Такова, например, структурная и секторная картодиаграмма, характеризующая порегионные различия в структуре посевных площадей.

Кроме рассмотренных видов диаграмм в статистической практике встречаются и другие, более сложные, графические объекты статистических данных.

Подводя итоги раздела, посвященного применению графических методов в статистике, следует особо подчеркнуть то, что графические методы важны как в предварительном анализе данных, так и в представлении окончательных выводов.

Правильно построенный график делает статистическую информацию более наглядной, запоминающейся и удобно воспринимаемой.

В производственной и коммерческой деятельности графический метод находит широкое применение для иллюстрации сложившегося положения с финансовыми, материальными, информационными и трудовыми ресурсами, а также положения дел на рынке товаров и услуг, конъюнктуры спроса и предложения, рекламы продуктов.

Наряду с аналитическим значением, наиболее важную роль статистические, графики играют в обобщении статистической информации.

Графический метод в статистике является продолжением и дополнением табличного метода. График позволяет сравнительно легко обнаружить на глаз ошибки расчетов, которые в табличной форме не были так заметны. Незамеченными при чтении таблиц могут оставаться не только ошибки, но какие-то тенденции. Все это обнаруживается на графике. Статистические графики отражают целостную картину изучаемого явления, его обобщенное представление.

При графическом изображении статистических данных становится более выразительной сравнительная характеристика изучаемых показателей, отчетливее проявляется тенденция развития изучаемого явления, лучше видны основные взаимосвязи.

Таким образом, обладая такими качествами, как наглядность, выразительность и запоминаемость, графический метод занимает важное значение среди других методов статистики.

Контрольные вопросы

1. Цели и задачи при графическом изображении статистических данных?
2. Понятие и основные элементы графика?
3. Каковы основные виды графиков?
4. Какие бывают виды структурных диаграмм?
5. Понятие гистограммы и полигона?
6. Что понимается под статистическими картами?
7. Сущность картограммы и картодиаграммы?

ГЛАВА 5. АБСОЛЮТНЫЕ И ОТНОСИТЕЛЬНЫЕ ВЕЛИЧИНЫ

5.1. Абсолютные величины

В итоге сводки статистических данных получают обобщающие показатели, в которых отражаются результаты познания количественной стороны изучаемых явлений. Исходной, первичной формой выражения статистических показателей, отражающих уровень развития явления, служат **абсолютные величины**.

Абсолютными в статистике называются суммарные обобщающие показатели, характеризующие размеры (уровни, объемы) общественных явлений в конкретных условиях места и времени. Они характеризуют экономическую мощь страны и социальную жизнь населения (ВВП, ВНП, НД, реальные располагаемые денежные доходы населения, объемы промышленного и сельскохозяйственного производств, объем выпуска важнейших видов продукции). Например, численность населения Российской Федерации на 1 января 1997 г. составила 147,5 млн чел.; в 1995 г. добыто 307 млн т нефти (включая газовый конденсат), 595 млрд м^3 естественного газа и т.д., за 1996 г. ВВП России составил в текущих ценах 2256 трлн руб., промышленностью страны за этот период произведено продукции (работ, услуг) в действующих ценах на сумму 1414 трлн руб.

Различают два вида абсолютных статистических величин: **индивидуальные и суммарные**.

Индивидуальными называют абсолютные статистические величины, характеризующие размеры признака у отдельных единиц совокупности (например, размер заработной платы отдельного работника, вклада гражданина в определенном банке и т.д.). Они получаются непосредственно в процессе статистического наблюдения и фиксируются в первичных учетных документах.

В отличие от индивидуальных **суммарные абсолютные статистические величины** характеризуют итоговое значение признака по определенной совокупности объектов, охваченных статистическим наблюдением. Они являются суммой количества единиц изучаемой совокупности **(численность совокупности)** или суммой значений варьирующего признака всех единиц совокупности **(объем варьирующего признака).**

Абсолютные статистические величины представляют собой **именованные числа,** т.е. имеют какую-либо **единицу измерения.**

В зависимости от сущности исследуемого социально-экономического явления абсолютные статистические величины выражаются в натуральных, стоимостных и трудовых **единицах измерения**.

Абсолютные статистические величины могут быть **положительными (доходы) и отрицательными (убытки, потери).**

Натуральные единицы измерения в свою очередь могут быть **простыми** (тонны, штуки, метры, литры) и **сложными**, являющимися комбинацией нескольких разноименных величин (грузооборот железнодорожного транспорта выражается в тонна – километрах, производство электроэнергии; – в киловатт-часах, затраты труда – в человеко-часах, человеко-днях). В статистике применяют и абсолютные показатели, выраженные **в условно-натуральных** единицах измерения (например, различные виды топлива пересчитываются в условное топливо, тракторный парк – в эталонные тракторы).

Стоимостные единицы измерения используются, например, для выражения объема разнородной продукции в стоимостной (денежной) форме **– рублях**. В стоимостных единицах выражают валовой выпуск продукции, доходы населения и пр.

При использовании стоимостных измерителей принимают во внимание изменение цен с течением времени. Этот недостаток стоимостных измерителей преодолевают применением «неизменных» или «сопоставимых» цен одного и того же периода.

В **трудовых** единицах измерения (человеко-днях, человеко-часах) учитываются общие затраты труда на предприятии, трудоемкость отдельных операций технологического цикла.

5.2. Относительные величины

Наряду с абсолютными статистическими величинами большое значение в статистике имеют **относительные величины**. В процессе выявления ряда важнейших для социально-экономической жизни вопросов возникает необходимость в изучении структуры явления, соотношения между отдельными его частями, развития во времени.

Относительная величина в статистике – это обобщающий показатель, который представляет собой частное от деления одного **абсолютного** показателя на другой и дает числовую меру соотношения между ними.

Основные условия правильного расчета относительной величины – сопоставимость сравниваемых показателей и наличие реальных связей между изучаемыми явлениями.

Величина, с которой производится сравнение (знаменатель дроби), обычно называется **базой сравнения или основанием.**

В зависимости от выбора базы сравнения относительный показатель может быть представлен в различных долях единицы: десятых; сотых (т.е. процентах – %); тысячных (десятая часть процента называется промилле – $^0/_{00}$); десятитысячных (сотая часть процента называется продецимилле – $^0/_{000}$).

Сопоставимые величины могут быть как **одноименными**, так и **разноименными** (в последнем случае их наименования образуются от наименований сравниваемых величин, например, руб./чел.; ц/га; руб./м^2).

По своему содержанию **относительные величины** подразделяются на **виды**:

1. **относительные величины динамики;**
2. **планового задания;**
3. **выполнения планового задания;**
4. **структуры;**
5. **интенсивности;**
6. **уровня экономического развития;**
7. **координации;**
8. **сравнения.**

Рассмотрим подробнее первые три относительные величины. Введем следующие обозначения. Допустим, имеется два изучаемых года:

2002 г.	2003 г.
$F_{ПП}$	$F_{ТП}$
	P

2002 г. – это предшествующий период. В этом году выпущено (произведено, добыто и т.д.) определенное количество некоторой продукции, которое обозначается через $F_{ПП}$ – факт предшествующего периода (абсолютная величина). 2003 г. – это текущий период. В этом году также выпущено (произведено, добыто и т.д.) определенное количество продукции, которое обозначается через $F_{ТП}$ – факт текущего периода (абсолютная величина). В 2002 г. запланировано выпустить продукции P – план текущего периода по отношению к 2002 г., абсолютная величина (в 2002 г. тоже есть план, но он по отношению к 2001 г.). Тогда:

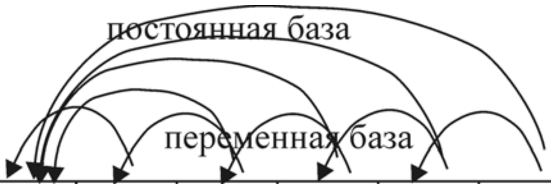
1. Относительная величина динамики ($О_Д$) рассчитывается как отношение факта (уровня) признака в текущем периоде к уровню этого же признака (факта) в предшествующем периоде, т.е. она характеризует изменение уровня какого-либо **явления во времени**

$$O_Д = \frac{F_{ТП}}{F_{ПП}} \tag{5.1}$$

Выбор базы сравнения при исчислении относительных показателей динамики определяется целью исследования. Если исследование проводится в течение ряда лет и в качестве базы используется только факт предшествующего года, то показатель динамики называется показатель с **переменной базой** (цепная динамика). Если же в качестве базы берется только начальный год – показатель динамики с **постоянной базой** (базовая динамика) (рис. 5.1)

Пример. 5.1.

Имеются данные по выпуску бумаги в РФ. Определить динамику с переменной и постоянной базой.



Годы	1992	1993	1994	1995	1996	1997
Бумага млн.т	3603	2882	2215	2771	3120	3651

Рис.5.1.

С переменной базой:

K_1(1993/1992) = 2882/3603 = 0,79 если умножить на 100, то получим в процентах – 79% т.е. в 1993 г. выпуск бумаги уменьшился на 21% по сравнению с 1992 г.;

K_2(1994/1993) = 2215/2885 = 0,77 или 77% – уменьшился на 23%;

K_3(1995/1994) = 2771/2215 = 1,25 или 125% – увеличился на 25%;

K_4(1996/1995) = 3120/2771 = 1,13 или 113% – увеличился на 13%;

K_5(1997/1996) = 3651/3120 = 1,17 или 117% – увеличился на 17%.

С постоянной базой:

M_1(1993/1992) = 2882/3603 = 0,79 или 79% – уменьшился на 2 %;

M_2(1994/1992) = 2215/3603 = 0,61 или 61%

M_3(1995/1992) = 2771/3603 = 0,77 или 77%

M_4(1996/1992) = 3120/3603 = 0,86 или 86%

M_5(1997/1992) = 3651/3603 = 1,01 или 101%

Причем произведение коэффициентов с переменной базой за определенный период равно коэффициенту с постоянной базой за весь период:

$K_1 \times K_2 \times K_3 \times K_4 \times K_5 = 0{,}79 \times 0{,}77 \times 1{,}25 \times 1{,}13 \times 1{,}17 = 1.01 = M_5$

Относительные величины динамики называются **темпами роста.**

2. Относительная величина планового задания ($O_{ПЭ}$) рассчитывается как отношение величины плана к его фактической величине (уровню) в предшествующем периоде

$$O_{ПЗ} = \frac{P}{F_{ПП}} \tag{5.2}$$

3. Относительная величина выполнения плана задания ($О_{ВПЗ}$) представляет собой отношение фактической величины в текущем периоде к запланированной величине

$$O_{ВПЗ} = \frac{F_{ТП}}{P} \qquad (5.3)$$

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

$$О_Д = О_{ПЗ} \times О_{ВПЗ} \qquad (5.4)$$

Пример 5.2.

По региону имеются следующие данные о вводе в эксплуатацию жилой площади (табл. 5.1.):

Определить: 1. Динамику ввода в эксплуатацию жилой площади по каждому виду жилых домов и в целом по региону;
2. Структуру введенной в эксплуатацию жилой площади;
3. Структуру представить в виде графика.

Таблица 5.1

Статистические данные о вводе жилья по региону

Вид жилых домов	Введено в эксплуатацию, тыс. $м^2$	
	2000 г.	2001 г.
Кирпичные	4400	4200
Панельные	2800	2100
Коттеджи	800	2100

Динамика:

$К_{Кр}$ = 4200/4400 = 0,955 или 95,5%

$К_П$ = 2100/2800 = 0,750 или 75%

$К_{Кт}$ = 2100/800 = 2,625 или 262,5%

В целом по региону $К_Р = \dfrac{4200+2100+2100}{4400+2800+800}$ = 1,05 или 105%.

Следовательно: ввод в эксплуатацию жилой площади в кирпичных домах уменьшился на 4,5%, а панельных – снизился на 25%, по коттеджам – увеличился на 162,5%, в целом по региону – возрос на 5%.

Структура введенной площади представлена в таблице 5.2 (стр. 62).

Таблица 5.2

Структура введенной жилой площади

Вид жилых домов	2000 г.		2001 г.	
	введено, тыс. $м^2$	в % к итогу	введено, тыс. $м^2$	в % к итогу
Кирпичные	4400	55	4200	50
Панельные	2800	35	2100	25
Коттеджи	800	10	2100	25
Итого	8000	100	8400	100

Процент к итогу подсчитывается как отношение, например, введенных квартир в кирпичных домах к общему итогу: 4400/8000 = 0,55 или 55%.

Структурная диаграмма представлена на рис. 5.2.

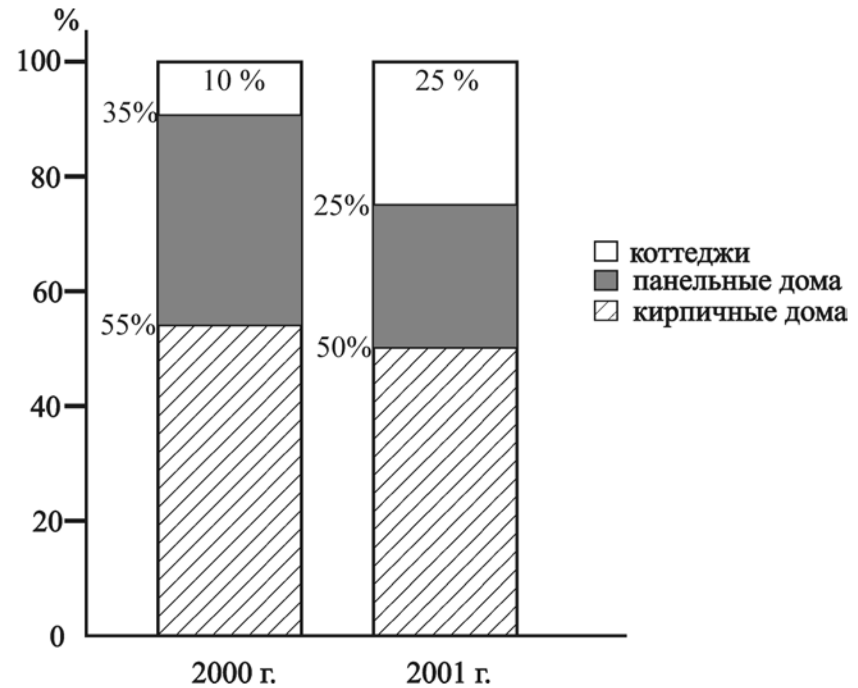


Рис. 5.2. Структурная диаграмма введенных квартир

Пример 5.3.

В прошлом году объем грузооборота по грузовому автотранспортному предприятию составил 210,0 млн ткм. Планом текущего года было предусмотрено довести объем грузооборота до 220,5 млн ткм; фактический объем грузооборота в текущем году составил 229,32 млн ткм.

Определить: 1) относительную величину планового задания по росту грузооборота;

2) относительную величину динамики грузооборота;

3) относительную величину выполнения плана по грузообороту.

Введем обозначения: $F_{ПП} = 210{,}0$; $P = 220{,}5$; $F_{ТП} = 229{,}32$

$O_{ПЗ} = ?$, $O_{Д} = ?$, $O_{ВПЗ} = ?$

Согласно формулам (5.1 – 5.3) проводим расчет:

$O_{Д} = 229{,}32/210 = 1{,}092$ или 109,2% фактически грузооборот увеличился на 9,2%

$O_{ПЗ} = 220{,}5/210 = 1{,}05$ или 105 планом предусмотрено увеличение грузооборота на 5%

$O_{ВПЗ} = 229{,}32/220{,}5 = 1{,}04$ или 104% план грузооборота превышен на 4%.

Пример 5.4.

Планом предусмотрено увеличение годовой производительности труда (ПТ) работников против прошлого года на 4,0%. Фактически против прошлого года производительность труда увеличилась на 6,2%. Определить: процент выполнения плана по уровню производительности труда.

Введем обозначения. Плановым заданием предусмотрено увеличение ПТ на 4%, т.е. это 104% или коэффициент 1,04. Тогда $O_{ПЗ} = 1{,}04$. Фактически ПТ увеличилась на 6,2%, показатель динамики равен 106,2% или 1,062, т.е. $O_{Д} = 1{,}062$. Тогда $O_{ВПЗ} = ?$

Согласно формулы (5.4) $O_{Д} = O_{ПЗ} \times O_{ВПЗ}$ Тогда $O_{ВПЗ} = O_{Д}/O_{ПЗ} = 1{,}062/1{,}04 = 1{,}021$ или 102,1% т.е. плановое задание перевыполнено на 2,1%.

Пример 5.5.

По плану объем продукции в отчетном году должен возрасти против прошлого года на 2,5%. План выпуска продукции перевыполнен на 3,0%.

Определить: фактический выпуск продукции в отчетном году, если известно, что объем продукции в прошлом году составил 25 300 тыс. руб.

Введем обозначения. По плану объем должен возрасти на 2,5%, это 102,5% или $1{,}025 = O_{ПЗ}$. План превышен на 3%, это 103% или $1{,}03 = O_{ВПЗ}$. $F_{ПП} = 25300$. $F_{ТП} = ?$.

Согласно формул (5.1 – 5.4) $O_{Д} = F_{ТП} / F_{ПП}$, или $F_{ТП} = O_{Д} \times F_{ПП}$.

$O_{Д} = O_{ПЗ} \times O_{ВПЗ}$. Тогда $F_{ТП} = O_{Д} \times F_{ПП} = O_{ПЗ} \times O_{ВПЗ} \times F_{ПП} = 1{,}025 \times 1{,}03 \times 25300 = 26710{,}5$ тыс. руб.

$F_{ТП} = 26710{,}5$ тыс. руб.

4. Относительными величинами структуры называются показатели, характеризующие долю отдельных частей изучаемой совокупности во всем ее объеме. Они рассчитываются делением числа единиц (или объема явления) в отдельных частях совокупности на общее число единиц совокупности (или объем явления). Выражаются они простым кратным отношением или в процентах. В качестве примера относительных величин структуры могут служить данные о доле городского населения в общей численности населения России: в 1913 г. – 18%, в 1996 г. – 73%.

5. Относительными величинами интенсивности называют показатели, характеризующие степень распространения или уровень развития того или иного явления в определенной среде. Они вычисляются путем сравнения разноименных величин, находящихся в определенной связи между собой. Эти показатели обычно определяются в расчете на 100, 1000 и т.д. единиц изучаемой совокупности (на 100 га земли, на 1000 человек населения и т.д.) и являются **именованными числами.** Примерами могут служить плотность населения, выражающаяся средним числом жителей на 1 км^2 территории (8,6 чел./км^2 в России в 1996 г.), обеспеченность населения медицинскими кадрами (численность врачей всех специальностей – 44,5 врача на 10000 россиян на начало 1996 г.), возрастные коэффициенты рождаемости (число родившихся в среднем за год на 1000 женщин по возрастным группам).

6. Показатели уровня экономического развития является разновидностью относительных величин интенсивности, характеризуют уровни ВВП, ВНП, НД и др. показатели на душу населения и играют важную роль в оценке развития экономики страны (уровень ВВП Российской Федерации на душу населения в 1995 г. составил 1 100 946 руб. в рыночных ценах).

7. Относительными величинами координации называют показатели, характеризующие соотношение отдельных частей целого между собой. Вычисление этого вида показателей производится путем деления одной части целого на другую часть целого. Таким образом, относительные показатели координации являются разновидностью относительных показателей интенсивности, с той лишь разницей, что они показывают степень распространения, развития разнородных признаков одной и той же совокупности (целого). В зависимости от поставленной задачи тот или иной признак может быть принят за базу. Поэтому для одной и той же совокупности можно исчислить несколько относительных показателей координации.

8. Относительными величинами сравнения называют показатели, представляющие собой частное от деления одноименных абсолютных статистических величин, характеризующих разные объекты (предприятия, фирмы, районы, области, страны и т.д.) и относящихся к одному и тому же периоду (или моменту) времени. Например, соотношение между уровнями себестоимости определенного вида продукции, выпущенной на двух предприятиях, между уровнями производительности труда в разных странах (при одинаковой методике счета).

Научная ценность относительных показателей высока, но их нельзя рассматривать в отрыве от абсолютных показателей, соотношения которых они выражают, иначе они не смогут точно характеризовать изучаемые явления.

Пользуясь в анализе относительными величинами, необходимо показать, какие абсолютные показатели за ними скрываются. В противном случае можно прийти к неправильным выводам. Например, при сравнении двух абсолютных статистических величин 2 тыс. и 5 тыс. руб. получили относительную – 40%, т.е. 2:5x100. Тот же результат получим, сравнивая 200 тыс. и 500 тыс. руб. Но абсолютное значение одного процента, например второго показателя, в том и другом случае будет разным: в первом – оно составит 50, во втором – 5000 руб. Таким образом, лишь комплексное применение абсолютных и относительных показателей выступает как важное средство информации и анализа самых различных явлений социально-экономической жизни.

Контрольные вопросы

1. Понятие абсолютной величины?
2. Понятие относительной величины?
3. Основные виды относительной величины?
4. Сущность относительной величины динамики, планового задания и выполнения планового задания?
5. Сущность относительной величины структуры, интенсивности и уровня экономического развития?
6. Сущность относительной величины координации и сравнения?

ГЛАВА 6. СРЕДНИЕ ВЕЛИЧИНЫ

6.1. Роль и значение средних величин

В результате **группировки** единиц совокупности по величине варьирующего признака получают **ряды распределения** – первичную характеристику массовой статистической совокупности. Чтобы охарактеризовать такую совокупность в целом, часто пользуются средней величиной.

Средняя – это один из распространенных приемов обобщений. Правильное понимание сущности средней определяет ее особую значимость в условиях рыночной экономики, когда средняя через единичное и случайное позволяет выявить общее и необходимое, выявить тенденцию закономерностей экономического развития.

Средняя величина – это обобщающая характеристика однородной совокупности явлений по определенному признаку.

Метод средних величин заключается в замене **индивидуальных значений** варьирующего признака единиц наблюдения, т.е. в замене $\mathbf{x_1, x_2, x_3, ..., x_n}$ некоторой уравненной величиной $\mathbf{x_{ср}}$. Например, индивидуальная выработка у 5 рабочих за месяц составила 135, 141, 153, 159, 162 детали. Чтобы получить среднюю выработку одного рабочего за месяц, нужно сложить эти индивидуальные показатели и полученную сумму разделить на 5:

$$x_{ср} = \bar{x} = \frac{135+141+153+159+162}{5} = \frac{750}{5} = 150 \text{ деталей}$$

Как видно из приведенного примера, средняя выработка не совпадает ни с одной из индивидуальных, так как ни один рабочий не сделал 150 деталей. Но если мы представим себе, что каждый рабочий выработал по 150 деталей, то общая сумма деталей не изменится, а будет также равна 750. Следовательно, средняя, заменяя фактическое значение отдельных, индивидуальных показателей, не может изменить размер всей суммы величин исследуемого признака.

Средняя величина является отражением значений изучаемого признака, следовательно, измеряется в той же размерности, что и этот признак.

Каждая средняя величина характеризует изучаемую совокупность по какому-либо одному признаку. Чтобы получить полное и всестороннее представление об изучаемой совокупности по ряду существенных признаков, в целом необходимо располагать системой средних величин, которые могут описать явление с разных сторон.

В экономических исследованиях и плановых расчетах применяются две категории средних:

- степенные средние;
- структурные средние.

6.2. Виды степенных средних величин и их вычисление

Существуют следующие степенные средние:

- средняя арифметическая ($\overline{x}_a, \overline{x}_{af}$);
- средняя геометрическая ($\overline{x}_{gm}, \overline{x}_{gmf}$);
- средняя гармоническая ($\overline{x}_{gr}, \overline{x}_{grf}$);
- средняя квадратическая ($\overline{x}_{kv}, \overline{x}_{kvf}$).

6.2.1. Средняя арифметическая простая $\overline{x}_a$ и взвешанная $\overline{x}_{af}$

Если имеется несколько различных индивидуальных величин одного и того же вида и надо вычислить среднюю, то необходимо найти сумму всех индивидуальных величин и поделить получаемую сумму на их число.

Обозначив индивидуальные значения признака через $\mathbf{x_1}$, $\mathbf{x_2}$ $\mathbf{x_3}$... $\mathbf{x_n}$ число индивидуальных величин – **n**, среднюю – $\overline{x}_a$, а сумму – знаком **Σ**, можно записать, что вычисленная таким образом средняя называется **средней арифметической простой.** Она равна частному от деления суммы индивидуальных значений признака на их количество.

Пример 6.1

В таблице 6.1 приведено распределение магазинов некоторой фирмы по торговой площади, м^2.

Таблица 6.1

Распределение магазинов некоторой фирмы по торговой площади

Магазины	Порядковый номер магазина																			
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Площадь магазина	60	80	40	100	60	70	50	120	100	60	50	120	100	60	80	60	70	40	50	50

Тогда,

$$\bar{x}_a = \frac{60+80+40+...+50}{20} = \frac{1420}{20} = 71m^2$$

т.е.,

$$\bar{x}_a = \frac{x_1 + x_2 + x_3 + ... + x_n}{n} = \frac{\sum_{i=1}^{n} x_i}{n} \qquad (6.1)$$

Средняя арифметическая простая применяется в тех случаях, когда каждое индивидуальное значение признака встречается один, одинаковое число раз, т.е. когда средняя рассчитывается по группированным единицам совокупности. Но чаще бывает так, что отдельные значения исследуемой совокупности встречаются не один, а много, причем неодинаковое число раз, т.е. представляют собой ряд распределения.

В этих случаях рассчитывают **среднюю арифметическую взвешенную.**

Рассмотрим предыдущий пример. Данные можно объединить в однородные группы, которые представлены в таблице 6.2.

Таблица 6.2

Распределение магазинов по торговой площади

Площадь магазинов	40	50	60	70	80	100	120
Число магазинов (f)	2	4	5	2	2	3	2

Тогда,

$$\bar{x}_{af} = \frac{40\times2+50\times4+60\times5+70\times2+80\times2+100\times3+120\times2}{2+4+5+2+2+3+2} = \frac{1420}{20} = 71м^2$$

Если индивидуальные значения признака (варианты) обозначить $\mathbf{x_1}$, $\mathbf{x_2}$, $\mathbf{x_3}$,..., $\mathbf{x_п}$, а числа, показывающие, сколько раз повторятся варианта (частоты, вес), – $\mathbf{f_1}$, $\mathbf{f_2}$, $\mathbf{f_3}$,..., $\mathbf{f_n}$, то средняя арифметическая взвешенная будет равна

$$\bar{x}_{af} = \frac{x_1 f_1 + x_2 f_2 + ... + x_n f_n}{f_1 + f_2 + ... + f_n} = \frac{\sum_{i=1}^{n} x_i f_i}{\sum_{i=1}^{n} f_i} \qquad (6.2)$$

Средняя арифметическая взвешенная равна сумме произведений вариант **(x)** на их частоты или веса **(f),** поделенной на сумму частот.

Средней арифметической интервального ряда. Как известно, вариационные ряды получаются в результате группировок, причем часто группировочные признаки показаны не одной величиной, а в определенных интервалах. Вычисление средней из интервального ряда имеет некоторые особенности. Для того чтобы рассчитать среднюю арифметическую интервального ряда, надо сначала определить среднюю для каждого интервала, а затем – среднюю для всего ряда.

Средняя для каждого интервала определяется как полусумма верхней и нижней границ, т.е. по **средней арифметической простой**.

Определение варианты как полусуммы верхней и нижней границ интервального ряда исходит из предположения, что индивидуальные значения признака внутри интервала распределяются равномерно и, следовательно, средние значения интервалов достаточно близко примыкают к средней арифметической в каждой группе. В действительности это не всегда так, поэтому средние, вычисленные из интервальных рядов, являются приближенными.

Для открытых интервалов в первой и последней группе значение признака определяется из значений последующего или предыдущего интервала. Их величина в первом случае отнимается от верхней границы и получают нижнюю, а во втором – прибавляют к нижней и получают верхнюю.

Пример 6.2.

Дано распределение рабочих предприятия по возрасту в табл. 6.3.

Таблица 6.3

Распределение рабочих предприятия по возрасту

Группы рабочих по возрасту, лет	Число рабочих, f_i	Середина интервала	$x_i f_i$
До 20	48	15	720
20–30	120	25	3000
30–40	75	35	2625
40–50	62	45	2790
Свыше 50	54	55	2970
Итого	359		12105

В первой группе нижняя граница будет: величина последующего интервала равна 10, 20 – 10 = 10. Получаем первую группу с границами 10 – 20, середина – 15. Для пятой группы: 50 + 10 = 60, группа с границами 50 – 60, середина – 55. Тогда

$$\bar{x}_{af} = \frac{\sum_{i=1}^{5} x_i f_i}{\sum_{i=1}^{5} f_i} = \frac{12105}{359} = 33.72$$

6.2.2. Свойства средней арифметической

Средняя арифметическая обладает рядом свойств, знание которых необходимо для понимания сущности средних, а также для упрощения их вычисления.

1. Средняя арифметическая суммы варьирующих величин равна сумме средних арифметических величин:

Если $\mathbf{x_i = y_i + z_i}$, то

$$\bar{x}_a = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{\sum_{i=1}^{n}(y_i + z_i)}{n} = \frac{\sum_{i=1}^{n} y_i}{n} + \frac{\sum_{i=1}^{n} z_i}{n} = \bar{y}_a + \bar{z}_a$$

Это правило показывает, в каких случаях можно суммировать средние величины. Если, например, выпускаемые изделия состоят из двух деталей y и z на изготовление каждой из них расходуется в среднем y = 3 ч, z = 5 ч, то средние затраты времени на изготовление одного изделия (x), будут равны: 3 + 5 = 8 ч, т.е. $x = y + z$.

2. Алгебраическая сумма отклонений индивидуальных значений варьирующего признака от средней равна нулю, так как сумма отклонений в одну сторону погашается суммой отклонений в другую сторону, т.е.

$$\sum_{i=1}^{n}\left(x_i - \bar{x}_{af}\right) = 0$$

Это правило показывает, что средняя является равнодействующей.

3. Если все варианты ряда уменьшить или увеличить на одно и то же число **a**, то средняя уменьшиться или увеличится на это же число **a**:

$$\bar{x}_{af} = \frac{\sum_{i=1}^{n} x_i f_i}{\sum_{i=1}^{n} f_i} = \frac{\sum_{i=1}^{n}(x_i \pm a) f_i}{\sum_{i=1}^{n} f_i} = \frac{\sum_{i=1}^{n} x_i f_i}{\sum_{i=1}^{n} f_i} \pm \frac{a\sum_{i=1}^{n} f_i}{\sum_{i=1}^{n} f_i} = \frac{\sum_{i=1}^{n} x_i f_i}{\sum_{i=1}^{n} f_i} \pm a = \bar{x}_{af} \pm a$$

4. Если все варианты ряда уменьшить или увеличить в **A** раз, то средняя также соответственно уменьшиться или увеличится в **A** раз:

$$\bar{x}_{af} = \frac{\sum_{i=1}^{n} x_i f_i}{\sum_{i=1}^{n} f_i} = \frac{\sum_{i=1}^{n}(A x_i) f_i}{\sum_{i=1}^{n} f_i} = \frac{A\sum_{i=1}^{n} x_i f_i}{\sum_{i=1}^{n} f_i} = A\bar{x}_{af}$$

5. Если все частоты (вес f) ряда разделить или умножить на одно и то же число d, то средняя не изменится:

$$\bar{x}_{af} = \frac{\sum\limits_{i=1}^{n} x_i f_i}{\sum\limits_{i=1}^{n} f_i} = \frac{\sum\limits_{i=1}^{n} x_i \frac{f_i}{d}}{\sum\limits_{i=1}^{n} \frac{f_i}{d}} = \frac{\sum\limits_{i=1}^{n} x_i f_i d}{\sum\limits_{i=1}^{n} f_i d} = \frac{d\sum\limits_{i=1}^{n} x_i f_i}{d\sum\limits_{i=1}^{n} f_i} = \frac{\sum\limits_{i=1}^{n} x_i f_i}{\sum\limits_{i=1}^{n} f_i} = \bar{x}_{af}$$

Это свойство показывает, что средняя зависит не от размеров весов, а от соотношения между ними. Следовательно, в качестве весов могут выступать не только абсолютные, но и относительные величины.

6.3. Другие формы средних величин

6.3.1. Средняя гармоническая

Для определения средней арифметической необходимо иметь ряд вариантов и частот, т.е. значения **x** и **f.** В некоторых случаях известны индивидуальные значения признака *x* и произведения **xf**, а частоты **f** неизвестны. Чтобы рассчитать среднюю, обозначим произведение **m = fx**, откуда **f = m/x.**

Теперь преобразуем формулу средней арифметической таким образом, чтобы по имеющимся данным **x**и**m** исчислить среднюю. Выразим в формуле средней арифметической **f** через **x** и **m** и получим:

$$\bar{x}_{grf} = \frac{\sum\limits_{i=1}^{n} x_i f_i}{\sum\limits_{i=1}^{n} f_i} = \frac{\sum\limits_{i=1}^{n} m_i}{\sum\limits_{i=1}^{n} \frac{m_i}{x_i}} \tag{6.3}$$

Средняя в такой форме называется **средней гармонической взвешенной** и обозначается $\mathbf{x_{gkf}}$. Следовательно, средняя гармоническая тождественна средней арифметической. Она применяется тогда, когда неизвестны действительные веса f, а известно произведение fx.

В тех случаях, когда произведения **fx** одинаковы или равны единице (m = 1), применяется **средняя гармоническая простая,** вычисляемая по формуле

$$\bar{x}_{gr} == \frac{\sum\limits_{i=1}^{n} m_i}{\sum\limits_{i=1}^{n} \frac{m_i}{x_i}} = \frac{1+1+...+1}{\frac{1}{x_1}+\frac{1}{x_2}+...+\frac{1}{x_n}} = \frac{n}{\sum\limits_{i=1}^{n} \frac{1}{x_i}} \tag{6.4}$$

где **x** – отдельные варианты; **n** – их число.

6.3.2. Средняя геометрическая

Если при замене индивидуальных величин признака на среднюю величину необходимо сохранить неизменным произведение индивидуальных величин, то следует применить **среднюю геометрическую:**

простую $$\bar{x}_{gm} = \sqrt[n]{x_1 \cdot x_2 \cdot \ldots \cdot x_n} = \sqrt[n]{\prod_{i=1}^{n} x_i} \qquad (6.5)$$

взвешенную $$\bar{x}_{gmf} = \sqrt[\sum_{i=1}^{n} f_i]{x_1^{f_1} \cdot x_2^{f_2} \cdot \ldots \cdot x_n^{f_n}} = \sqrt[\sum_{i=1}^{n} f_i]{\prod_{i=1}^{n} x_i^{f_i}} \qquad (6.6)$$

Основное применение геометрическая средняя находит при определении средних темпов роста. Пусть, например, в результате инфляции за первый год цена товара возросла в 2 раза к предыдущему году, а за второй год еще в 3 раза к уровню предыдущего года. Ясно, что за два года цена выросла в 6 раз. Каков средний темп роста цены за год? Арифметическая средняя здесь непригодна, ибо если за год цены возросли бы в (2 + 3)/2 = 2.5 раза, то за два года цена возросла бы в 2,5×2,5 = 6,25 раза, а не в 6 раз. Геометрическая средняя дает правильный ответ: $\sqrt{2 \cdot 3} = \sqrt{6} = 2.45$ раза. Геометрическая средняя величина дает наиболее правильный по содержанию результат осреднения, если задача состоит в нахождении такого значения признака, который качественно был бы равно удален как от максимального, так и от минимального значения признака. Например, если максимальный размер выигрыша в лотерее составляет миллион рублей, а минимальный – сто рублей, то какую величину выигрыша можно считать средней между миллионом и сотней? Арифметическая средняя явно непригодна, она составляет 500 050 руб., а это, как и миллион, крупный, никак не средний выигрыш; он качественно однороден с максимальным и резко отличен от минимального. Не дают верного ответа ни рассматриваемые далее квадратическая средняя (707 107 руб.), ни кубическая (793 699 руб.), ни гармоническая средняя (199,98 руб.), слишком близкая к минимальному значению. Только геометрическая средняя дает верный с точки зрения экономики и логики ответ: $\sqrt{100 \cdot 1000000} = 10000$ руб. Десять тысяч – не миллион, и не сотня! Это, действительно, нечто среднее между ними.

6.3.3. Средняя квадратическая

В тех случаях, когда осреднению подлежат величины, выраженные в виде квадратных функций, применяется средняя квадратическая. Так, средние диаметры колес, труб, стволов, средние стороны квадратов и др. определяются при помощи средней квадратической.

Средняя квадратическая простая рассчитывается путем извлечения квадратного корня из частного отделения суммы квадратов отдельных значений признака на их число:

$$\bar{x}_{kv} = \sqrt{\frac{x_1^2 + x_2^2 + \ldots + x_n^2}{n}} = \sqrt{\frac{\sum_{i=1}^{n} x_i^2}{n}} \tag{6.7}$$

Средняя квадратическая взвешенная равна

$$\bar{x}_{kvf} = \sqrt{\frac{x_1^2 f_1 + x_2^2 f_2 + \ldots + x_n^2 f_n}{f_1 + f_2 + \ldots + f_n}} = \sqrt{\frac{\sum_{i=1}^{n} x_i^2 f_i}{\sum_{i=1}^{n} f_i}} \tag{6.8}$$

где **f** – вес.

Соотношение между формами степенных средних величин. Рассмотренные выше средние величины могут быть представлены в форме некоторой системы величин. Различаются они лишь показателем. Можно записать в общем виде

$$\bar{x} = \sqrt[k]{\frac{\sum_{i=1}^{n} x_i^k}{n}} \tag{6.9}$$

В зависимости от целых значений **k** можно получить ту или иную среднюю:

k = –1 – средняя гармоническая;

k = 0 – средняя геометрическая (после преобразований);

k = 1 – средняя арифметическая;

k = 2 – средняя квадратическая.

При расчете различных степенных средних по одним и тем же данным статистического наблюдения средние не будут одинаковы. Чем выше степень **k** средней, тем больше ее величина. Математически доказано, что между величинами степенных средних, рассчитанных по одной и той же совокупности единиц статистического наблюдения и одному и тому же признаку, существует следующее соотношение:

$$\bar{x}_{gr} < \bar{x}_{gm} < \bar{x}_a < \bar{x}_{kv}$$

Например, имеются данные о выработке деталей пятью рабочими: 20; 16; 22; 28 и 14 деталей.

Соответственно,

средняя арифметическая: $\bar{x}_a = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{20+16+22+28+14}{5} = \frac{100}{5} = 20$

средняя гармоническая: $\bar{x}_{gr} = \frac{n}{\sum_{i=1}^{n} \frac{1}{x_i}} = \frac{5}{\frac{1}{20}+\frac{1}{16}+\frac{1}{22}+\frac{1}{28}+\frac{1}{14}} = 18.8$

средняя геометрическая: $\bar{x}_{gm} = \sqrt[n]{\prod_{i=1}^{n} x_i} = \sqrt[5]{20 \cdot 16 \cdot 22 \cdot 28 \cdot 14} = 19.4$

средняя квадратическая: $\bar{x}_{kv} = \sqrt{\frac{\sum_{i=1}^{n} x_i^2}{n}} = \sqrt{\frac{20^2+16^2+22^2+28^2+14^2}{5}} = 21.5$

Таким образом, получаем указанное выше соотношение между степенными средними:

$$18{,}8 < 19{,}4 < 20 < 21{,}2$$

6.4. Структурные средние. Мода и медиана

Для характеристики величины варьирующего признака пользуются так называемыми **структурными средними – модой и медианой**.

Мода – это наиболее часто встречающееся значение ряда. Мода применяется, например, при определении размера одежды, обуви, пользующихся наибольшим спросом у покупателей, наиболее распространенной цены на тот или иной товар на рынке, и т.д. (пример).

Величина моды, как правило, отличается от величины средней, совпадая с ней только в случае симметрии вариационного ряда.

Пример 6.3. Имеются данные обследования потребляемой женщинами обуви (табл. 6.4)

Таблица 6.4

Данные обследования потребляемой женщинами обуви

Размер обуви	32	33	34	35	36	37	38	39	40	41	42
Количество опрошенных женщин	6	33	247	910	2093	2696	1923	1196	283	51	55

Как видно из приведенного вариационного ряда, наиболее часто встречающейся величиной, т.е. **модой** этого ряда, является размер обуви 37, который носит 2696 женщин из опрошенных 9493 человек.

При расчете **моды** для **интервального** вариационного ряда необходимо вначале определить модальный интервал, в пределах которого находится мода, а затем значение модальной величины признака. В этом случае моду рассчитывают по следующей формуле:

$$M_0 = x_0 + h\frac{(f_m - f_{m-1})}{(f_m - f_{m-1}) + (f_m + f_{m+1})} \quad (6.10)$$

где $\mathbf{x_0}$ – нижняя граница модального интервала; **h** – величина модального интервала; $\mathbf{f_m}$ – частота (вес) модального интервала; $\mathbf{f_{m\_1}}$ – частота (вес) интервала, предшествующего модальному; $\mathbf{f_{m+1}}$ – частота (вес) интервала, следующего за модальным.

Пример 6.4.

Имеются группировочные данные по торговой площади магазинов (табл. 6.5).

Таблица 6.5

Группировочные данные по торговой площади магазинов

Торговая площадь магазинов, м^2	Число магазинов
До 100	3
100–120	13
120–140	15
140–160	20
160–180	8
Свыше 180	1
Итого	60

Произвести расчет моды из интервального ряда.

Модальным интервалом будет – (140–160). Тогда

$$M_O = 140 + 20\frac{(20-15)}{(20-15)+(20-8)} = 140 + 20\frac{5}{5+12} = 140 + 20 \cdot 0.29 = 145.8 \text{ м}^2$$

Но мода и средняя величина по-разному характеризуют совокупность. Мода определяет непосредственно размер признака, свойственный хотя и значительной части, но все же не всей совокупности. Мода по своему обобщающему значению менее точна по сравнению со средней арифметической, характеризующей совокупность в целом с учетом всех без исключения элементов совокупности.

Медианой является значение элемента, который больше или равен и одновременно меньше или равен половине остальных элементов ряда распределения. Медиана делит ряд на две равные части.

Пример 6.5.

А. Дан нечетный вариационный ряд роста студенток

156 158 160 $\boxed{161}$ 166 168 172

Из приведенного ряда видно, что медианой (центральным членом) данного ряда является рост студентки – 161 см.

В. Дан четный вариационный ряд роста студенток

155 156 158 $\boxed{160 \quad 161}$ 166 168 172

$$M_e = \frac{160+161}{2} = 160.5 \text{ см}$$

При нахождении медианы интервального вариационного ряда вначале определяют медианный интервал, в пределах которого находится медиана, а затем – приближенное значение медианы по формуле:

$$M_e = x_0 + h\frac{\frac{\Sigma f}{2} - S_{m-1}}{f_m} \qquad (6.11)$$

где $\mathbf{x_0}$ – нижняя граница интервала, который содержит медиану; **h** – величина медианного интервала; **Σf** – сумма частот всего ряда или число членов ряда, $\mathbf{S_{m-1}}$ – сумма накопленных частот интервалов, предшествующих медианному; f_m – частота медианного интервала.

Пример 6.6.

Имеются данные по группам семей по среднемесячному доходу на 1 человека (табл. 6.6).

Таблица 6.6

Статистические данные по группам семей по среднемесячному доходу на 1 человека

Группы семей по среднемес. доходу на 1 чел., руб	Число семей
До 900	10
900–1200	20
1200–1500	40
1500–1800	10
Свыше 1800	20
Итого	10

Медианный интервал – (1200–1500). Тогда

$$M_e = 1200 + 300\frac{\frac{100}{2} - 30}{40} = 1200 + 300\frac{20}{40} = 1350 \text{ руб.}$$

Медиана не зависит ни от амплитуды колебаний ряда, ни от распределения частот в пределах двух равных частей ряда, поэтому ее применение позволяет получить более точные результаты, чем при использовании других форм средних.

Контрольные вопросы

1. Роль и значение средних величин?
2. Виды степенных средних? Общая характеристика?
3. Сущность средней арифметической простой и взвешенной?
4. Сущность средней арифметической интервального ряда?
5. Каковы свойства средней арифметической?
6. Сущность средней гармонической? Пример применения.
7. Сущность средней геометрической? Пример применения.
8. Сущность средней геометрической? Пример применения.
9. Каковы соотношения между формами степенных средних величин?
10. Понятие структурных средних: мода, медиана?

ГЛАВА 7. ПОКАЗАТЕЛИ ВАРИАЦИИ

При изучении явлений и процессов общественной жизни статистика встречается с разнообразной вариацией (т.е. изменчивостью) признаков, характеризующих отдельные единицы совокупности. Величины признаков изменяются под действием различных факторов. Очевидно, что чем разнообразнее условия, влияющие на размер данного признака, тем больше его вариация. Так, размер заработной платы рабочих зависит от ряда факторов: специальности, разряда, стажа работы, образования, состояния здоровья и т.д. Чем больше различия между значениями факторов, тем больше вариация в уровне заработной платы рабочих.

Пример 7.1.

Имеется два различных ряда:

I ряд: 6, 10, 14, 26, 34 $\overline{x}_a = 18$

II ряд: 14, 16, 18, 20, 22 $\overline{x}_a = 18$

Среднее значение одинаково, но ряды различны.

При характеристике колеблемости признака применяют систему абсолютных и относительных показателей.

7.1. Абсолютные показатели вариации

К абсолютным показателям вариации относятся:

1) размах вариации R = $\mathbf{x_{max} - x_{min}}$**;**

2) среднее линейное отклонение d;

3) дисперсия $\boldsymbol{\sigma^2}$ **(D);**

4) среднеквадратическое отклонение $\sigma = \sqrt{D}$ **,**

Эти показатели (кроме дисперсии) измеряются в тех же единицах, что и сам признак: в тоннах, метрах, секундах, рублях и т.д.

7.1.1. Размах вариации

Наиболее простым способом измерения колеблемости является определение размаха вариации, т.е. разности между максимальным и минимальным значениями варьирующего признака.

$$R = x_{max} - x_{min} \qquad (7.1).$$

Величина **R** показывает, в каких пределах колеблется размер признака, образующего ряд распределения. Для примера 7.1: R(I ряда) = 28, R(II ряда) = 8. Таким образом, можно сделать вывод, что первый ряд распределения имеет значительно большую амплитуду вариант, чем второй ряд.

Показатель **R** выражается в тех же единицах измерения, что и варианты ряда. Но размах вариации как показатель колеблемости имеет существенный недостаток. Его величина определяется двумя крайними значениями признака, в то время как колеблемость последнего в целом складывается из суммы всех его значений. Поэтому размах вариации может в ряде случаев неправильно характеризовать колеблемость признака. Если, например, на большой посевной площади с равномерной в целом урожайностью встречаются отдельные небольшие участки с исключительно высокой и низкой урожайностью, то размах вариации будет иметь значительный размер, хотя колеблемость урожайности в целом незначительна. Следовательно, размах вариаций не отражает варьирования признака основной массы единиц совокупности.

В связи с тем, что каждое индивидуальное значение признака отклоняется от средней на определенную величину, очевидно, что мерой вариации может служить среднее из отклонений каждой отдельной варианты от их средней. Такими показателями являются среднее линейное отклонение, дисперсия и среднее квадратическое отклонение.

7.1.2. Среднее линейное отклонение

Среднее линейное отклонение представляет собой среднюю из абсолютных значений отклонений отдельных вариант от их средней. При вычислении среднего линейного отклонения принимаются во внимание только абсолютные значения отклонений без учета знаков «+» или «–». Среднее линейное отклонение рассчитывается по формуле средней арифметической простой

$$\bar{d} = \frac{\sum_{i=1}^{n} \left| x_i - \bar{x}_a \right|}{n} \qquad (7.1)$$

или средней арифметической взвешенной

$$\bar{d}_f = \frac{\sum_{i=1}^{n} \left| x_i - \bar{x}_{af} \right| f_i}{\sum_{i=1}^{n} f_i} \qquad (7.3)$$

Для примера7.1: $\quad$ d(I ряда) = $\dfrac{12+8+4+8+16}{5}$ = **9,6**;

d(II ряда) = $\dfrac{4+2+0+2+4}{5}$ = **2,4**.

Простота расчета и интерпретации составляют положительные стороны данного показателя. Однако, математические свойства модулей «плохие»: их нельзя поставить в соответствие с каким-либо вероятностным законом, в том числе и с нормальным распределением, параметром которого является не средний модуль отклонений, а среднее квадратическое отклонение.

7.1.3. Среднее квадратическое отклонение

Среднее квадратическое отклонение равно корню квадратному из суммы квадратов отклонений индивидуальных значений признака от их средней:

простое
$$\sigma = \sqrt{\frac{\sum_{i=1}^{n}\left(x_i - \bar{x}_a\right)^2}{n}} \qquad (7.4)$$

взвешенное
$$\sigma_f = \sqrt{\frac{\sum_{i=1}^{n}\left(x_i - \bar{x}_{af}\right)^2 f_i}{\sum_{i=1}^{n} f_i}} \qquad (7.5)$$

Для примера 7.1:

$$\sigma(\text{I ряда}) = \sqrt{\frac{12^2 + 8^2 + 4^2 + 8^2 + 16^2}{5}} = \sqrt{\frac{544}{5}} = \sqrt{108.8} \quad = 10,4$$

$$\sigma(\text{II ряда}) = \sqrt{\frac{4^2 + 2^2 + 0^2 + 2^2 + 4^2}{5}} = \sqrt{\frac{40}{5}} = \sqrt{8} = 2,8.$$

7.1.4. Дисперсия

Дисперсией называется средний квадрат отклонений индивидуальных значений признака от их средней величины. Дисперсия обозначается греческой буквой $\boldsymbol{\sigma^2}$ (или **D**) и равна

простая
$$\sigma^2 = \frac{\sum_{i=1}^{n}\left(x_i - \bar{x}_a\right)^2}{n} = D \qquad (7.6)$$

взвешенная
$$\sigma_f^2 = \frac{\sum_{i=1}^{n}\left(x_i - \bar{x}_{af}\right)^2 f_i}{\sum_{i=1}^{n} f_i} = D_f \qquad (7.7)$$

Для примера 7.1: D(I ряда) = **108.8**; D(II ряда) = **8**.

Пример 7.2.

Имеются данные по стажу продавцов в двух магазинах (таблица 7.1).

Таблица 7.1

Распределение продавцов в двух магазинах по стажу

№ п/п	1-й магазин			2-й магазин		
	Стаж продавцов, лет	Откл. от среднего	Кв. отклонения	Стаж продавцов, лет	Откл. от среднего	Кв. отклонения
1	1	–6,2	38,44	6	–1,2	1,44
2	2	–5,2	27,04	6	–1,2	1,44
3	3	–4,2	17,64	7	–0,2	0,04
4	3	–4,2	17,64	7	–0,2	0,04
5	4	–3,2	10,24	7	–0,2	0,04
6	9	1,8	3,24	7	–0,2	0,04
7	10	2,8	7,84	8	0,8	0,64
8	12	4,8	23,04	8	0,8	0,64
9	13	5,8	33,64	8	0,8	0,64
10	15	7,8	60,84	8	0,8	0,64
	72	0	239,60	72	0	5,6

Рассчитать: x_a, R, d, σ, D

$$\bar{x}_a(I) = \frac{72}{10} = 7.2; \bar{x}_a(II) = \frac{72}{10} = 7.2$$

R(I) = 15 – 1 = 14; R(II) = 8 – 6 = 2.

Значения отклонения от среднего и квадраты отклонения приведены в таблице 7.1. Тогда

$$d(I) = \frac{6.2+5.2+4.2+4.2+3.2+1.8+2.8+4.8+5.8+7.8}{10} = \frac{46}{10} = 4.6$$

$$d(II) = \frac{1.2+1.2+0.2+0.2+0.2+0.2+0.8+0.8+0.8+0.8}{10} = \frac{6.4}{10} = 0.64$$

$$D(I) = \frac{239.6}{10} = 23.96; \qquad D(II) = \frac{5.6}{10} = 0.56$$

$$\sigma(I) = \sqrt{D} = \sqrt{23.96} = 4.9 \qquad \sigma(II) = \sqrt{D} = \sqrt{0.56} = 0.75$$

7.1.5. Основные свойства дисперсии и упрощенные приемы ее вычисления

Дисперсия обладает рядом математических свойств, использование которых значительно упрощает и облегчает ее вычисление. Рассмотрим основные из них.

1. Дисперсия постоянной величины равна нулю: $D_{(a)} = 0$

2. Если все значения признака уменьшить или увеличить на какое-то постоянное число ***a***, то дисперсия от этого не изменится: $D_{(x+a)} = D_{(x)}$

3. Если все значения признака уменьшить или увеличить в **а** раз, то дисперсия от этого соответственно изменится в $\mathbf{a^2}$ раз. Следовательно, при исчислении дисперсии можно все значения признака уменьшить в **а** раз, исчислить дисперсию, а затем умножить ее на это постоянное число в квадрате ($\mathbf{a^2}$): $D_{(ax)} = a^2 D_x$

4. Сумма квадратов отклонений индивидуальных значений признака $\mathbf{x_i}$ от их средней $\bar{x}$ меньше суммы квадратов отклонений индивидуальных значений признака от любого данного числа **а** при условии, что $\mathbf{a \neq x_i}$, т.е.

$$\frac{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2}{n} < \left(x_i - a\right)^2 \qquad (7.8)$$

Это свойство дает возможность упрощать расчеты среднего квадратического отклонения путем замены громоздких отклонений индивидуальных значений признака от средней отклонениями от любого произвольно взятого числа, удобного для проведения расчетов, с последующей поправкой.

5. Дисперсия признака равна разности между средним квадратом значений признака и квадратом их средней, т.е.

$$D = \overline{x^2} - \bar{x}^2 \qquad (7.9)$$

Этот способ исчисления **D** называется **способом моментов** или **способом отсчета от условного нуля.**

Каждое свойство при расчете дисперсии может быть применено самостоятельно или в сочетании с другими.

Порядок расчета дисперсии простой:

1) определяют среднюю арифметическую $\bar{x} = \frac{\sum x}{n}$

2) возводят в квадрат среднюю арифметическую $\bar{x}^2 = \left(\frac{\sum x}{n}\right)^2$

3) возводят в квадрат каждую варианту ряда x_i^2

4) находим сумму квадратов вариант $\sum x_i^2$

5) делят сумму квадратов вариант на их число, т.е. определяют средний квадрат

$$\overline{x^2} = \frac{\sum x_i^2}{n}$$

6) определяют разность между средним квадратом признака и квадратом средней

$$\overline{x^2} - \bar{x}^2$$

Пример 7.3.

Имеются следующие данные о производительности труда рабочих (табл. 7.2):

Произведем следующие расчеты:

$$\bar{x} = \frac{50}{5} = 10 \text{ шт.}$$

$$D = \overline{x^2} - \bar{x}^2 = \frac{\sum x^2}{n} - \left(\frac{\sum x}{n}\right)^2 = \frac{510}{5} - \left(\frac{50}{5}\right)^2 = 102 - 100 = 2{,}0$$

Таблица 7.2

Статистические данные производительности труда рабочих

Табельный номер рабочего	Произведено продукции, шт	x_i^2
1	8	64
2	9	81
3	10	100
4	11	121
5	12	144
ИТОГО	50	510

Дисперсия альтернативного признака. В ряде случаев возникает необходимость измерить вариацию альтернативного признака. Обозначив отсутствие интересующего нас признака через 0; его наличие – через 1; долю единиц, обладающих данным признаком – через **p**; не обладающих – через **q**, исчислим среднее значение альтернативного признака и его дисперсию.

Среднее значение альтернативного признака равно

$$\bar{x} = \frac{(1 \cdot p) + (0 \cdot q)}{p + q} = p \qquad (7.10)$$

так как $p + q = 1$ (сумма долей единиц, обладающих и не обладающих данным признаком, равна единице).

Дисперсия альтернативного признака определяется следующим образом:

$$D = \frac{(1-p)^2 p + (0-p)^2 q}{p+q} \qquad (7.11)$$

Подставим в формулу дисперсии вместо 1 – p значение q = 1 – p и получим

$$D = pq \qquad (7.11)$$

Таким образом, $\boldsymbol{\sigma^2 = D = pq = p(1 - p)}$, т.е. дисперсия альтернативного признака равна произведению доли единиц, обладающих данным признаком, и доли единиц, им не обладающих.

7.1.6. Правило сложения дисперсий

Изучая дисперсию интересующего нас признака в пределах исследуемой совокупности и опираясь на общую среднюю в расчетах, нельзя оценить влияние отдельных факторов, определяющих колеблемость индивидуальных значений (вариант) признака. Это можно сделать при помощи метода группировок, когда единицы изучаемой совокупности подразделяются на однородные группы по признаку-фактору. При этом кроме общей средней для всей совокупности исчисляются средние по отдельным группам (групповые или частные средние) и три показателя дисперсии:

- общая дисперсия;
- межгрупповая дисперсия;
- средняя внутригрупповая дисперсия.

Величина общей дисперсии (σ_0^2) характеризует вариацию признака под влиянием всех факторов, формирующих уровень признака у единиц данной совокупности, и определяется по формуле (7.7), либо (7.6).

Межгрупповая дисперсия (дисперсия групповых средних $\boldsymbol{\sigma^2}$) отражает систематическую вариацию, т.е. те различия в величине изучаемого признака, которые возникают под влиянием фактора, положенного в основу группировки. Межгрупповая дисперсия определяется по формуле

$$\delta^2 = \frac{\sum_{i=1}^{n} \left(\bar{x}_i - \bar{x}_{af}\right)^2 n_i}{\sum_{i=1}^{n} n_i} \qquad (7.12)$$

где x_i – средняя по отдельной группе; n_i – число единиц в определенной группе.

Средняя внутригрупповая дисперсия ($\overline{\sigma^2}$) характеризует случайную вариацию, возникающую под влиянием других, неучтенных факторов, и не зависит от условия (признака-фактора), положенного в основу группировки.

Средняя внутригрупповая дисперсия определяется по формуле

$$\overline{\sigma^2} = \frac{\sum\limits_{i=1}^{n} \sigma_i^2 n_i}{\sum\limits_{i=1}^{n} n_i} \qquad (7.13)$$

где σ_i^2 – дисперсия по отдельной группе (согласно формуле (7.7)).

Указанные дисперсии взаимосвязаны между собой следующим равенством: величина общей дисперсии равна сумме межгрупповой дисперсии и средней внутригрупповой дисперсии:

$$\sigma_0^2 = \delta^2 + \overline{\sigma^2} \qquad (7.14)$$

Это тождество отражает закон (правило) **сложения дисперсий**. Опираясь на это правило, можно определить, какая часть (доля) общей дисперсии складывается под влиянием признака-фактора, положенного в основу группировки.

7.2. Относительные показатели вариации

К относительным показателям вариации относятся:

1) **коэффициент осцилляции**

$$V_R = \frac{R}{\overline{\overline{x}}_a} \cdot 100 \quad \text{или} \quad V_R = \frac{R}{\overline{\overline{x}}_{af}} \cdot 100 \qquad (7.15)$$

2) **линейный коэффициент вариации**

$$V_d = \frac{d}{\overline{\overline{x}}_a} \cdot 100 \quad \text{или} \quad V_d = \frac{d}{\overline{\overline{x}}_{af}} \cdot 100 \qquad (7.16)$$

3) **коэффициент вариации**

$$V_\sigma = \frac{\sigma}{\overline{\overline{x}}_a} \cdot 100 \quad \text{или} \quad V_\sigma = \frac{\sigma}{\overline{\overline{x}}_{af}} \cdot 100 \qquad (7.17)$$

Эти показатели выражаются в процентах.

Коэффициент вариации. Дисперсия и среднее квадратическое отклонение недостаточно полно характеризуют колеблемость признака, так как показывают абсолютный размер отклонений, что затрудняет сравнение изменчивости различных признаков.

Для характеристики колеблемости явлений среднее квадратическое отклонение сопоставляют с его средней величиной и выражают в процентах. Такой показатель называется коэффициентом вариации, обозначается буквой $\mathbf{V_\sigma}$ и рассчитывается по формуле (7.16)

Коэффициент вариации представляет собой отношение среднего квадратического отклонения к средней арифметической. Выражая его в процентах, различные абсолютные среднеквадратические отклонения приводят к одному основанию и дают возможность сравнивать, оценивать колеблемость величин различных признаков. При помощи коэффициента вариации возможно, например, сравнение размера колеблемости производительности труда групп рабочих, занятых производством различных видов продукции, размера колеблемости урожаев различных сельскохозяйственных культур и т.д. Чем меньше коэффициент вариации, тем меньше колеблемость признака, и наоборот.

Самым распространенным относительным показателем колеблемости является коэффициент вариации. Он более точно, чем абсолютный, характеризует различие колеблемости признаков.

По величине коэффициента вариации можно судить о степени вариации признаков совокупности. Чем больше его величина, тем больше разброс значений вокруг средней, тем менее однородна совокупность по своему составу и тем менее представительна средняя.

Коэффициент вариации важен и в тех случаях, когда нужно сравнивать средние квадратические отклонения, выраженные в разных единицах измерения.

Контрольные вопросы

1. Что представляют собой абсолютные показатели вариации?
2. Понятие размаха вариации и средне линейного отклонения?
3. Сущность среднеквадратического отклонения и дисперсии?
4. Основные свойства дисперсии?
5. Способы вычисления дисперсии?
6. Каковы правила сложения дисперсий?
7. Сущность относительных показателей вариации?
8. Понятие коэффициента вариации?

ГЛАВА 8. ВЫБОРОЧНЫЕ НАБЛЮДЕНИЯ

8.1. Понятие о выборочном методе

Статистика далеко не всегда имеет дело с данными **сплошного наблюдения**. Из всех видов несплошного наблюдения главным является **выборочное наблюдение**, так как только выборочный метод имеет статистико-математическое обоснование распространения данных, полученных по выборке, на всю совокупность.

Причин использования выборочного метода несколько.

Во-первых, как это ни парадоксально, это **повышение точности данных**; уменьшение числа единиц наблюдения в выборке резко снижает ошибки регистрации. Правда, за счет неполноты охвата единиц возникает ошибка репрезентативности, т.е. представительности выборочных данных. Но даже взятые вместе ошибка наблюдения для выборки плюс ошибка репрезентативности обеспечивают большую точность выборочных данных по сравнению с массовым сплошным наблюдением.

При ограничении объема работы можно привлечь более квалифицированных исполнителей (интервьюеров, счетчиков-регистраторов). Это положительно сказывается на качестве данных выборочного обследования.

Во-вторых, обращение к выборкам обеспечивает **экономию материальных, трудовых, финансовых ресурсов и времени.** Например, для составления баланса, денежных доходов и расходов населения, для изучения денежного обращения, выявления дифференциации населения по уровню жизни, определения черты бедности и т.д. необходимы данные о бюджетах домохозяйств. Сбор этих данных осуществляется государственной статистикой, но один статистик в состоянии курировать ежедневные записи доходов, расходов, потребления не более чем в 20–25 домохозяйствах. Если бы решили собирать данные о бюджетах всех домохозяйств, то только для этой цели (не учитывая потребности последующей обработки) потребовалось бы примерно два миллиона статистиков. Так что использование выборочного наблюдения является единственным экономически выгодным решением, тем более что по результатам изучения сравнительно небольшой части можно получить с достаточно высокой степенью уверенности данные о всей совокупности. Подобная ситуация возникает и при аудиторских проверках крупных фирм, когда вместо детального изучения каждого платежного документа ограничиваются анализом выборки документов, и в других областях применения статистики.

В-третьих, без выборки не обойтись, когда наблюдение связан с **порчей наблюдаемых объектов.** Это относится прежде всего к изучению качества продукции, которое основано на испытаниях образцов на вибрацию, упругость, разрыв и т.д. Всю продукцию, конечно же, таким испытаниям не подвергают, только отобранные образцы. То же можно сказать об исследовании молока на жирность, зерна на содержание белка, влажность, чистоту и всхожесть семян, электрических лампочек – на длительность горения и т.д. На выборках основаны маркетинговые исследования, оценки качества поставок.

Практика применения выборочного метода очень разнообразна. Иногда, проведя сплошное наблюдение, применяют выборочный метод при разработке данных: отбирают часть данных для более подробной разработки по расширенной программе. Так поступают, например, при разработке данных переписи населения о составе и типах семей. Нередко в процессе сбора данных применяют совместно сплошное и несплошное наблюдение. При переписях населения в нашей стране (1959, 1970, 1979 гг.) собирались сведения о каждом лице по 11 признакам, а 25% населения давали более подробную информацию (18 вопросов).

Выборки используются при опросах общественного мнения, при выяснении потребительских предпочтений, формировании доходов и расходов населения, при определении урожайности сельскохозяйственных культур и продуктивности скота. С 20-х гг. нашего века выборочный метод стал использоваться для контроля и анализа качества продукции. Сейчас методы статистической выборки все шире внедряются в самые различные области. В 1994 г. в Российской Федерации была проведена 5%-ная микроперепись населения с целью уточнения демографического и социального состава населения, уровня благосостояния, включая жилищные условия, источники дохода и др.

Та совокупность, из которой производится отбор, называется *генеральной совокупностью;* отобранные данные составляют *выборочную совокупность.* Эти данные представляют интерес постольку, поскольку дают основание для суждений о параметрах и свойствах генеральной совокупности.

Таким образом, **выборочный метод обладает следующими достоинствами:**

- относительно небольшие (по сравнению со сплошным наблюдением) материальные, трудовые и стоимостные затраты на сбор данных (включая затраты на планирование и формирование выборки);
- оперативность получения результатов;
- широкая область применения;
- высокая достоверность результатов.

Все эти достоинства проявляются лишь **при условии правильного решения проблем выборочного обследования.** К ним относятся:

1) определение границ генеральной совокупности;

2) разработка программы наблюдения и инструкций;

3) определение основы для проведения выборки – списка единиц генеральной совокупности, сведений об их размещении и т.д.;

4) установление допустимого размера погрешности и определение объема выборки;

5) определение вида выборочного наблюдения;

6) установление сроков проведения наблюдения;

7) определение потребности в кадрах для проведения выборочного наблюдения, их подготовка;

8) оценка точности и достоверности данных выборки, определение порядка их распространения на генеральную совокупность.

Представление о статистических данных как о выборочных может относиться не только к собственно выборке, но и к данным сплошного наблюдения, которые иногда рассматриваются как выборка из всех возможных реализаций изучаемого процесса. Это имеет смысл в случае малого числа единиц совокупности. Кроме того, трактовка данных как выборочных используется применительно к результатам эксперимента, которые рассматриваются как некая выборка из потенциально бесконечного числа повторений экспериментальных наблюдений.

Трактовка данных как выборочных является **основой деления** статистики на **описательную (дескриптивную) и выводную**.

Методы описательной статистики включают сбор данных по всем единицам изучаемой совокупности, их обработку, получение сводных показателей, которые являются характеристиками только наблюдаемой совокупности. Например, если наша задача состоит в изучении успеваемости группы студентов, включающей 25 человек, вычисленный средний балл по этой группе, процент отличных оценок и т.д. являются описаниями этой совокупности. Если же мы будем рассматривать эту группу студентов с точки зрения оценки успеваемости всех студентов данного колледжа или университета, то эта группа предстанет как выборка из общего числа студентов. В этом случае средний балл для группы будет являться оценкой средней успеваемости студентов колледжа в целом.

Генеральная совокупность может быть **реальной**, а может быть **гипотетической,** включающей случаи, которые реально не существуют, например все возможные результаты эксперимента.

В выводной статистике принято строго различать **параметры** и свойства генеральной совокупности и их **оценки** по данным выборки. С этой целью принята следующая система обозначений:

	Генеральная совокупность	Выборка
Средняя величина	$\bar{x}$	$\tilde{x}$
Относительная величина (доля)	p	ω
Дисперсия	σ^2(D)	σ^2 (S)
Сред. квадратическое	σ	σ

p – генеральная доля (доля единиц, обладающих данным значением признака в общем числе единиц генеральной совокупности)

Объем генеральной совокупности обозначают **N,** объем выборочной совокупности – **n**.

Выборочные оценки отличаются от генеральных параметров за счет ошибки наблюдения и ошибки выборки:

Выборочная		Генеральный		Ошибка		Ошибка
оценка	=	параметр	±	наблюдения	±	выборки

Подводя итоги, можно сказать, что описательная статистика является инструментом описания совокупности, по которой у нас полностью имеются исходные данные. **Метод статистического вывода** позволяет по данным выборок делать заключение о более большой совокупности, по которой не имеется исчерпывающих наблюдений.

Доля выборки есть отношение числа единиц выборочной совокупности к числу единиц генеральной совокупности:

$$K_B = \frac{n}{N} \tag{8.1}$$

Так, при 5%-ной выборке из партии деталей в 1000 ед. объем выборки *n* составляет 50 ед., а при 10%-ной выборке – 100 ед., и т.д. При правильной научной организации выборки ошибки **репрезентативности** можно свести к минимальным значениям, в результате – выборочное наблюдение становится достаточно точным.

8.2. Виды и схемы отбора

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

Различают **четыре вида отбора** совокупности единиц наблюдения:

1) случайный;
2) механический;
3) типический;
4) серийный (гнездовой).

Случайный отбор. Под этим видом отбора понимают наиболее распространенный способ отбора в случайной выборке, так называемый **метод жеребьевки**, при котором на каждую единицу совокупности заготовляется жетон или билет с порядковым номером. Затем в случайном порядке отбирается необходимое количество единиц совокупности. При этих условиях каждая из них имеет одинаковую вероятность попасть в выборку.

Примером случайного отбора могут служить тиражи выигрышей, когда из общего количества выпущенных билетов в случайном порядке наугад отбирается определенная часть номеров, на которые приходятся выигрыши. При этом всем номерам обеспечивается равная возможность попасть в выборку.

Механический отбор. Вся совокупность разбивается на равные по объему группы по случайному признаку. Затем из каждой группы, как правило, берется одна единица. Все единицы изучаемой совокупности предварительно располагаются в определенном порядке – например, по алфавиту, местоположению и т.п., а потом, в зависимости от объема выборки, механически, через определенный интервал, отбирается необходимое количество единиц. Так, если надо провести 10%-ную механическую выборку студентов, то составляется список их фамилий по алфавиту и механически отбирается каждый десятый студент, например: 1-й, 11-й, 21-й, 31-й или 7-й, 17-й, 27-й, 37-й и т.д. Если выборка 5%-ная, то отбирается каждый 20-й студент, т.е. интервал зависит от объема выборки. Чем меньше выборка, тем больше интервал.

Типический отбор. Изучаемая совокупность разбивается по существенному, типическому признаку на качественно однородные, однотипные группы. Затем из каждой группы случайным способом отбирается количество единиц, пропорциональное удельному весу группы во всей совокупности.

Например, необходимо провести типический отбор 1500 студентов из 10 тыс., обучающихся на 4 факультетах института. Для этого их группируют в однородные группы по факультетам, а затем по каждой из них отбирают число студентов пропорционально удельному весу числа студентов института по факультетам.

Типический отбор дает более точные результаты, чем случайный или механический, потому что при нем в выборку в такой же пропорции, как и в генеральной совокупности, попадают представители всех типических групп.

Серийный (гнездовой) отбор. Отбору подлежат не отдельные единицы совокупности, а целые группы (серии, гнезда), отобранные случайным или механическим способом. В каждой такой группе, серии проводится сплошное наблюдение, а результаты переносятся на всю совокупность.

Так, например, 10 тыс. студентов института занимаются группами по 25 человек. Для проведения 15%-ного выборочного наблюдения серийным (гнездовым) способом необходимо в случайном порядке отобрать 60 групп (1500:25 = 60) из 400 (10 000:25 = 400) и результаты наблюдения перенести на всю совокупность.

Точность выборки зависит и от схемы отбора. Выборка может быть проведена по **схеме повторного и бесповторного отбора**.

Повторный отбор. Каждая отобранная единица или серия возвращается во всю совокупность и может вновь попасть в выборку. Это так называемая схема возвращенного шара.

Бесповторный отбор. Каждая обследованная единица изымается и не возвращается в совокупность, поэтому она не попадает в повторное обследование. Эта схема получила название **невозвращенного шара**.

Бесповторный отбор дает более точные результаты по сравнению с повторным, так как при одном и том же объеме выборки наблюдение охватывает большее количество единиц изучаемой совокупности.

Комбинированный отбор. Рассмотренные виды отбора могут применяться и в комбинации. Комбинированный отбор может проходить одну или несколько ступеней. Выборка называется **одноступенчатой,** если отобранные однажды единицы совокупности подвергаются изучению. Тиражи денежно-вещевых лотерей, спортлото проводятся случайным отбором и представляют собой одноступенчатую выборку, как и проверка качества выпускаемой продукции путем отбора отдельных партий при серийном отборе.

Выборка называется **многоступенчатой,** если отбор совокупности проходит по ступеням, последовательным стадиям, причем каждая ступень, стадия отбора имеет свою единицу отбора. Так, отбор студентов для обследования успеваемости можно провести методом двухступенчатого отбора: вначале отобрать необходимое число академических групп, а затем в каждой выбранной группе – число студентов. Каждая ступень имеет свою единицу отбора: группа и студент.

Многофазная выборка характеризуется тем, что на всех ступенях выборки сохраняется одна и та же единица отбора, но проводится несколько стадий, фаз выборочных обследований, которые различаются между собой широтой программы обследования и объемом выборки. Важной особенностью многофазной выборки является возможность использовать данные первой фазы наблюдения для дополнительной характеристики и уточнения результатов, полученных на второй фазе, а эти данные, в свою очередь, – для третьей фазы и т.д. Например, на первой фазе по краткой программе (пять вопросов) обследуется 25% генеральной совокупности, на второй фазе по более широкой программе (включающей еще десять вопросов) – 15% генеральной совокупности, на третьей фазе по расширенной программе (включающей еще десять вопросов) – 5% и т.д.

8.3. Ошибки выборки

Ошибка выборки – это объективно возникающее расхождение между характеристиками выборки и генеральной совокупности. Она зависит от ряда факторов: степени вариации изучаемого признака, численности выборки, методом отбора единиц в выборочную совокупность, принятого уровня достоверности результата исследования.

Ошибки выборочного наблюдения, которые иначе называют **ошибками репрезентативности**, возникают вследствие специфики самого метода и именно потому, что обследуется не вся совокупность, а лишь его часть, отобранная в случайном порядке.

Определение средней величины этих ошибок и их возможных границ, а следовательно, определение достоверности данных выборочного наблюдения, является основной задачей теории выборочного исследования.

Рассмотрим некоторые вопросы теории выборочного метода и формулы ошибок для простой **случайной** выборки.

Применяя выборочный метод в статистике, обычно используют два основных вида обобщающих показателей: **среднюю величину количественного признака и относительную величину альтернативного признака** (долю или удельный вес единиц в статистической совокупности, которые отличаются от всех других единиц этой совокупности только наличием изучаемого признака).

Выборочная доля ω, или частость (вес), определяется отношением числа единиц, обладающих изучаемым признаком **m**, к общему числу единиц выборочной совокупности **n**:

$$\omega = \frac{m}{n} \tag{8.2}$$

Например, если из 100 деталей выборки (**n** = 100), 95 деталей оказались стандартными (**m** = 95), то **выборочная доля**

$$\omega = 95\,/100 = 0{,}95.$$

Для характеристики надежности выборочных показателей различают **среднюю и предельную** ошибки выборки.

Ошибка выборки ε или ошибка **репрезентативности** представляет собой разность соответствующих выборочных и генеральных характеристик:

для средней количественного признака

$$\varepsilon_x = \left|\bar{x} - \tilde{x}\right| \tag{8.3}$$

для доли (альтернативного признака)

$$\varepsilon_\omega = \left|\omega - p\right| \tag{8.4}$$

Ошибка выборки свойственна только выборочным наблюдениям. Чем больше значение этой ошибки, тем в большей степени выборочные показатели отличаются от соответствующих генеральных показателей.

Выборочная средняя и выборочная доля по своей сути являются **случайными величинами**, которые могут принимать различные значения в зависимости от того, какие единицы совокупности попали в выборку. Следовательно, ошибки выборки также являются случайными величинами и могут принимать различные значения. Поэтому определяют среднюю из возможных ошибок – **среднюю ошибку выборки μ**.

Средняя ошибка выборки зависит от степени варьирования изучаемого признака. Степень варьирования, как известно, характеризуется дисперсией $\boldsymbol{\sigma^2}$ или $\mathbf{D = \omega(1-\omega)}$ – для альтернативного признака (формула 7.11). Чем **меньше** вариация признака, а следовательно, и дисперсия, тем меньше **средняя ошибка выборки**, и наоборот. При нулевой дисперсии (признак не варьирует), средняя ошибка выборки равна нулю, т.е. любая единица генеральной совокупности будет совершенно точно характеризовать всю совокупность по этому признаку.

При **случайном повторном отборе средние ошибки теоретически** рассчитывают по следующим формулам:

для средней **количественного** признака

$$\mu_x = \sqrt{\frac{\sigma^2}{n}} \tag{8.5}$$

для **доли** (**альтернативного** признака)

$$\mu_p = \sqrt{\frac{p(1-p)}{n}} \tag{8.6}$$

Применяя формулу (8.4), получим следующую величину средней ошибки для примера: **n** = 100, **σ** = 45, тогда **μ** = ± 4,5

Поскольку практически дисперсия признака в генеральной совокупности $\boldsymbol{\sigma^2}$ точно неизвестна, на практике пользуются значением дисперсии **S**, рассчитанным для выборочной совокупности на основании закона больших чисел, согласно которому выборочная совокупность при достаточно большом объеме выборки достаточно точно воспроизводит характеристики генеральной совокупности.

Таким образом**,** расчетные формулы средней ошибки выборки при **случайном повторном отборе** будут следующие:

для средней количественного признака

$$\mu_x = \sqrt{\frac{S}{n}} \tag{8.7}$$

для доли (альтернативного признака)

$$\mu_{\omega} = \sqrt{\frac{\omega(1-\omega)}{n}} \qquad (8.8)$$

Однако **дисперсия выборочной совокупности** не равна **дисперсии генеральной совокупности**, и, следовательно, средние ошибки выборки, рассчитанные по формулам, будут приближенными. Но в теории вероятностей доказано, что генеральная дисперсия выражается через выборную следующим соотношением:

$$\sigma^2 = S\frac{n}{n-1} \qquad (8.9)$$

Так как **n/(n – 1)** при достаточно больших **n** – величина, близкая к единице, то можно принять, что $\sigma^2 \approx S$, а следовательно, в практических расчетах средних ошибок выборки можно использовать формулы. И только в случаях малой выборки (когда объем выборки не превышает 30) необходимо учитывать коэффициент **n/(n – 1)** и исчислять среднюю ошибку малой выборки по формуле:

$$\mu_{cp} = \sqrt{\frac{S}{n-1}} \qquad (8.10)$$

При **случайном бесповторном отборе** в приведенные выше формулы расчета средних ошибок выборки необходимо подкоренное выражение умножить на **1–(n/N),** поскольку в процессе **бесповторной выборки сокращается** численность единиц генеральной совокупности. Следовательно, для бесповторной выборки расчетные формулы средней ошибки выборки примут такой вид:

для средней количественного признака $\mu_x = \sqrt{\frac{S}{n}\left(1-\frac{n}{N}\right)}$ (8.11)

для доли (альтернативного признака) $\mu_{\omega} = \sqrt{\frac{\omega(1-\omega)}{n}\left(1-\frac{n}{N}\right)}$ (8.12)

Так как **n** всегда меньше **N**, то дополнительный множитель **1–(n/N)** всегда будет меньше единицы. Отсюда следует, что средняя ошибка при бесповторном отборе всегда будет меньше, чем при повторном.

На практике при применении выборочного метода обычно ставится задача определения пределов, за которые не выйдет величина конкретной ошибки выборочного наблюдения.

Величина пределов конкретной ошибки зависит от степени вероятности, с которой измеряется ошибка выборки.

Ошибка выборки, вычисленная с **заданной степенью вероятности**, представляет **предельную** ошибку выборки.

Если через **Δ** обозначим предельную ошибку, частное от деления **Δ** на **μ** приравняем к **t**, тогда можно записать $\frac{\Delta}{\mu} = t$, отсюда

$$\Delta = \mu t, \text{ а так как } \mu = \frac{\sigma}{\sqrt{n}}, \text{ то } \Delta = \frac{\sigma \cdot t}{\sqrt{n}} \quad (8.13)$$

Следовательно, величина предельной ошибки зависит от величины средней ошибки и коэффициента **t** (коэффициент доверия, связан с интегралом вероятностей, имеется специальная таблица). Коэффициент зависит от **степени вероятности**, с которой производится **выборочное наблюдение**.

Величину вероятности для различных значений **t** можно определить на основе теоремы Ляпунова. На практике пользуются готовыми таблицами значений этой функции, вычисленных для различных значений **t**. С увеличением значения **t** вероятность **P** быстро приближается к единице, так что практически обычно ограничиваются значениями **t**, не превышающими 2–3 единицы.

При значении t, равном 1, вероятность равна 0,683.

При значении t, равном 2, вероятность равна 0,954.

При значении t, равном 3, вероятность равна 0,997.

При значении t, равном 3,59, вероятность равна 0,9999999998.

Уже при значении t, равном 3, вероятность очень близка к единице. Это означает, что если бы из одной и той же генеральной совокупности было произведено большое число случайных выборок одинаковой численности, то в среднем на 1000 выборок приходилось бы 997 таких, в которых отклонение выборочной средней от генеральной не превышало бы 3 μ, и только в трех выборках отклонение могло бы выйти за эти пределы.

Указывая вероятные пределы случайной ошибки выборки, мы тем самым указываем и те пределы, за которые не выйдет характеристика генеральной совокупности, т.е. решаем ту задачу, которая, собственно, и ставится при выборочном наблюдении.

При проведении выборочного наблюдения часто возникает необходимость предварительного определения численности выборочной совокупности. Предположим, что мы хотим получить ошибку выборки вдвое меньшую, чем мы получили, т.е. ставим определенные условия: величина μ должна быть равна 2,15 вместо 4,30. Чтобы добиться уменьшения ошибки вдвое, нужно увеличить число наблюдений. Но на какое количество? Формула средней ошибки выборки позволяет ответить на этот вопрос:

$$\mu = \frac{\sigma}{\sqrt{n}}, \text{ или } \quad \mu^2 = \frac{\sigma^2}{n}, \text{ а отсюда } \; n = \frac{\sigma^2}{\mu^2}, \text{ или } \; \frac{\mu}{2} = \frac{\sigma}{2\sqrt{n}} = \frac{\sigma}{\sqrt{4n}}$$

при сокращении ошибки вдвое численность выборки должна быть увеличена в четыре раза, при сокращении втрое объем выборки должен быть увеличен в девять раз и т.д.

Для определения доли (удельного веса), изучаемого признака пользуются формулой средней ошибки выборки, которая имеет следующий вид:

$$\mu = \sqrt{\frac{p(1-p)}{n}} \tag{8.14}$$

где p – доля единиц, обладающих данным признаком в генеральной совокупности.

Но этот показатель неизвестен, и его как раз нужно определить на основе выборочного наблюдения. Поэтому величина доли **P** заменяется частостью ω:

$$\mu = \sqrt{\frac{\omega(1-\omega)}{n}} \tag{8.15}$$

Пример 8.1.

В таблице 8.1 приведены данные испытания крепости шерстяной пряжи. Определить:

- среднюю выборочную;
- среднее квадратическое отклонение выборочной совокупности;
- среднюю ошибку выборки;
- предельную ошибку с вероятностью 0,997;
- долю нитей, имеющих крепость 190 г. и больше.

Таблица 8.1

Крепость нити, в г.	Число проб	Средняя крепость в г.	$x_i f_i$	$(x_i - x_0)$	$\frac{x_i - x_0}{k}$	$\left(\frac{x_i - x_0}{k}\right)^2 \cdot f_i$
А	1	2	3	4	5	6
80 – 100	3	90	270	-100	-5	75
100 – 120	5	110	550	-80	-4	80
120 – 140	8	130	1040	-60	-3	72
140 – 160	10	150	1500	-40	-2	40
160 – 180	18	170	3060	-20	-1	18
180 – 200	26	190	4940	0	0	0
200 – 220	12	210	2520	20	1	12
220 – 240	8	230	1840	40	2	32
240 – 260	5	250	1250	60	3	45
260 – 280	3	270	810	80	4	48
280 - 300	2	290	580	100	5	50
Итого	100	-	18360	-		472

$x_0 = 190$

k – число, кратное величине интервала

Средняя выборочная: $\tilde{x} = \frac{\sum_{i=1}^{n} x_i f_i}{\sum_{i=1}^{n} f_i} = \frac{18360}{100} = 183.6$

Среднее квадратическое отклонение:

$$\sigma = \sqrt{\sigma_0^2 - (\bar{x} - x_0)^2} = \sqrt{\frac{\sum_{i=1}^{n}\left(\frac{x_i - x_0}{k}\right)^2 f_i k^2}{\sum_{i=1}^{n} f_i} - \left(\bar{x} - x_0\right)^2} =$$

$$= \sqrt{1888.0 - 40.96} = \sqrt{1847.04} = 43.0$$

Средняя ошибка выборки: $\mu = \frac{\sigma}{\sqrt{n}} = \frac{43}{10} = \pm 4.3$

Предельная ошибка (с вероятностью 0,997, t = 3): $\Delta = \mu \times t = \pm\, 4{,}3 \times 3 =$ $= \pm\, 12{,}9$

Доля нитей, имеющих прочность 190 г. и выше: Доля $\omega = \frac{56}{100} = 0.56$

Отсюда средняя ошибка для доли:

$$\mu = \sqrt{\frac{\omega(1-\omega)}{n}} = \sqrt{\frac{0.56(1-0.56)}{100}} = \sqrt{0.002464} = \pm 0.04956$$

При заданной степени вероятности (0,997) предельная ошибка доли равна:

$$\Delta_p = t\sqrt{\frac{\omega(1-\omega)}{n}} = \pm 0.04956 \cdot 3 = \pm 0.14868$$

Пределы генеральной доли определяем по формуле:

$$p = \omega \pm \Delta_p$$

Отсюда $p = 0{,}56 \pm 0{,}14868$

8.4. Малая выборка

При контроле качества товаров в экономических исследованиях эксперимент может проводиться на основе малой выборки.

Под малой выборкой понимается несплошное статистическое обследование, при котором выборочная совокупность образуется из сравнительно небольшого числа единиц генеральной совокупности. Объем малой выборки обычно не превышает 30 единиц и может доходить до 4–5 единиц.

Средняя ошибка малой выборки $\mu_{\text{М.В}}$ вычисляется по формуле:

$$\mu_{M.B} \approx \sqrt{\frac{S_{M.B}}{n}} \qquad (8.16)$$

где $S_{\text{М.В}}$ – дисперсия малой выборки.

При определении дисперсии s^2 число степеней свободы равно n–1:

$$S_{M.B} = \frac{\sum (x_i - \tilde{x})^2}{n-1} \qquad (8.17)$$

Предельная ошибка малой выборки $\Delta_{\text{М.В}}$ определяется по формуле

$$\Delta_{\text{М.В}} = t\,\mu_{\text{М.В}}$$

При этом значение коэффициента доверия t зависит не только от заданной доверительной вероятности, но и от численности единиц выборки n. Для отдельных значений t и n доверительная вероятность малой выборки определяется по специальным таблицам Стьюдента (табл. 8.2), в которых даны распределения стандартизированных отклонений:

$$t = \frac{\tilde{x} - \bar{x}}{S_{M.B}} \qquad (8.18)$$

Таблица 8.2

Значения доверительная вероятность

n	t				
	0,5	1,0	1,5	2,0	3,0
4	0,347	0,609	0,769	0,861	0,942
6	0,362	0,637	0,806	0,898	0,970
8	0,368	0,649	0,823	0,914	0,980
10	0,371	0,657	0,832	0,923	0,985
15	0,376	0,666	0,846	0,936	0,992
20	0,377	0,670	0,850	0,940	0,993

Поскольку при проведении малой выборки в качестве доверительной вероятности практически принимается значение 0,59 или 0,99, то для определения предельной ошибки малой выборки $\Delta_{\text{М.В}}$ используются следующие показания распределения Стьюдента (табл. 8.3, стр. 100)

Таблица 8.3

Значения предельной ошибки малой выборки

n	S_1	
	0,95	**0,99**
4	3,183	5,841
5	2,777	4,604
6	2,571	4,032
7	2,447	3,707
8	2,364	3,500
9	2,307	3,356
10	2,263	3,250
15	2,119	2,921
20	2,078	2,832

Пример 8.2.

При контрольной проверке качества поставленной в торговлю колбасы получены данные о содержании поваренной соли в пробах. По данным выборочного обследования нужно установить с вероятностью 0,95 предел, в котором находится средний процент содержания поваренной соли в данной партии товара.

Составляем расчетную таблицу (табл. 8.4) и по ее итогам определяем среднюю пробу малой выборки.

$$\tilde{x} = \frac{\sum x_i}{n} = \frac{41}{10} = 4{,}1\%$$

Определяем дисперсию малой выборки:

$$s^2_{M.B} = \frac{\sum (x_i - \tilde{x})^2}{n-1} = \frac{0{,}68}{10-1} = 0{,}075$$

Определяем среднюю ошибку малой выборки:

$$\mu_{M.B} \approx \sqrt{\frac{s_{M.B}}{n}} = \sqrt{\frac{0{,}075}{10}} = \pm 0{,}087$$

Исходя из численности выборки (n = 10) и заданной вероятности S_t = 0,95, устанавливается по распределению Стьюдента (см. табл. 8.3.) значение коэффициента доверия t = 2,263.

Предельная ошибка малой выборки составит:

$$\Delta_{\text{М.В}} = t\,\mu_{\text{М.В}} = 2{,}263(\pm 0{,}087) \approx \pm 0{,}02$$

Таблица 8.4

Расчетные значения

Пробы, % X_i	$x_i - \widetilde{x}$	$(x_i - \widetilde{x})^2$
4,3	0,2	0,04
4,2	0,1	0,01
3,8	0,3	0,09
4,3	0,2	0,04
3,7	–0,4	0,16
3,9	–0,2	0,04
4,5	0,4	0,16
4,4	0,3	0,09
4,0	–0,1	0,01
3,9	–0,2	0,04
Σ 41,0	–	0,68

Следовательно, с вероятностью 0,95 можно утверждать, что во всей партии колбасы содержание поваренной соли находится в пределах:

$$\bar{x} = \widetilde{x} \pm \Delta_{M.B} = 4{,}1\% \pm 0{,}2\%,$$

т.е. от 4,1% – 0,2% = 3,9% до 4,1% + 0,2% = 4,3%.

8.5. Определение необходимой численности выборки

Прежде чем приступить к проведению выборочного наблюдения, надо установить необходимую численность выборки, т.е. объем выборки, необходимый для того, чтобы обеспечить результаты выборочного наблюдения с заранее установленной точностью.

Необходимая численность выборки (**n**) определяется на основе формул предельной выборки. Так, если выборка повторная, то **n** при собственно-случайном и механическом отборах определяется из формулы

$$\Delta_x = t\sqrt{\frac{\sigma^2}{n}} \qquad (8.19)$$

где t – коэффициент доверия, вычисляемый по таблицам в зависимости от вероятности.

Чтобы найти n, возведем обе части уравнения в квадрат и получим:

$$\Delta_x^2 = \frac{t^2\sigma^2}{n}, \text{ откуда } \Delta_x^2 n = t^2\sigma^2, \text{ а } n = \frac{t^2\sigma^2}{\Delta_x^2} \qquad (8.20)$$

Объем выработки при исчислении доли определяется по этой же формуле, только вместо σ^2 берется $\omega(1 - \omega)$, т.е.

$$n = \frac{t^2\omega(1-\omega)}{\Delta_x^2\omega} \tag{8.21}$$

При бесповторном отборе численность выборки определяется из формулы

$$\Delta_x = t\sqrt{\frac{\sigma^2}{n}\left(1-\frac{n}{N}\right)} \tag{8.22}$$

Возведя обе части уравнения в квадрат, после преобразований получим:

$$n = \frac{t^2\sigma^2 N}{\Delta_x^2 N + t^2\sigma^2} \tag{8.23}$$

Аналогично исчисляется объем выборки и при определении доли, только вместо σ^2 берется $\omega(1 - \omega)$.

Пример 8.3.

В регионе имеется 2500 коров. Требуется определить необходимый объем **случайной** выборки для **повторного и бесповторного** отборов для того, чтобы с вероятностью 0,954 предельная ошибка выборки при определении среднего годового убоя не превышала 20 кг при среднем квадратическом отклонении 300 кг.

По условию задачи N = 2500, t = 2, Δ_x = 20 и σ = 300.

Для повторного отбора:

$$n = \frac{t^2\sigma^2 N}{\Delta^2 N + t^2\sigma^2} = \frac{2^2 300^2 2500}{20^2 2500 + 2^2 300^2} = \frac{4 * 90000 * 2500}{400 * 2500 + 4 * 90000} = 662$$

Таким образом, чтобы с вероятностью 0,954 получить предельную ошибку выборки не более 20 кг при среднем квадратическом отклонении 300 кг, необходимо отобрать из 2500 коров при повторном отборе 900, а при бесповторном – 662. Как видно, при прочих равных условиях объем выборки при бесповторном отборе меньше, чем при повторном.

Так же исчисляется объем выборки и при определении доли. Определим по данным нашего примера, сколько нужно отобрать породных коров для выборочного наблюдения, чтобы ошибка доли с вероятностью 0,954 (по приложению 1 t = 2) не превышала 3% (Δ_ω = 0,03) при удельном весе породных коров в выборке, равном

80% (ω = 0,8), N = 2500.

Для повторного отбора:

$$n = \frac{t^2\omega(1-\omega)}{\Delta_x^2\omega} = \frac{2^2 0{,}80{,}2}{0{,}03^2} = \frac{6400}{9} = 711$$

Для бесповторного отбора:

$$n = \frac{t^2\omega(1-\omega)N}{\Delta_x^2 N + t^2\omega(1-\omega)} = \frac{1600}{2{,}25 + 0{,}64} = 554$$

При определении необходимой численности выборки d^2 и p(ω) генеральной и выборочной совокупностей неизвестны, причем σ^2 и ω выборочной совокупности могут быть получены в результате проведения выборочного наблюдения. А без них нельзя установить необходимую численность выборки. В таких случаях фактическое значение дисперсии заменяют приближенным, полученным в результате проведения аналогичного выборочного наблюдения или пробного для ориентировочного суждения о ее размерах. Если признак альтернативный, то исходят из того, что ω = 0,5, а произведение ω(1 – ω) = 0,5 × 0,5 = 0,25. Вообще, при определении выборочных данных для вычисления необходимой численности выборки исходят из максимально возможных значений.

Рассмотрим другой пример выборки при исчислении выборочной доли для бесповторного отбора.

Пример 8.4.

Предполагается, что партия деталей содержит 8% бракованных. Необходимо определить нужный объем выборки, чтобы с вероятностью 0,954 можно было установить долю брака с погрешностью не более 2%. Исследуемая партия содержит 5 тыс. деталей.

Объем выборки при исчислении выборочной доли для бесповторного отбора определяется по формуле:

$$n = \frac{t^2\omega(1-\omega)N}{\Delta_x^2 N + t^2\omega(1-\omega)} \tag{8.24}$$

По условию задачи t = 2, доля бракованных деталей ω = 0,08, ω(1 – ω) = 1 – 0,08 = 0,92. Предельная ошибка доли по условию равна $\Delta_\omega = 0{,}02$, а N = 5000.

Подставляем эти данные в формулу и получаем:

$$\text{n} = \frac{2^2 * 0.08 * 0.92 * 5000}{0.0004 * 5000 + 4 * 0.0736} = 642$$

Чтобы с вероятностью 0.954 можно было утверждать, что предельная ошибка доли брака не превышает 2%, необходимо из 5000 деталей отобрать 642.

8.6. Способы распространения выборочных данных

Распространение выборочных данных на генеральную совокупность является конечной задачей выборочного наблюдения.

Обычно применяется два способа такого распространения: **способ прямого пересчета и способ коэффициентов.**

Способ прямого пересчета состоит в том, что средняя величина признака, найденная посредством выборки, умножается на число единиц генеральной совокупности.

Например, необходимо определить средний процент брака в партии консервов, состоящей из 10 000 банок. Для выборочного наблюдения в случайном порядке было отобрано 900 банок. Анализ качества отобранных банок консервов показал, что средний процент брака в данной совокупности составил 1,5%. Среднее квадратическое отклонение равно 0,3%. Максимальная ошибка выборочного наблюдения с вероятностью 0,997 равна 0,3%.

Таким образом, средний процент брака в генеральной совокупности находится в пределах 1,5±0,3%, т.е. колеблется от 1,2% до 1,8%.

Имея данные об общей величине партии, определяем общее количество бракованных банок, которое будет колебаться в пределах 1,8–1,2% от 10 000, или 180–120 единиц. Можно пределы не указывать, а пользоваться средней выборочной как генеральной средней. Тогда среднее количество бракованных банок в генеральной совокупности составит 1,5% от 10 000, т.е. 150 единиц.

Второй способ, или способ коэффициентов, применяется тогда, когда выборочное обследование проводится в целях проверки данных сплошного наблюдения.

Сущность этого метода заключается в том, что на основании сопоставления данных сплошного и данных выборочного наблюдений устанавливают **процент расхождений** (**процент недоучета**), который и служит **коэффициентом поправки**, налагаемой на данные сплошного наблюдения.

Пример 8.5.

Имеются данные о количестве скота, находящегося в личном пользовании согласно переписи, а также согласно контрольному обходу (табл. 8.5).

Таблица 8.5

Количество скота,
находящегося в индивидуальном пользовании населения

Группы скота	Учтено во всех хозяйствах по переписи	Учтено в хозяйствах, подвергнутых контрольному обходу		За время прошедшее от переписи до контрольного обхода в хозяйствах, подвергнутых контрольному обходу	
		по переписи	при контрольном обходе	прибыло	убыло
А	1	2	3	4	5
Коровы	9200	850	863	6	2
Нетели и телки, рожденные в прошлом году и старше	1200	140	144	4	1
Телки, рожденные в этом году	800	80	87	2	–
ИТОГО	10200	1070	1094	12	3

Чтобы определить процент недоучета, нужно найти разность между данными контрольного обхода и данными сплошного наблюдения, а затем полученную величину разделить на данные сплошного наблюдения.

При переписи недоучтено:

коров $863 - 850 - 6 + 2 = 9$;

нетелей $144 - 140 - 4 + 1 = 1$;

телок $87 - 80 - 2 = 5$

Отсюда коэффициент недоучета коров равен $\frac{9 \cdot 100}{850} = 1{,}06\%$

Коэффициент недоучета нетелей $\frac{1 \cdot 100}{140} = 0{,}72\%$

Коэффициент недоучета телок $\frac{5 \cdot 100}{80} = 6{,}25\%$

Полученные результаты выборочного наблюдения (проценты недоучета) распространяются на всю совокупность. Для этого поправочные коэффициенты (проценты недоучета) умножаем на данные сплошного наблюдения, полученные в результате переписи скота во всех хозяйствах (табл. 8.6, стр. 106).

Таблица 8.6.

Расчет фактического количества поголовья скота при помощи поправочных коэффициентов (процент недоучета)

Группы скота	Поправочные коэффициенты	Учтено во всех хозяйствах	Количество скота с поправкой
А	1	2	3
Коровы Нетели Телки	1,06 0,72 6,25	9200 1200 800	9297 1209 850
Итого	–	11200	11356

Контрольные вопросы

1. Понятие о выборочном методе наблюдения.
2. Достоинства и ошибки выборочного наблюдения. Общая характеристика.
3. Основные параметры генеральной и выборочной совокупности.
4. Виды отбора в выборочную совокупность.
5. Схемы отбора в выборочную совокупность.
6. Ошибки выборки.
7. Малая выборка.
8. Определение необходимой численности выборки.
9. Способы распространения выборочных данных.

ГЛАВА 9. ИЗМЕРЕНИЕ СВЯЗИ

9.1. Понятие связи в статистике

Происходящие явления и процессы органически связаны между собой, зависят друг от друга и обуславливают друг друга. Взаимосвязь и взаимообусловленность проявляются в работе любой фирмы, компании, предприятия и т.д. Так, замена одних станков на другие с коэффициентом полезного действия в два раза выше заменяемых станков приводит к снижению себестоимости единицы продукции, а следовательно, к увеличению прибыли, повышению материальной заинтересованности работников и т.д. Поэтому одной из важнейших задач статистики является изучение, измерение и количественное выражение взаимосвязей между явлениями жизни, установленными на основе качественного анализа.

Невозможно управлять явлениями, предсказывать их развитие без изучения характера, силы и других особенностей связей. Поэтому методы исследования, измерения связей составляют чрезвычайно важную часть методологии научного исследования, в том числе и статистического.

Различают два типа связей между различными явлениями и их признаками: **функциональную** или **жестко детерминированную,** с одной стороны, и **статистическую** или **стохастически детерминированную** – с другой. Строго определить различие этих типов связи можно тогда, когда они получают математическую формулировку. Для простоты будем говорить о связи двух явлений или двух признаков, математически отображаемой в форме уравнения связи двух переменных.

Если с изменением значения одной из переменных вторая изменяется строго определенным образом, т.е. значению одной переменной обязательно соответствует одно или несколько точно заданных значений другой переменной, связь между ними является **функциональной**.

Нередко говорят о строгом соответствии лишь одного значения второй из переменных каждому значению первой из них, но это неверно. Например, связь между **y** и **x** является строго функциональной, если $y = \sqrt{x}$ **;** но значению **x = 4** соответствует не одно, а два значения: $\mathbf{y_1 = +2;\ y_2 = -2}$**.** Уравнения более высоких степеней могут иметь несколько коней, связь, разумеется, остается функциональной.

Функциональная связь двух величин возможна лишь при условии, что вторая из них зависит *только* от первой и ни от чего более. Между тем все явления и процессы безграничного реального мира связаны между собой, и нет такого конечного числа переменных **k,** которые абсолютно полно определяли бы собою зависимую величину **y**. Некоторые науки (механика, электротехника, акустика, политическая экономия и другие) успешно используют представление связей как функциональных не только в аналитических целях, но нередко и в целях прогнозирования. Это возможно потому, что в простых системах интересующая нас переменная величина зависит в основном (скажем, на 99% или даже на 99,99%) от немногих других переменных или только от одной переменной. Например, длина года (период обращения Земли вокруг Солнца) почти функционально зависит только от массы Солнца и расстояния Земли от него. На самом деле она зависит в очень слабой степени и от масс, и расстояния других планет от Земли, но вносимые ими (и тем более в миллионы раз более далекими звездами) искажения функциональной связи для всех практических целей, кроме космонавтики, пренебрежимо малы.

Стохастически детерминированная связь не имеет ограничений и условий, присущих функциональной связи. Если с изменением значения одной из переменных вторая может в определенных пределах принимать любые значения с **некоторыми вероятностями**, но ее среднее значение или иные статистические (массовые) характеристики изменяются по определенному закону – **связь является статистической**. Иными словами, **при статистической связи** разным значениям одной переменной соответствуют разные распределения значений другой переменной.

В настоящее время наука не знает более широкого определения связи. Все связи, которые могут быть измерены и выражены численно, подходят под определение «статистические связи», в том числе и функциональные. Последние представляют собой частный случай статистических связей, когда значениям одной переменной соответствуют «распределения» значений второй, состоящие из одного или нескольких значений и имеющие вероятность, равную единице.

Корреляционной связью называют важнейший частный случай статистической связи, состоящий в том, что разным значениям одной переменной соответствуют различные **средние значения** другой. С изменением значения признака **x** закономерным образом изменяется среднее значение признака **y;** в то время как в каждом отдельном случае значение признака **y** (с различными вероятностями) может принимать множество различных значений.

Если же с изменением значения признака **x** среднее значение признака **y** не изменяется закономерным образом, но закономерно изменяется другая статистическая характеристика (показатели вариации, асимметрии, эксцесса и т.п.), то связь является не корреляционной, хотя и статистической.

Статистическая связь между двумя признаками (переменными величинами) предполагает, что каждый из них имеет случайную вариацию индивидуальных значений относительно средней величины. Если же такую вариацию имеет лишь один из признаков, а значения другого являются жестко детерминированными, то говорят лишь о регрессии, но не о статистической (тем более корреляционной) связи. Например, при анализе динамических рядов можно измерять регрессию уровней ряда урожайности (имеющих случайную колеблемость) на номера лет. Но нельзя говорить о корреляции между ними и применять показатели корреляции с соответствующей им интерпретацией.

Само слово **корреляция** ввел в употребление в статистику английский биолог и статистик **Френсис Гальтон** в конце XIX в. Тогда оно писалось как **«correlation»** (соответствие), но не просто «связь» **(relaton),** а «как бы связь», т.е. связь, но не в привычной в то время функциональной форме. В науке вообще, а именно в палеонтологии, термин «корреляция» применил еще раньше, в конце XVIII в., знаменитый французский палеонтолог (специалист по ископаемым останкам животных и растений прошлых эпох) **Жорж Кювье**. Он ввел даже «закон корреляции» частей и органов животных. «Закон корреляции» помогает восстановить по найденным в раскопках черепу, костям и т.д. облик всего животного и его место в системе: если череп с рогами, то это было травоядное животное, а его конечности имели копыта; если же лапа с когтями – то хищное животное без рогов, но с крупными клыками.

Корреляционная связь между признаками может возникать разными путями.

Причинная зависимость результативного признака (его вариации) от вариации факторного признака. Например, признак **x** – балл оценки плодородия почв, признак **y** – урожайность сельскохозяйственной культуры. Здесь совершенно ясно логически, какой признак выступает как независимая переменная (фактор) **x,** какой – как зависимая переменная (результат) **y.**

Совершенно иная интерпретация необходима при изучении корреляционной связи между двумя **следствиями общей причины**. Известен классический пример, приведенный крупнейшим статистиком России начала XX в. А.А. Чупровым: если в качестве признака **x** взять число пожарных команд в городе, а за признак **y** – сумму убытков за год в городе от пожаров, то между признаками **x** и **y** в совокупности городов России существенна прямая корреляция – в среднем, чем больше пожарников в городе, тем больше и убытков от пожаров! Уж не занимались ли пожарники поджигательством из боязни потерять работу? Но дело в другом. Данную корреляцию нельзя интерпретировать как связь причины и следствия; оба признака – **следствия общей причины** – размера города. Вполне логично, что в крупных городах больше пожарных частей, но больше и пожаров, и убытков от них за год, чем в мелких городах.

Третий путь возникновения корреляции – **взаимосвязь признаков**, каждый из которых и причина, и следствие. Такова, например, корреляция между уровнями производительности труда рабочих и уровнем оплаты 1 ч труда (тарифной ставкой). С одной стороны, уровень зарплаты – следствие производительности труда: чем она выше, тем выше и оплата. Но с другой стороны, установленные тарифные ставки и расценки играют стимулирующую роль: при правильной системе оплаты они выступают в качестве фактора, от которого зависит производительность труда. В такой системе признаков допустимы обе постановки задачи – каждый признак может выступать и в роли независимой переменной **x**, и в качестве зависимой переменной **y.**

Итак, при **функциональной связи** изменение результативного признака **y** всецело зависит от изменения факторного признака **x**: **y = f(x).**

При **корреляционной связи** изменение результативного признака **y** не всецело зависит от факторного признака **x**, а лишь частично, так как возможно влияние прочих факторов ε:

$$y = \varphi\,(x) + \varepsilon.$$

При корреляционной связи каждому определенному значению влияющего фактора соответствует ряд различных, не имеющих строго определенной величины, значений рассматриваемого признака. Эти значения колеблются вокруг средней из них. Корреляционная зависимость проявляется только в средних величинах и выражает числовое соотношение между ними в виде тенденции к возрастанию или убыванию одной переменной величины при возрастании или убывании другой. **Корреляционная связь** – понятие более узкое, чем статистическая связь.

Пример 9.1.

Среднее относительное число мальчиков среди детей в семьях любой величины близко к 0,5. Но в больших семьях, где 7–8 детей или больше, редко все дети – мальчики или все девочки. В семьях же с 2–3 детьми это бывает чаще. Таким образом, процент мальчиков среди детей находится в статистической зависимости от их общего числа.

Пример 9.2.

Себестоимость единицы продукции зависит от уровня производительности труда; чем выше производительность труда, тем ниже себестоимость. Но себестоимость зависит также и от ряда других факторов: стоимости сырья и материалов, топлива, электроэнергии, их расходов на единицу продукции, общепроизводственных и общехозяйственных расходов и т.д. Поэтому нельзя утверждать, что при повышении производительности труда, допустим на 10%, себестоимость снизится также на 10%. Может случиться, что, несмотря на рост производительности труда, себестоимость не только не снизится, но даже несколько повысится, если на нее окажут более сильное влияние действующие в обратном направлении другие факторы.

При корреляционной связи мы имеем дело не с приращением функции в зависимости от факторных признаков, а с сопряженной вариацией результативных и факторных признаков, выражающейся в их взаимосопряженных отклонениях от соответствующих средних значений. По этой причине корреляционная связь может быть установлена только в общем, в среднем при прочих равных условиях путем элиминирования (уничтожения) влияния многочисленных факторов.

Различают **прямую и обратную корреляционную связь**. Если с увеличением аргумента **x** функция **y** также увеличивается без всяких **единичных** исключений, то такая связь называется **полной прямой связью**. Если с увеличением аргумента **x** функция **y** уменьшается без всяких **единичных** исключений, то такая связь называется **полной обратной.**

При наличии исключений, которые, однако, не нарушают общей тенденции, имеет место **частичная связь** – прямая или обратная.

Когда признаки варьируют независимо друг от друга, это свидетельствует о полном отсутствии связи.

Пример 9.3.

Имеются два взаимосвязанных признака **x** и **y**, которые ведут себя в разных случаях следующим образом:

x	3	5	7	10	
y	15	17	20	22	– полная прямая связь
y	22	20	17	15	– полная обратная связь
y	15	20	17	22	– частичная прямая связь
y	22	17	20	15	– частичная обратная связь
y	20	15	22	17	– полное отсутствие связи

Линейная и нелинейная связь. По аналитическому выражению корреляционная связь может быть линейная и нелинейная. **Линейной называется связь, когда величина явления изменяется приблизительно равномерно в соответствии с изменением величины влияющего фактора.** Математически линейная связь может быть выражена уравнением прямой $y = a_0 + a_1x$.

Если происходит неравномерное изменение явления в связи с изменением величины влияющего фактора, то такая связь называется нелинейной. Математически нелинейная зависимость может быть выражена уравнением кривой второго порядка. В экономическом анализе часто пользуются уравнением параболы: $y = a_0 + a_1x + a_2x^2$. Кроме того, уравнение нелинейной связи может быть выражено в виде дробной функции: $y = a_0 / x + a_1$, показательной – $y = a_0a_1^x$ и т.д.

Точное аналитическое выражение имеют только функциональные связи. Корреляционные связи могут быть выражены лишь приближенно, при наличии определенных условий.

Характерной особенностью корреляционных связей является то, что они проявляются не в единичных случаях, а в массе, при больших значениях единиц совокупности.

9.2. Основные методы изучения взаимосвязей

Для изучения, измерения и количественного выражения взаимосвязей между явлениями в статистике применяются различные методы, важнейшими из которых являются: метод сопоставления, метод параллельных рядов, балансовый, графический, методы аналитических группировок, дисперсионного и корреляционного анализа.

Наиболее простой способ иллюстрации зависимости между двумя величинами – построение таблиц, показывающих, как при изменении одной величины меняется другая.

Пример 9.4.

Таблица 9.1

Данные о выработке молока на одного работающего

Производство молока в год, тыс. т	Выработка продукции на 1 работающего, тыс. руб.
До 31	34,2
31–50	37,3
51 и выше	42,7

Таблица 9.1 показывает лишь согласованность в изменении двух величин, наличие связи. Но она не определяет ни тесноту связи, ни форму этой связи.

Метод параллельных рядов. Чтобы установить связь между явлениями, достаточно расположить полученные в результате сводки и обработки материалы в виде параллельных рядов и сопоставить их между собой. Такое сопоставление, проведенное после теоретического анализа, показавшего возможность связи между изучаемыми явлениями, позволяет проследить числовые соотношения сопоставляемых признаков и направление их изменений, т.е. позволяет установить наличие связи и получить представление о ее характере (приведенная выше таблица).

Балансовый метод. Для характеристики взаимосвязи между явлениями в статистике применяется балансовый метод. Сущность его заключается в том, что данные взаимосвязанных показателей изображаются в виде таблицы и располагаются таким образом, чтобы итоги между отдельными частями были равны, т.е. чтобы был баланс. Балансовый метод используется для характеристики взаимосвязи между производством и реализацией продукции, денежными доходами и расходами населения и т.д. То есть:

$$O_H + P = B + O_K$$

O_H – остаток товаров на начало отчетного периода;

P – поступление товаров за период;

B – выбытие товаров в изучаемом периоде;

O_K – остаток товаров на конец отчетного периода.

Левая часть формулы характеризует предложение товаров – ($O_H + P$), а правая часть использование товарных ресурсов – ($B + O_K$).

Метод аналитических группировок. При наличии массовых статистических данных для изучения массовых явлений широко используются методом аналитических группировок. Аналитические группировки позволяют установить наличие связи между двумя и более признаками и ее направление. Метод группировок сочетается с методом средних и отдельных величин. Сущность метода аналитических группировок заключается в том, что единицы статистической совокупности группируются, как правило, по факторному признаку и для каждой группы исчисляется средняя или относительная величина по результативному признаку. Затем изменения средних или относительных значений результативного признака составляются с изменением факторного признака для выявления характера связи между ними.

Компонентный метод. Показателей производственной деятельности характеризуются тем, что изменение статистического показателя определяется изменением компонентов, входящих в этот показатель, как множители: $a = b \cdot c$

В статистике производственной деятельности компонентные связи используются в индексном методе. Например, индекс товарооборота в фактических ценах I_{qp} представляет произведение двух компонентов – индекса товарооборота в сопоставимых ценах I_q и индекса цен I_p, т.е. $I_{qp} = I_p \cdot I_q$.

Важное значение компонентной связи состоит в том, что она позволяет определять величину одного из неизвестных компонентов:

$$I_q = \frac{I_{pq}}{I_p} \qquad \text{или} \qquad I_p = \frac{I_{pq}}{I_q}$$

9.3. Дисперсионный анализ

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

Дисперсионной анализ дает, прежде всего, возможность определить роль систематической и случайной вариации в общей вариации и, следовательно, установить роль изучаемого фактора в изменении результативного признака. Для этого пользуются правилом сложения дисперсий, согласно которому общая дисперсия равна сумме двух дисперсий: средней из внутригрупповых и межгрупповой $\sigma^2 = \bar{\sigma}^2 + \delta^2$.

Для характеристики тесноты корреляционной связи между признаками в аналитических группировках межгрупповую дисперсию сопоставляют с общей. Это **отношение** называется **корреляционным** и обозначается:

$$\eta^2 = \frac{\delta^2}{\sigma^2}$$

Оно характеризует долю вариации результативного признака, вызванного воздействием факторного признака, положенного в основание группировки. Корреляционное отношение по своему абсолютному значению колеблется в пределах от 0 до 1. Чем ближе корреляционное отношение к 1, тем большее влияние оказывает факторный признак на результативный. Если же факторный признак не влияет на результативный, то вариация, обусловленная им, будет равна нулю ($\sigma^2 = 0$) и корреляционное отношение также равно нулю ($\eta^2 = 0$), что свидетельствует о полном отсутствии связи. И наоборот, если результативный признак изменяется только под воздействием одного факторного признака, то вариация, обусловленная этим признаком, будет равна общей вариации ($\delta^2 = \sigma^2$), и корреляционное отношение будет равно единице ($\eta^2 = 1$), что говорит о наличии полной связи.

Пример 9.5.

Определить при помощи корреляционного отношения тесноту связи между числом обслуживаемых станков и средней выработкой одной ткачихи (табл. 9.2).

Таблица 9.2

Дневная выработка ткачих (м)

№ п/п ткачихи	Дневная выработка		Отклон. $(x_i - \bar{x})$		Кв. отклон. $(x_i - \bar{x})^2$	
	32 станка	48 станков	32 станка	48 станков	32 станка	48 станков
1	40	62	–14	+8	196	64
2	48	66	–6	+12	36	144
3	43	60	–11	+6	121	36
4	45	68	–9	+14	81	196
5	44	64	–10	+10	100	100
Итого	220	320			534	540
Ср. выраб.	44	64				

Общая средняя выработка ткачих равна:

$$\bar{x} = \frac{220 + 320}{10} = 54 \text{м}$$

Вычислим общую дисперсию, характеризующую общую вариацию под влиянием всех факторов:

$$\sigma^2 = \frac{\sum_{i=1}^{n} \left(x_i - \bar{x}\right)^2}{n} = \frac{534 - 540}{10} = 107.4$$

Межгрупповая дисперсия, характеризующая факторную вариацию, т.е. различия в выборке, обусловленной неодинаковым числом обслуживаемых станков, определяются по формуле:

$$\delta^2 = \frac{\sum_{i=1}^{m} \left(\tilde{x}_i - \bar{x}\right)^2}{m} = \frac{(44 - 54)^2 + (64 - 54)^2}{2} = 100$$

Рассчитаем корреляционное соотношение:

$$\eta^2 = \frac{\delta^2}{\bar{\sigma}^2} = \frac{100}{107.4} = 0.931 \text{ или } 93{,}1\%$$

Следовательно, 93,1% всей вариации объясняется тем, что часть ткачих работала на 32 станках, а часть – на 48 и только 6,9% вариации является результатом действия прочих случайных факторов, не положенных в основание группировки.

Дисперсионный анализ позволяет не только определить роль случайной и систематической вариации, но и оценить **достоверность вариации**, обнаруженной методом аналитических группировок. Определение достоверности вариации дает возможность с заданной степенью вероятности установить, чем вызвана межгрупповая вариация – признаком, положенным в основание группировки, или является результатом действия случайных причин. Для оценки существенности корреляционного отношения пользуются критическими значениями корреляционного отношения η^2 при разных уровнях вероятности или значимости **а**.

Уровень значимости – это достаточно малое значение вероятности, отвечающее событиям, которые в данных условиях исследования будут считаться практически невозможными. Появление такого события является указанием на неправильность начального предположения. Чаще всего пользуются уровнями $a = 0{,}05$ или $a = 0{,}01$. Критические значения корреляционного отношения содержатся в специальных таблицах.

В этих таблицах распределение η^2 при случайных выборках зависит от числа степеней свободы факторной и случайной дисперсий. Число степеней свободы факторной дисперсии $K_1 = m - 1$, где m – число групп, а для случайной дисперсии $K_2 = n - m$, где n – число вариант, m – число групп. В нашем примере 10 ткачих сгруппированы в две группы по числу обслуживаемых станков. Поэтому $K_1 = m - 1 = 2 - 1 = 1$, а $K_2 = n - m = 10 - 2 = 8$. Из таблицы (**см. приложение 4**) находим критическое значение η^2, соответствующее $K_1 = 1$ и $K_2 = 8$ для уровней значимости $a = 0{,}05$, которое равно: $\eta^2_{(0.05)} = 0{,}399$. Это значит, что только в пяти случаях из 100 может случайно возникнуть корреляционное отношение, превышающее 0,399, а в 95 случаях из 100 корреляционное отношение, не может быть больше 0,399. Теперь фактическое значение корреляционного отношения надо сравнить с критическим, табличным. Если оно окажется больше критического, то связь между результативным и факторным признаками считается существенной, если же фактическое значение корреляционного η^2 меньше табличного, то связь между указанными признаками считается несущественной. В рассматриваемом нами примере фактическое значение корреляционного отношения $\eta^2 = 0{,}93$ больше табличного $\eta^2_{(0.05)} = 0{,}399$. Поэтому связь между числом обслуживаемых станков и выработкой является существенной.

При проверке существенной связи чаще пользуются **критерием Фишера**, потому что при больших числах степеней свободы его табличные значения мало изменяются, в отличие от корреляционного отношения, которое требует более громоздких таблиц. Критерий Фишера представляет собой отношение межгрупповой дисперсии к средней из среднегрупповых дисперсий, исчисленных с учетом числа степеней свободы:

$$F = \frac{\delta^2}{\overline{\sigma}^2} \cdot \frac{n-m}{m-1} \tag{9.1}$$

Для этих отношений Фишером (отсюда название «критерий Фишера») были составлены таблицы, по которым можно определить, какая величина F при данном числе степеней свободы по факторной вариации (K_1) и остаточной вариации (K_2) дает основание утверждать с определенной вероятностью (например 0,95 · 0,399), что положенный в основание группировки признак является несущественным.

В нашем примере $\overline{\sigma}^2 = 107{,}4$, $\delta^2 = 100$. По правилу сложения вариаций $\overline{\sigma}^2 = \sigma^2 - \delta^2 = 107{,}4 - 100{,}0 = 7{,}4$ и вычислим F:

$$F = \frac{\delta^2}{\overline{\sigma}^2} \cdot \frac{n-m}{m-1} = \frac{100.0}{7.4} \cdot \frac{10-2}{2-1} = 108.1$$

При уровне значимости $a = 0{,}05$, $K_1 = 1$ и $K_2 = 8$ критическое табличное (**см. приложение 5**) значение $F = 5{,}32$. Значит, уже при значении $F = 5{,}32$ можно с вероятностью 0,95 утверждать, что группировочный признак (число обслуживаемых станков) является весьма существенным.

Зная корреляционное отношение, можно определить критерий Фишера по следующей формуле:

$$F = \frac{\eta^2}{1-\eta^2} \cdot \frac{n-m}{m-1} \qquad (9.2)$$

В нашем примере: $F = \frac{\eta^2}{1-\eta^2} \cdot \frac{n-m}{m-1} = \frac{0.931}{1-0.931} \cdot \frac{10-2}{2-1} = 108.1$

9.4. Корреляционный анализ

9.4.1. Условия применения и ограничения корреляционного метода. Задачи корреляционного анализа

Поскольку корреляционная связь является статистической, **первым условием** возможности ее изучения является общее условие всякого статистического исследования: **наличие данных по достаточно большой совокупности явлений**. По отдельным явлениям можно получить совершенно превратное представление о связи признаков, ибо в каждом отдельном явлении значения признаков кроме закономерной составляющей имеют случайное отклонение (вариацию). Например, сравнивая два хозяйства, одно из которых имеет лучшее качество почв, по уровню урожайности, можно обнаружить, что урожайность выше в хозяйстве с худшими почвами. Ведь урожайность зависит от сотен факторов и при том же самом качестве почв может быть и выше, и ниже. Но если сравнивать большое число хозяйств с лучшими почвами и большое число – с худшими, то средняя урожайность в первой группе окажется выше и станет возможным измерить достаточно точно параметры корреляционной связи.

Какое именно число явлений достаточно для анализа корреляционной и вообще статистической связи, зависит от цели анализа, требуемой точности и надежности параметров связи, от числа факторов, корреляция с которыми изучается. Обычно считают, что число наблюдений должно быть не менее чем в 5–6, а лучше – не менее чем в 10 раз больше числа факторов. Еще лучше; если число наблюдений в несколько десятков или в сотни раз больше числа факторов, тогда закон больших чисел, действуя в полную силу, обеспечивает эффективное взаимопогашение случайных отклонений от закономерного характера связи признаков.

Вторым условием закономерного проявления корреляционной связи служит условие, обеспечивающее надежное выражение закономерности в средней величине. Кроме уже указанного большого числа единиц совокупности для этого **необходима достаточная качественная**

однородность совокупности. Нарушение этого условия может извратить параметры корреляции. Например, в массе зерновых хозяйств уровень продукции с гектара растет по мере концентрации площадей, т.е. он выше и крупных хозяйствах. В массе овощных и овоще-молочных хозяйств (пригородный тип) наблюдается та же прямая связь уровня продукции с размером хозяйства. Но если соединить в общую неоднородную совокупность те и другие хозяйства, то связь уровня продукции с размером площади пашни (или посевной площади) получится обратной. Причина в том, что овощные и овоще-молочные хозяйства, имея меньшую площадь, чем зерновые, производят больше продукции с гектара ввиду большей интенсивности производства в данных отраслях, чем в производстве зерна.

Иногда как **условие** корреляционного анализа выдвигают **необходимость подчинения распределения совокупности по результативному и факторным признакам нормальному закону распределения вероятностей**. Это условие связано с применением метода наименьших квадратов при расчете параметров корреляции: только при нормальном распределении метод наименьших квадратов дает оценку параметров, отвечающую принципам максимального правдоподобия. На практике эта предпосылка чаще всего выполняется приближенно, но и тогда метод наименьших квадратов дает неплохие результаты.

Однако при значительном отклонении распределений признаков от нормального закона нельзя оценивать надежность выборочного коэффициента корреляции, используя параметры нормального распределения вероятностей или распределения Стьюдента.

Еще одним **спорным вопросом** является допустимость **применения корреляционного анализа к функционально связанным признакам**. Можно ли, например, построить уравнение корреляционной зависимости размеров выручки от продажи картофеля, от объема продажи и цены? Ведь произведение объема продажи и цены равно выручке в каждом отдельном случае. Как правило, к таким жестко детерминированным связям применяют только индексный метод анализа. Однако на этот вопрос можно взглянуть и с другой точки зрения. При индексном анализе выручки предполагается, что количество проданного картофеля и его цена независимы друг от друга, потому-то и допустима абстракция от изменения одного фактора при измерении влияния другого, как это принято в индексном методе. В реальности количество и цена не являются вполне независимыми друг от друга.

Корреляционный анализ учитывает межфакторные связи, следовательно, дает нам более полное измерение роли каждого фактора: прямое, непосредственное его влияние на результативный признак; косвенное влияние фактора через его влияние на другие факторы; влияние всех факторов на результативный признак. Если связь между факторами несущественна, индексным анализом можно ограничиться. В противном случае его полезно дополнить корреляционно-регрессионным измерением влияния факторов, даже если они функционально связаны с результативным признаком.

Измерение связи между событиями не ограничивается установлением связи между ними или определением роли систематической вариации в общей вариации. Изучение взаимосвязей между признаками статистической совокупности заключается в определении формы и количественной характеристики связи, а также степени тесноты (сопряженности) связи. Корреляционный анализ и решает эти **две основные задачи**.

Первая задача заключается в определении формы связи, установлении математической формы, в которой выражается данная связь, т.е. измерение параметров уравнения, выражающего связь средних значений зависимой переменной со значениями независимой переменной (зависимость средних величин результативного признака от значений одного или нескольких факторных признаков). Это очень важно, так как от правильного выбора формы связи зависит конечный результат изучения взаимосвязи между признаками. Определение формы связи не может быть произведено только при помощи математических методов. Корректно и наиболее полно определить ее возможно только на основе предварительного качественного анализа изучаемых явлений.

Вторая задача состоит в измерении тесноты, т.е. меры связи между признаками с целью установить степень влияния данного фактора на результат. Она решается математическими методами путем определения параметров корреляционного уравнения.

Вторая задача специфична для статистических связей, а первая разработана для функциональных связей и является общей. Основным методом решения задачи нахождения параметров уравнения связи является метод наименьших квадратов (МНК), разработанный К.Ф. Гауссом (1777–1855). Он состоит в минимизации суммы квадратов отклонений фактически измеренных значений зависимой переменной **y** от ее значений, вычисленных по уравнению связи с факторным признаком (многими признаками) **x**.

В заключение проводится оценка и анализ полученных результатов при помощи специальных показателей корреляционного метода (коэффициентов детерминации, линейной и множественной корреляции и т.д.), а также проверка существенности связи между изучаемыми признаками.

Выбор формы связи. Определяющая роль в выборе формы связи между явлениями принадлежит теоретическому анализу. Так, например, выпуск продукции и стоимость основных фондов, урожайность и количество внесенных удобрений взаимосвязаны между собой. Анализ показывает, что чем больше имеет фирма, компания, предприятие основных фондов (факторный признак), тем больше, при прочих равных условиях, оно выпускает продукции (результативный признак). С ростом факторного признака здесь, как правило, равномерно растет и результативный, поэтому зависимость между ними может быть выражена уравнением прямой $\mathbf{y = a_0 + a_1x}$, которое называется **линейным уравнением регрессии**.

Параметр $\mathbf{a_1}$ называется **коэффициентом регрессии** и показывает, насколько в среднем отклоняется величина результативного признака **y** при отклонении величины факторного признака **x** на одну единицу. При $\mathbf{x = 0,\ a_0 = y}$. Увеличение количества внесенных удобрений приводит, при прочих равных условиях, к росту урожайности, но чрезмерное внесение их без изменения других элементов к дальнейшему повышению урожайности не приводит, а, наоборот, снижает ее. Такая зависимость может быть выражена уравнением кривой второго порядка – параболы: $\mathbf{y = a_0 + a_1x + a_2x^2}$. Параметр $\mathbf{a_2}$ характеризует степень ускорения или замедления кривизны параболы и при $\mathbf{a_2 > 0}$ парабола имеет минимум, а при $\mathbf{a_2 < 0}$ – максимум. Параметр $\mathbf{a_1}$ характеризует крутизну кривой, а параметр $\mathbf{a_0}$ – вершину кривой.

На практике изучение взаимосвязи между признаками часто базируется на значительном числе наблюдений, материалы которых группируются по двум взаимосвязанным признакам (**x** и **y**). Результаты группировки оформляются в виде корреляционной таблицы, или решетки.

Корреляционная решетка представляет собой комбинационную таблицу, в подлежащем которой располагаются значения одного признака, как правило, факторного, а в сказуемом – другого, результативного. В клетках, образовавшихся при пересечении строк и граф, указываются частоты, т.е. число случаев, в которых одни значения сочетаются с другими.

Пример 9.6.

Имеются данные о количестве внесенных удобрений (в пересчете на действующие вещества) и урожайности зерновых культур по 240 фермерским хозяйствам области. Чтобы изучить связь между количеством внесенных удобрений (**x**) и урожайностью зерновых (**y**), составим корреляционную таблицу (табл. 9.3).

Цифры, стоящие на пересечении строк и граф, показывают связь количества фермерских хозяйств с данным количеством внесенных удобрений и урожайностью.

Таблица 9.3

Данные о связи количества фермерских хозяйств, внесенных удобрений и урожайностью

Внесено удобрений (х), кг/га	Урожайность (у), ц/га							
	16	18	21	25	26	30	всего	сред. ур. (у)
До 50	5	6	5	–	–	–	16	18,3
50–75	7	–	18	12	–	–	37	21,4
75–100	6	12	36	18	10	–	82	21,7
100–125	–	–	19	30	14	8	71	24,7
125–150	–	–	–	6	12	10	28	27,2
150 и выше	–	–	–	–	–	6	6	30,0
Всего	18	18	78	66	36	24	240	22,7

По корреляционной таблице можно сделать некоторые выводы о форме и направлении связи, о степени тесноты связи. Если значения **x** и **y** расположены в возрастающем порядке, то сосредоточение частот около диагонали таблицы, идущей с левого верхнего угла в правый нижний, свидетельствует о прямой связи между изучаемыми признаками, а с правого верхнего угла в левый нижний – об обратной связи, причем связь будет тем теснее, чем плотнее концентрируются частоты у диагонали. Если частоты расположены по всей таблице равномерно, то это говорит о слабой связи между признаками или об отсутствии ее.

Данные корреляционной таблицы можно также изобразить графически (рис. 9.1). Для этого результаты группировки единиц совокупности по факторному признаку и средние значения результативного признака по каждой группе наносятся на график в виде точек, которые затем соединяются, в результате чего получают ломаную линию, как это показано на рис. 9.1 (средние значения результативного признака даны в последней графе табл. 9.2).

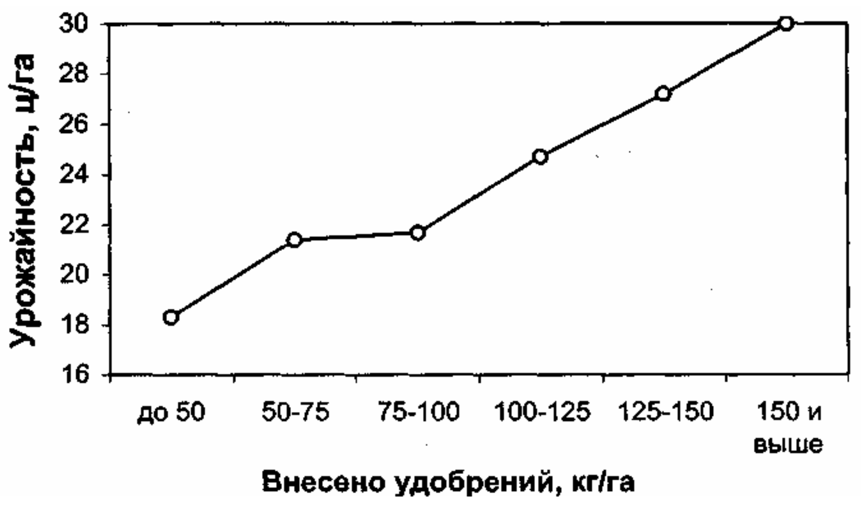


Рис. 9.1.

Полученная ломаная линия называется **эмпирической линией регрессии.**

Необходимо сказать и о других задачах применения корреляционного метода, имеющих не формально математический, а содержательный характер.

1. Задача выделения важнейших факторов, влияющих на результативный признак (т.е. на вариацию его значений в совокупности). Эта задача решается в основном на базе мер тесноты связи факторов с результативным признаком.

2. Задача оценки хозяйственной деятельности по эффективности использования имеющихся факторов производства. Эта задача решается путем расчета для каждой единицы совокупности тех величин результативного признака, которые были бы получены при средней по совокупности эффективности использования факторов и сравнения их с фактическими результатами производства,

3. Задача прогнозирования возможных значений результативного признака при задаваемых значениях факторных признаков. Такая задача решается путем подстановки ожидаемых, или планируемых, или возможных значений факторных признаков в уравнение связи и вычисления ожидаемых значений результативного признака.

Приходится решать и обратную задачу: вычисление необходимых значений факторных признаков для обеспечения планового или желаемого значения результативного признака в среднем по совокупности. Эта задача обычно не имеет единственного решения в рамках данного метода и должна дополняться постановкой и решением оптимизационной задачи на нахождение наилучшего из возможных вариантов ее решения (например, варианта, позволяющего достичь требуемого результата с минимальными затратами).

4. Задача подготовки данных, необходимых в качестве исходных для решения оптимизационных задач. Например, для нахождения оптимальной структуры производства в районе на перспективу исходная информация должна включать показатели производительности на предприятиях разных отраслей и форм собственности. В свою очередь, эти показатели могут быть получены на основе корреляционно-регрессионной модели либо на основании тренда динамического ряда (а тренд – это тоже уравнение регрессии).

При решении каждой из названных задач нужно учитывать особенности и ограничения корреляционного метода. Всякий раз необходимо специально обосновать возможность причинной интерпретации уравнения как **объясняющего** связь между вариацией фактора и результата. Трудно обеспечить раздельную оценку влияния каждого из факторов. В этом отношении корреляционные методы глубоко противоречивы.

С одной стороны, их идеал – измерение чистого влияния каждого фактора. С другой стороны, такое измерение возможно при отсутствии связи между факторами и случайной вариации признаков. А тогда связь является функциональной, и корреляционные методы анализа излишни. В реальных системах связь всегда имеет статистический характер, и тогда идеал методов корреляции становится недостижимым. Но это не значит, что эти методы не нужны.

Данное противоречие означает попросту недостижимость абсолютной истины в познании реальных связей. Приближенный характер любых результатов корреляционно-регрессионного анализа не является поводом для отрицания их полезности. Всякая научная истина – относительна. Забыть об этом и абсолютизировать параметры регрессионных уравнений, меры корреляции было бы ошибкой, так же как и отказаться от использования этих мер.

9.4.2. Аналитическое выражения связи

Применение методов корреляционного анализа дает возможность выражать связь между признаками аналитически – в виде уравнения – и придавать ей количественное выражение. Рассмотрим применение приемов корреляционного анализа на конкретном примере (табл. 9.4).

Пример 9.7.

Таблица 9.4

Стоимость основных фондов и выпуск продукции по группе предприятий

№ п/п	Стоимость ОФ (x), млн. руб.	Выпуск продукции (y), млн. руб.	xy	x^2	y_T
1	6	2,4	14,4	36	2,692
2	8	4,0	32,0	64	3,537
3	9	3,6	32,4	81	3,958
4	10	4,0	40,0	100	4,380
5	10	4,5	45,0	100	4,380
6	11	4,6	50,6	121	4,802
7	12	5,6	67,2	144	5,224
8	13	6,5	84,5	169	5,646
9	14	7,0	98,0	196	6,068
10	15	5,0	75,0	225	6,490
Итого	108	47,2	539,1	1236	47,177

Анализ данных табл. 9.3 показывает, что с увеличением стоимости основных фондов (ОФ) растет, как правило, и выпуск продукции. Однако мы не можем утверждать, что увеличение стоимости основных фондов, например на 1 млн. руб., приводит к фактическому увеличению выпуска продукции на определенную сумму. Здесь можно говорить только о каких-то средних тенденциях. Чтобы установить, насколько повышается в среднем выпуск продукции при увеличении основных фондов на 1 млн. руб., прежде всего, определим форму связи. Допустим, что между стоимостью основных фондов и выпуском продукции существует линейная связь, которая выражается уравнением прямой $\mathbf{y = a_0 + a_1x}$. Необходимо найти параметры $\mathbf{a_0}$ и $\mathbf{a_1}$, что позволит определить теоретические значения **y** для разных значений $\mathbf{x_i}$. Причем $\mathbf{a_0}$, $\mathbf{a_1}$ должны быть такими, чтобы было достигнуто максимальное приближение к первоначальным (эмпирическим) значениям $\mathbf{y_i}$ теоретических значений $\mathbf{y_T}$. Эта задача решается при помощи способа наименьших квадратов, основное условие которого сводится к определению параметров $\mathbf{a_0}$ и $\mathbf{a_1}$ таким образом, чтобы

$$\sum_{i=1}^{n}(y_i - y_T)^2 = \min$$

Математически доказано, что условие минимума обеспечивается, если параметры a_0 и a_1 определяются при помощи системы двух нормальных уравнений, отвечающих требованию способа наименьших квадратов:

$$\begin{cases} \sum_{i=1}^{n} y_i = na_0 + a_1 \sum_{i=1}^{n} x_i \\ \sum_{i=1}^{n} x_i y_i = a_0 \sum_{i=1}^{n} x_i + a_1 \sum_{i=1}^{n} x_i^2 \end{cases} \tag{9.3}$$

Для определения параметров уравнения регрессии воспользуемся приведенными в таблице 9.3 расчетами и подставим в систему нормальных уравнений (9.3) соответствующие данные:

$$\begin{cases} 47.2 = 10a_0 + 108a_1 \\ 539.1 = 108a_0 + 1236a_1 \end{cases}$$

Решая систему, получим $a_0 = 0{,}16$, $a_1 = 0{,}422$ т.е. линейное уравнение корреляции связи будет иметь вид:

$$y = 0{,}16 + 0{,}422x \tag{9.4}$$

Подставляя значения x_i в уравнение (9.4) находим теоретические, выровненные значения y_T, приведенные в последней графе таблицы 9.3.

Если в результате качественного анализа установлена зависимость, принимающая форму кривой второго порядка, то связь выражается уравнением кривой

$$y = a_0 + a_1x + a_2x^2 \tag{9.5}$$

Задача сводится к нахождению параметров a_0, a_1, a_2. Для этого необходимо решить систему трех нормальных уравнений:

$$\begin{cases} \sum\limits_{i=1}^{n} y_i = na_0 + a_1 \sum\limits_{i=1}^{n} x_i + a_2 \sum\limits_{i=1}^{n} x_i^2 \\ \sum\limits_{i=1}^{n} x_i y_i = a_0 \sum\limits_{i=1}^{n} x_i + a_1 \sum\limits_{i=1}^{n} x_i^2 + a_2 \sum\limits_{i=1}^{n} x_i^3 \\ \sum\limits_{i=1}^{n} x_i^2 y_i = a_0 \sum\limits_{i=1}^{n} x_i^2 + a_1 \sum\limits_{i=1}^{n} x_i^3 + a_2 \sum\limits_{i=1}^{n} x_i^4 \end{cases} \tag{9.6}$$

Кроме рассмотренных функций связи, в экономическом анализе часто применяются **степенная, показательная и гиперболическая функции.**

Степенная функция имеет вид: $y = a_0 x^{a_1}$ Параметр a_1 степенного уравнения называется показателем **эластичности** и показывает, на сколько процентов изменится **у** при возрастании **х** на 1%. При $x = 1$ $a_0 = y_0$. Для определения параметров степенной функции в начале ее приводят к линейному виду путем логарифмирования: $\lg y = \lg a_0 + a_1 \lg x_i$, а затем строят систему нормальных уравнений:

$$\begin{cases} \sum\limits_{i=1}^{n} \lg y = n \lg a_0 + a_1 \sum\limits_{i=1}^{n} \lg x_i \\ \sum\limits_{i=1}^{n} \lg y \lg x = \lg a_0 \sum\limits_{i=1}^{n} \lg x + a_1 \sum\limits_{i=1}^{n} \lg x_i^2 \end{cases} \tag{9.7}$$

Решив систему двух нормальных уравнений (9.7), находят логарифмы параметров логарифмической функции a_0 и a_1, а затем и сами параметры a_0 и a_1. При помощи степенной функции определяют, например, зависимость между фондом оплаты труда и выпуском продукции, затратами труда и выпуском продукции и т.д.

Если факторный признак **х** растет в **арифметической прогрессии**, а результативный **у** – в геометрической, то такая зависимость выражается **показательной функцией у** $= a_0 a_1^x$. Для определения параметров показательной функции ее также в начале приводят к линейному виду путем логарифмирования: $\lg y = \lg a_0 + x \cdot \lg a_1$, а затем строят систему нормальных уравнений:

$$\begin{cases} \sum\limits_{i=1}^{n} \lg y_i = n \lg a_0 + (\lg a_1) \sum\limits_{i=1}^{n} x_i \\ \sum\limits_{i=1}^{n} x_i \lg y_i = (\lg a_0) \sum\limits_{i=1}^{n} x_i + (\lg a_1) \sum\limits_{i=1}^{n} x_i^2 \end{cases} \tag{9.8}$$

Вычислив соответствующие данные и решив систему двух нормальных уравнений, находят параметры показательной функции a_0 и a_1

В ряде случаев **обратная связь** между факторным и результативным признаками может быть выражена уравнением **гиперболы**: $y = a_0 + a_1 / x$. И здесь задача заключается в нахождении параметров a_0 и a_1 при помощи системы двух нормальных уравнений:

$$\begin{cases} \sum_{i=1}^{n} y_i = na_0 + a_1 \sum_{i=1}^{n} \frac{1}{x_i} \\ \sum_{i=1}^{n} y_i \frac{1}{x_i} = a_0 \sum_{i=1}^{n} \frac{1}{x_i} + a_1 \sum_{i=1}^{n} \frac{1}{x_i^2} \end{cases} \tag{9.9}$$

При помощи гиперболической функции изучают, например, связь между выпуском продукции и себестоимостью, уровнем издержек обращения (в процентах к товарообороту) и товарооборотом в торговле, сроками уборки и урожайностью и т.д.

Таким образом, применение различных функций в качестве уравнения связи сводится к определению параметров уравнения по способу наименьших квадратов при помощи системы нормальных уравнений.

В малых совокупностях значение коэффициента регрессии подвержено случайным колебаниям. Поэтому возникает необходимость в определении **достоверности коэффициента регрессии.** Достоверность коэффициента линейной регрессии определяется так же, как и в выборочном наблюдении, т.е. устанавливаются средняя и предельная ошибки для выборочной средней и доли. Средняя ошибка коэффициента регрессии определяется по формуле:

$$\mu_B = \sqrt{\frac{\sigma_0^2}{\sigma^2 (n-1)}} \tag{9.10}$$

где σ_0^2 – случайная дисперсия, σ^2 – общая дисперсия, а **n** – число коррелируемых пар.

Например, имеются следующие уравнение корреляционной связи между стажем работы и выработкой: **y = 34,1 + 5,4x** при $\sigma_0^2 = 0{,}25$ и $\sigma^2 = 2$. Так как совокупность незначительна (10 групп рабочих), то коэффициент регрессии 5,4 подвержен случайным колебаниям и необходимо проверить его достоверность, для чего исчислим среднюю ошибку коэффициента регрессии по формуле (9.10):

$$\mu_B = \sqrt{\frac{0.25}{2(10-2)}} = \sqrt{\frac{0.25}{16}} = 0.125$$

Предельная ошибка коэффициента регрессии зависит от вероятности, с которой мы гарантируем, что коэффициент регрессии не выйдет за границы интервалов. При P = 0,954, t = 2, предельная ошибка коэффициента регрессии будет равна: $\Delta = t\mu_B = 2 \cdot 0{,}125 = 0{,}25$. Таким образом, коэффициент регрессии будет находиться в пределах: $a_1 = 5{,}4 \pm \Delta = 5{,}4 \pm 0{,}25$, т.е. не меньше 5,15 и не больше 5,65, что можно гарантировать с вероятностью 0,954.

9.5. Измерение тесноты связи

При изучении корреляционной связи важно выяснить не только форму, но и тесноту (**сопряженность**) связи между факторным и результативным признаками. Для этого статистикой установлен объективный числовой показатель, который вычисляется по определенным правилам. Чтобы измерить тесноту прямолинейной связи между двумя признаками, пользуются парным коэффициентом корреляции, который обозначается r_{xy}.

Так как, при корреляционной связи имеют дело не с приращением функции в связи с изменением аргумента, а с сопряженной вариацией результативных и факторных признаков, то определение тесноты связи, по существу, сводится к изучению этой сопряженности, т.е. того, в какой мере отклонение от среднего уровня одного признака сопряжено с отклонением другого. Это значит, что при наличии **полной прямой связи** все значения $(x_i - \overline{x})$ и $(y_i - \overline{y})$ должны иметь одинаковые знаки, при **полной обратной** – разные, при частичной связи знаки в преобладающем числе случаев будут совпадать, а при отсутствии связи – совпадать примерно в равном числе случаев.

Математически доказано, что $\sum\limits_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})$ достигает максимума при полной прямой связи и минимума при полной обратной связи. Причем

$$\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y}) = \sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2 \sum_{i=1}^{n}(y_i - \overline{y})^2}$$

поэтому теснота связи определяется как:

$$r_{xy} = \frac{\sum\limits_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum\limits_{i=1}^{n}(x_i - \overline{x})^2 \sum\limits_{i=1}^{n}(y_i - \overline{y})^2}} \qquad (9.11)$$

где r_{xy} – **линейный коэффициент корреляции**.

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

В зависимости от того, насколько r_{xy} приближается к ±1, различают связь слабую, умеренную, заметную, высокую, тесную и весьма тесную.

Пример 9.8.

Продолжим анализ данных, приведенных в таблице 9.4. Рассмотрим вычисление коэффициента корреляции по себестоимости ОФ и выпуску продукции по 10 предприятиям.

Таблица 9.5

Вычисление коэффициента корреляции

№ п/п	Стоимость ОФ (х), млн. руб.	Выпуск продукции (у), млн. руб.	$(x_i-\overline{x})$	$(y_i-\overline{y})$	$(x_i-\overline{x})\times(y_i-\overline{y})$	$(x_i-\overline{x})^2$	$(y_i-\overline{y})^2$
1	6	2,4	–4,8	–2,32	+11,136	23,04	5,3824
2	8	4,0	–2,8	–0,72	+2,016	7,84	0,5184
3	9	3,6	–1,8	–1,12	+2,016	3,24	1,2544
4	10	4,0	–0,8	–0,72	+0,576	0,64	0,5184
5	10	4,5	–0,8	–0,22	+0,176	0,64	0,0484
6	11	4,6	+0,2	–0,12	–0,024	0,04	0,0144
7	12	5,6	+1,2	+0,88	+1,056	1,44	0,7744
8	13	6,5	+2,2	+1,78	+3,916	4,84	3,1684
9	14	7,0	+3,2	+2,28	+7,296	10,24	5,1984
10	15	5,0	+4,2	+0,28	+1,176	17,64	0,0784
Итого	108	47,2			+29,34	69,60	16,956

$$\overline{x}=\frac{\sum_{i=1}^{n}x_i}{n}=\frac{108}{10}=10.8;\qquad \overline{y}=\frac{\sum_{i=1}^{n}y_i}{n}=\frac{47.2}{10}=4.72$$

$$r_{xy}=\frac{\sum_{i=1}^{n}(x_i-\overline{x})(y_i-\overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\overline{x})^2\sum_{i=1}^{n}(y_i-\overline{y})^2}}=\frac{29.34}{\sqrt{69.6\cdot 16.956}}=\frac{29.34}{\sqrt{1180.1376}}=\frac{29.34}{34.35}=+0.854$$

Таким образом, связь между стоимость основных фондов и выпуском продукции прямая и высокая.

Коэффициент корреляции r_{xy} применяется только в тех случаях, когда между явлениями существует прямолинейная связь. Если же связь не линейная, то пользуются **индексом корреляции**, который вычисляется по формуле:

$$R=\sqrt{1-\frac{\sum_{i=1}^{n}(y_i-y_T)^2}{\sum_{i=1}^{n}(y_i-\overline{y})^2}} \tag{9.12}$$

где $\mathbf{y}_i$ – первоначальные значения; $\overline{y}$ – среднее значение; y_T – теоретические (выравненные) значения переменной величины.

Показатель $\sigma_o^2 = \frac{\sum_{i=1}^{n}(y_i - y_T)^2}{n}$ – остаточная, случайная дисперсия. Она характеризует размер отклонений эмпирических значений результативного признака y от теоретических, т.е. случайную вариацию. $\sigma^2 = \frac{\sum_{i=1}^{n}(y_i - \overline{y})^2}{n}$ – общая дисперсия. По правилу сложения дисперсия общая дисперсия равна $\sigma^2 = \sigma_y^2 + \sigma_0^2$, где σ_y^2 – доля факторной вариации. Тогда, подставляя все в выражение (9.10), получаем

$$R = \sqrt{1 - \frac{\sum_{i=1}^{n}(y_i - y_T)^2}{\sum_{i=1}^{n}(y_i - \overline{y})^2}} = \sqrt{1 - \frac{\sigma_0^2}{\sigma^2}} = \sqrt{\frac{\sigma_y^2}{\sigma^2}} \qquad (9.13)$$

Таким образом, индекс корреляции характеризует долю факторной корреляции в общей с той только разницей, что вместо групповых средних берутся теоретические значения.

Если индекс корреляции возвестив квадрат, то получим **коэффициент детерминации** $R^2 = \frac{\sigma_y^2}{\sigma^2}$. Он характеризует роль факторной вариации в общей вариации и по построению аналогичен корреляционному отношению η^2. Критические значения коэффициента детерминации R^2 определяются по тем же таблицам, что и для корреляционного отношения η^2. Чтобы оценить существенность связи при помощи критерия Фишера, пользуются формулой

$$F = \frac{R^2}{1 - R^2} \cdot \frac{n - m}{m - 1} \qquad (9.14)$$

где m и n – имеют тоже значение, что и в формуле (9.2).

9.6. Методы измерения тесноты связи

Измерение тесноты (сопряженности) связи при помощи дисперсионного и корреляционного анализа связано с определенными сложностями и требует громоздких вычислений. Для ориентировочной оценки тесноты связи пользуются приближенными показателями, не требующими сложных, трудоемких расчетов. К ним относятся: **коэффициент корреляции знаков Фехнера**, **коэффициент корреляции рангов**, **коэффициент ассоциации** и **коэффициент взаимной сопряженности**.

Коэффициент корреляции знаков. Основан на сопоставлении знаков отклонений от средней и подсчете числа случаев совпадения и несовпадения знаков, а не на сопоставлении попарно размеров отклонений индивидуальных значений факторного и результативного признаков от средней $(x_i - \overline{y})$ и $(y_i - \overline{y})$.

Коэффициент корреляции знаков определяется по формуле

$$i = \frac{u - v}{u + v} \tag{9.15}$$

где **u** и **v** – число пар с одинаковыми и разными знаками, соответственно, отклонений **x** и **y** от $\overline{x}$ и $\overline{y}$. Коэффициент корреляции знаков колеблется в пределах от –1 до +1. Чем ближе коэффициент к 1, тем теснее связь. Если u > v, то i > 0, так как число согласованных знаков больше, чем несогласованных, и связь прямая. При u < v, i < 0, потому что число несогласованных знаков больше, чем согласованных, и связь обратная. Если u = v, то i = 0, и связи нет.

Пример.

Вычислим коэффициент корреляции знаков по десяти промышленным предприятиям, согласно данным, приведенных в таблице 9.4. Результат сведем в таблицу 9.6.

Таблица 9.6

Данные о стоимости ОФ и выпуску продукции некоторых предприятий

№ п/п	Стоимость ОФ (x) млн. руб.	Выпуск продукции (y) млн. руб.	Знак отклонения		Ранг			
			$(x_i - \overline{x})$	$(y_i - \overline{y})$	по x	по y	разность рангов	кв. разности рангов
A	1	2	3	4	5	6	7	8
1	6,0	2,4	–	–	1,0	1,0	0	0
2	8,0	4,0	–	–	2,0	3,5	–1,5	2,25
3	9,0	3,6	–	–	3,0	2,0	+1,0	1,00
4	10,0	4,0	–	–	4,5	3,5	+1,0	1,00
5	10,0	4,5	–	–	4,5	5,0	–0,5	0,25
6	11,0	4,6	+	–	6,0	6,0	0	0
7	12,0	5,6	+	+	7,0	8,0	–1,0	1,00
8	13,0	6,5	+	+	8,0	9,0	–1,0	1,00
9	14,0	7,0	+	+	9,0	10,0	–1,0	1,00
10	15,0	5,0	+	+	10,0	7,0	+3,0	9,00
Итого	108,0	47,2						16,5
Средняя	10,8	4,72						

Таким образом, u = 9; v = 1, $i = \frac{u-v}{u+v} = \frac{9-1}{9+1} = \frac{8}{10} = 0.8$. Это значит, что связь между стоимостью основных фондов и выпуском продукции прямая и высокая. Парный коэффициент корреляции, вычисленный выше по этим данным, равен +0,854 (см. табл. 9.4). Как видим, коэффициент корреляции знаков исчисляется очень просто и в этом его преимущество. Однако он не точен, так как учитывает только знаки отклонений, а не числовые значения отклонений.

Коэффициент корреляции рангов. Коэффициент корреляции рангов исчисляется не по первичным данным, а по рангам (порядковым номерам), которые присваиваются всем значениям изучаемых признаков, расположенных в порядке их возрастания. Если значения признаков совпадают, то определяется средний ранг путем деления суммы рангов на число значений.

Разберем порядок вычисления этого показателя на примере.

Пример 9.9.

Изучается товарооборот и суммы издержек обращения по ряду магазинов (в тыс. руб.). Данные представлены таблицей 9.7.

Таблица 9.7

Данные о товарообороте и издержках некоторых магазинов

№ магазина	Товарооборот	Издержки обращения
1	480	30
2	510	25
3	530	31
4	540	28
5	570	29
6	590	32
7	620	36
8	640	36
9	650	37
10	660	38

Из таблицы видно, что с ростом товарооборота растут и издержки обращения. График еще раз это подтверждает (рис. 9.2., стр. 132).

Но в ряде случаев увеличение товарооборота ведет и к уменьшению издержек обращения, поскольку, помимо двух названных величин, в реальном процессе торговли участвуют и другие факторы, которые в рассмотрение не включены и носят случайный характер. Рассмотрим критерий тесноты связи, названный показателем корреляции рангов

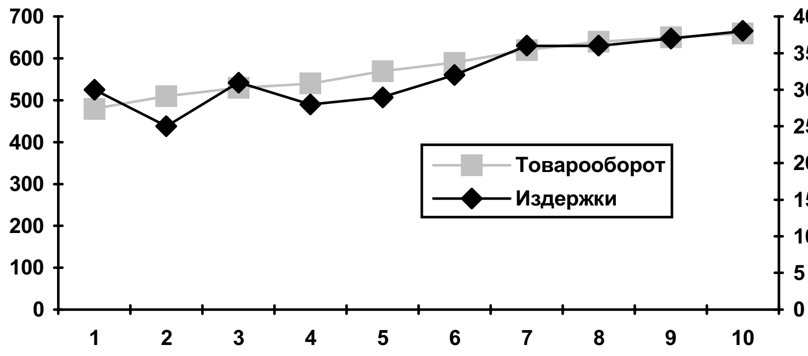


Рис. 9.2.

(порядковый номер, обозначается – r_i). От величин абсолютных перейдем к рангам по такому правилу: самое меньшее значение – ранг 1, затем 2 и т.д. Если встречаются одинаковые значения, то каждое из них заменяется средним. Итак:

Товарооборот (r_{ix})	1	2	3	4	5	6	7	8	9	10
Издержки (r_{iy})	4	1	5	2	3	6	7,5	7,5	9	10

Построим разности между рангами и возведем их в квадрат.

1. Если ранги совпадают, то ясно, что сумма их квадратов равна 0.

$d_i = r_{i_x} - r_{i_y}$; $\sum\limits_{i=1}^{n} d_i^2 = 0$ – связь полная, прямая.

2. Ранги образуют обратную последовательность

r_{ix}	1	2	3	4	5	6	7	8	9	10
r_{iy}	10	9	8	7	6	5	4	3	2	1

В этом случае $\sum\limits_{i=1}^{n} d_i^2 = \frac{n(n-1)}{3}$ (n – число единиц в совокупности) – связь полная, обратная.

3. Среднее значение из двух крайних означает полное отсутствие связи:

$$\left[0 + \frac{n(n^2-1)}{3}\right] : 2 = \frac{n(n^2-1)}{6}$$

4. Показатель корреляции рангов:

$$\rho = 1 - \frac{6\sum\limits_{i=1}^{n} d_i^2}{n(n^2-1)} \qquad (9.16)$$

Показатель показывает, как отличается полученная при наблюдении сумма квадратов разностей между рангами от случая отсутствия связи.

Проанализируем показатель корреляции рангов.

1. Связь полная и прямая, $\frac{6\sum d^2}{n\left(n^2-1\right)}=0$ и $\rho = 1$

2. Связь полная и обратная, $\frac{6\sum d^2}{n\left(n^2-1\right)}=2$ и $\rho = -1$

3. Все остальные значения лежат между –1 и +1.

Построим показатель корреляции рангов для нашего примера:

Товарооборот (ранг)	Издержки (ранг)	d	d^2
1	4	–3	9
2	1	1	1
3	5	–2	4
4	2	2	4
5	3	2	4
6	6	0	0
7	7,5	–0,5	0,25
8	7,5	0,5	0,25
9	9	0	0
10	10	0	0
			$\sum d^2 = 22{,}5$

$$\rho = 1-\frac{6\cdot 22{,}5}{10(100-1)}=+0{,}864$$

Полученный показатель свидетельствует о достаточно тесной связи между товарооборотом и издержками.

Пример 9.10.

Вычислим коэффициент корреляции рангов по данным о стоимости основных фондов и выпуске продукции (млн. руб.) (таблица 9.5). Соответствующие расчеты приведены в табл. 9.5 в колонках 5–8. Ранги стоимости основных фондов для предприятий №3, 4 определяются как средние между 4 и 5 рангами (4.5). Также для предприятий №2, 4 – среднее между 3 и 4 рангом. Тогда:

$$\rho = 1-\frac{6\sum_{i=1}^{n} d_i^2}{n\left(n^2-1\right)}=1-\frac{6.0\cdot 16.5}{10(100-1)}=1-\frac{99}{990}=1-0.1=0.9$$

Полученный ранговый коэффициент корреляции свидетельствует о наличии прямой тесной связи между величиной основных фондов и выпуском продукции.

Ранговый коэффициент корреляции более точный по сравнению с коэффициентом корреляции знаков, потому что он учитывает не только знаки отклонений, но и место величины признака в данном ряду.

Коэффициент ассоциации. Коэффициент ассоциации применяется для установления меры связи между двумя качественными альтернативными признаками. Для его вычисления строится комбинационная четырехклеточная таблица, которая выражает связь между двумя альтернативными явлениями.

Пример 9.11.

В табл. 9.8 приведены данные зависимости наличия отдельной квартиры от семейного положения. Вычислить коэффициенты, обозначив данные каждой клетки буквами a, b, c и d.

Таблица 9.8

Зависимость наличия отдельной квартиры от семейного положения

Семейное положение	Имеют отдел. квартиру	Не имеют отдел. квартиру	Всего
Семейные Одинокие	300 (a) 15 (c)	115 (b) 70 (d)	415 (a + b) 85 (c + d)
Всего	315 (a + c)	185 (b + d)	500

Коэффициент ассоциации рассчитывается по формуле:

$$A = \frac{ad - bc}{\sqrt{(a+b)(c+d)(a+c)(b+d)}} \qquad (9.17)$$

Подставив в формулу (9.17) необходимые значения из таблицы, получим

$$A = \frac{300 \cdot 70 - 15 \cdot 115}{485 \cdot 85 \cdot 315 \cdot 185} = 0.425$$

Это значит, что между семейным положением и обеспеченностью отдельной квартирой существует прямая умеренная связь.

Коэффициент ассоциации также изменяется от –1 до +1. Чем **А** ближе к единице, тем сильнее связаны между собой изучаемые признаки. При ad > bc – **связь прямая**, при ad < bc – **связь обратная**, при ad = bc, А = О и связь отсутствует.

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

Пример 9.12.

В результате опроса, проведенного в 1998 году ВЦИОМом (Всероссийский центр изучения общественного мнения) по отношению к личности Президента Российской Федерации мужчин и женщин, получены результаты, приведенные в табл. 9.9.

В данном случае распределение мужчин и женщин по отношению к Президенту России называется **условным распределением,** а итоговое распределение – **безусловным.**

Таблица 9.9

Данные общественного мнения

Пол	Положит.	Безразлич.	Отрицат.	Итого (f_1)	Вес				Итого
						Положит.	Безразлич.	Отрицат.	
Жен	10	5	65	80	w_{1j}	0,125	0,073	0,802	1
Муж	15	15	90	120	w_{2j}	0,125	0,125	0,75	1
Итого	25	20	155	200	w_j	0,125	0,1	0,775	1

Если бы в нашем примере пол не влиял на отношение к президенту, т.е. если бы между отношением к президенту и полом не было никакой связи, то частости (вес f) каждой строки были бы одинаковыми и частости условных и безусловного распределения совпадали.

Различия между условными распределениями и безусловным свидетельствует о влиянии факторного признака на распределение совокупности по результативному признаку, т.е. о наличии связи между факторным и результативным признаками, причем чем больше эти различия, тем в большей мере признаки связаны между собой, тем теснее связь между ними.

Для определения степени тесноты связи вычисляется специальный показатель, который называется **коэффициентом взаимной сопряженности**. Он определяется по следующей формуле:

$$C = \sqrt{\frac{\chi^2}{n(m_1 - 1)(m_2 - 1)}} \qquad (9.18)$$

где n – число единиц совокупности; m_1 и m_2 – число групп по первому и второму признакам ($m_1 = 2$ – муж, жен; $m_2 = 3$ – положит, отрицат, безразлич.); χ^2 – показатель **абсолютной квадратической сопряженности**

Пирсона, характеризует близость условных распределений к безусловным, который вычисляется по формуле:

$$\chi^2 = \sum_{i=1}^{m_1} f_j \sum_{i=1}^{m_1} \sum_{j=1}^{m_2} \frac{\left(w_{i,j} - w_j\right)^2}{w_j} \tag{9.19}$$

где f_i – общий вес в группах по первому признаку; $w_{i,j}$ – частости условного распределения в i – строке, w_j – частости безусловного распределения; j – номер столбца.

Если признаки независимы, то $w_{i,j} = w_j$, откуда $\chi^2 = 0$ и C = 0. Если же связь функциональная, то коэффициент взаимной сопряженности будет равен единице.

Рассчитаем коэффициент взаимной сопряженности для примера, приведенного в табл. 9.8.

$$\chi^2 = \left(\frac{(0{,}125 - 0{,}125)^2}{0{,}125} + \frac{(0{,}073 - 0{,}1)^2}{0{,}1} + \frac{(0{,}802 - 0{,}775)^2}{0{,}775}\right) \cdot 80 +$$

$$+\left(\frac{(0{,}125 - 0{,}125)^2}{0{,}125} + \frac{(0{,}125 - 0{,}1)^2}{0{,}1} + \frac{(0{,}125 - 0{,}775)^2}{0{,}775}\right) \cdot 120 =$$

$$= 6{,}282 + 66{,}169 = 72{,}475$$

$$C = \sqrt{\frac{\chi^2}{n(m_1 - 1)(m_2 - 1)}} = \sqrt{\frac{72{,}475}{200(2-1)(3-1)}} = \sqrt{\frac{72{,}475}{400}} = \sqrt{0{,}181} = 0{,}425$$

Таким образом, пол имеет существенное значение для определения отношения к Президенту РФ.

Контрольные вопросы

1. Понятие связи в статистике.
2. Основные методы изучения взаимосвязи.
3. Определение взаимосвязи с использованием дисперсионного анализа.
4. Критерий Фишера.
5. Корреляционный анализ, его задачи.
6. Аналитические выражения связи.
7. Использование линейной функции при определении связи.
8. Использование квадратической функции при определении связи.
9. Использование степенной, показательной и гиперболической функции при определении связи.
10. Коэффициент регрессии.
11. Измерение тесноты связи. Общее понятие.

12. Линейный коэффициент корреляции.
13. Индекс корреляции.
14. Коэффициент детерминации.
15. Методы измерения тесноты связи. Коэффициент корреляции знаков Фехнера.
16. Методы измерения тесноты связи. Коэффициент корреляции рангов.
17. Методы измерения тесноты связи. Коэффициент ассоциации.
18. Методы измерения тесноты связи. Коэффициент взаимной сопряженности. Коэффициент Пирсона.

ГЛАВА 10. РЯДЫ РАСПРЕДЕЛЕНИЯ

10.1. Виды рядов распределения

В результате обработки и систематизации первичных статистических данных получают ряды цифровых показателей, которые характеризуют изменение изучаемых явлений. Эти ряды называются **статистическими рядами.**

По своему содержанию статистические ряды подразделяются на ряды распределения и ряды динамики.

Рядом динамики называется ряд чисел, характеризующих последовательное изменение явления во времени.

Рядом распределения называется ряд чисел, характеризующих распределение единиц совокупности на группы по одному признаку.

Распределение единиц совокупности на группы по атрибутивному (качественному) группировочному признаку называется атрибутивным рядом распределения.

Распределение единиц совокупности на группы по количественному признаку, по степени возрастания или убывания числового значения признака называется **вариационным рядом**.

Если распределение произведено по дискретному (прерывному), выраженному целыми числами признаку, то такой вариационный ряд называется **дискретным**.

Если распределение совокупности произведено по количественному признаку, выраженному в виде материалов «от» и «до», то такой вариационный ряд называется **интервальным**.

Составными элементами каждого вариационного ряда являются два ряда чисел: 1) ряд вариантов и 2) ряд частот (вес) или частостей.

Варианты – отдельные числовые значения варьирующего признака**.**

Частоты (вес) – это абсолютные числа, показывающие, сколько раз встречается та или иная варианта в данной совокупности или, иначе говоря, это абсолютное число единиц в каждой группе. Частоты, выраженные в долях единицы или в процентах к итогу, называются **частостями**.

Пример 10.1.

Таблица 10.1

Распределение совокупности магазинов по числу рабочих мест

Число рабочих мест	Число магазинов	В % к итогу
1	4	20
2	8	40
3	5	25
4	2	10
5	1	5
Итого	20	100

В первой колонке показано число рабочих мест (1, 2, 3, 4, 5) – это расположены варианты, во второй колонке соответственно этим вариантам размещены частоты (веса), а в третьей – частости.

На основе вариационного ряда могут быть вычислены показатели вариации: средняя, дисперсия, среднеквадратическая, коэффициенты вариации, мода, медиана и т.д.

Вариационные ряды графически обычно изображаются при помощи гистограмм распределения или полигона распределения.

10.2. Моменты распределения

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

Кроме рассмотренных выше важной характеристикой рядов распределения являются также **моменты распределений**.

Моментом распределения ($M_К$) называется средняя арифметическая из отклонений значений признака **x** от некоторой постоянной величины **a** в степени **к**. Порядок момента определяется величиной **к.** Эмпирический момент **к**-го порядка определяется

по формуле
$$M_K = \frac{\sum_{i=1}^{n} (x_i - a)^k f_i}{\sum_{i=1}^{n} f_i} \qquad (10.1)$$

В зависимости от постоянной величины **a** различают **начальные, центральные и условные моменты**.

Если **a** = 0, то моменты называются **начальными** и определяются по формуле
$$V_K = \frac{\sum_{i=1}^{n} x_i^k f_i}{\sum_{i=1}^{n} f_i} \qquad (10.2)$$

В этом случае при **к** = 0 получим начальный момент нулевого порядка, который равен:

$$V_0 = 1$$

($x_i^0 = 1$, а f_i – сокращаются).

При **к** = 1 получим начальный момент первого порядка, который равен:

$$V_1 = \frac{\sum_{i=1}^{n} x_i^1 f_i}{\sum_{i=1}^{n} f_i} = \bar{x} \tag{10.3}$$

При **к** = 2 - начальный момент второго порядка, равный

$$V_2 = \frac{\sum_{i=1}^{n} x_i^2 f_i}{\sum_{i=1}^{n} f_i} = \overline{x^2} \tag{10.4}$$

и т.д. Начальные моменты используются, в частности, при расчете дисперсии: $\sigma^2 = V_2 - V_1^2$, откуда $\sigma^2 = \overline{x^2} - \bar{x}^2$.

Если постоянная величина $a = \bar{x}$, то получим **центральные моменты,** которые определяются по формуле

$$M_K = \frac{\sum_{i=1}^{n} \left(x_i - \bar{x}\right)^k f_i}{\sum_{i=1}^{n} f_i} \tag{10.5}$$

В этом случае при **к** = 0 получим центральный момент нулевого порядка, который равен

$$M_0 = \frac{\sum_{i=1}^{n} \left(x_i - \bar{x}\right)^0 f_i}{\sum_{i=1}^{n} f_i} = 1,$$

при **к** = 1 – центральный момент первого порядка, равный:

$$M_1 = \frac{\sum_{i=1}^{n} \left(x_i - \bar{x}\right)^1 f_i}{\sum_{i=1}^{n} f_i} = 0,$$

(по свойству – сумма всех отклонений равна нулю), при **к** = 2 – центральный момент второго порядка, равный:

$$M_2 = \frac{\sum_{i=1}^{n} \left(x_i - \bar{x}\right)^2 f_i}{\sum_{i=1}^{n} f_i} = 2,$$

(дисперсии) и являющийся мерой колеблемости признака и т.д.

Если постоянная величина равна **а**, то моменты называются **условными** и определяются по формуле (10.1).

В связи с тем, что вычисление центральных моментов, которыми довольно часто пользуются для характеристики рядов распределений, довольно громоздко, вначале вычисляют условные моменты, а затем по специальным формулам переходят от условных моментов к центральным. Так, математически доказано, что:

$$\sigma^2 = m_3 = M_2 - M_1^2;$$

$$\sigma^3 = m_3 = M_3 - 3M_2 M_1 + 2M_1^2;$$

$$\sigma^4 = m_4 = M_4 - 4M_3 M_2 + 6M_2 M_1^2 - 3M_1^3$$

Условные моменты используются для определения дисперсий высоких степеней σ^2, σ^3, σ^4. Практически используются моменты первых четырех порядков. Если в качестве весов взять не частоты, а **вероятности**, то получим **теоретические моменты распределения**.

10.3. Кривые распределения

Кривая распределения – линия на плоскости, отражающая зависимость между значениями рассматриваемой случайной величины и соответствующими им числами наблюдений.

Если на оси абсцисс откладывать значения варьирующего признака, а на оси ординат частоты, то, соединяя эти точки, получаем **эмпирическую кривую распределения**. Пользуются также **кумулятивной кривой** распределения, указывающей для каждого данного значения **x** частоту тех значений, которые не превосходят **x**. Кривая распределения служит отправным пунктом статистического исследования варьирующего признака. Она является обобщенной характеристикой особенностей формы распределения. Кривая распределения выражает закономерность распределения единиц совокупности по величине варьирующего признака. Различают **эмпирические и теоретические** кривые распределения**. Эмпирическая кривая –** это фактическая кривая распределения, полученная по данным наблюдения, в которой отражаются как общие, так и случайные условия, определяющие распределение. **Теоретическая кривая распределения –** это кривая, выражающая функциональную связь между изменением варьирующего признака и изменением частот и характеризующая определенный тип распределения**.**

Кривые распределения бывают **симметричными и асимметричными**. В зависимости от того, какая ветвь кривой распределения вытянута – правая или левая – различают **правостороннюю** или **левостороннюю асимметрию**. Кривые распределения могут быть **одно-, двух- и многовершинными**.

В нормальном ряду распределения размах вариации R = 6σ; $\sigma = 1{,}25\bar{d}$; $\bar{x} = M_0 = M_e$. Если указанные соотношения, нарушены, то это свидетельствует о наличии асимметрии распределения. Так, при $M_0 < M_e < \bar{x}$ разности между $\bar{x} - M_0$ и $\bar{x} - M_e$ положительные и асимметрия правосторонняя, а при $M_0 > M_e > \bar{x}$, наоборот, разности между $\bar{x} - M_Q$ и $\bar{x} - M_e$ отрицательные и симметрия левосторонняя.

В симметричном распределении центральный момент третьего порядка $m_3 = 0$, поэтому чем он больше, тем больше и асимметрия. Эта особенность и используется для характеристики асимметрий. Коэффициент асимметрии равен отношению центрального момента третьего порядка к среднему квадратическому отклонению в кубе:

$$A_S = \frac{m_3}{\sigma^3} \qquad (10.6)$$

Если A > 0, то асимметрия правосторонняя, а если A < 0, то асимметрия левосторонняя. Чем числитель ближе к 0, тем асимметрия меньше (рис. 10.1).

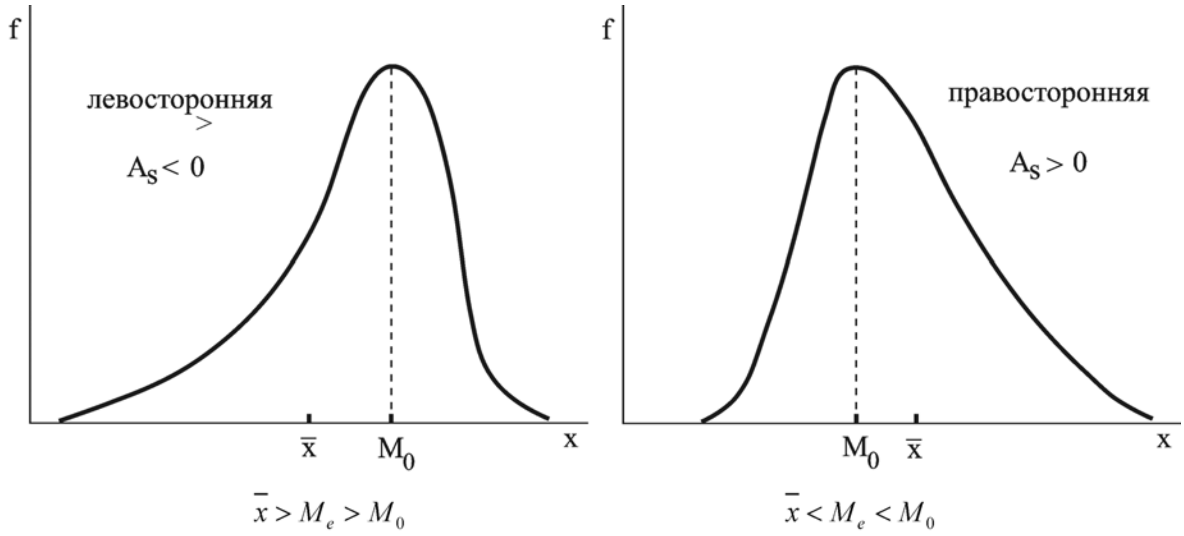


Рис. 10.1.

Английский статистик К. Пирсон на основе разности между средней величиной и модой предложил другой показатель асимметрии

$$A_{SП} = \frac{\bar{x} - M_0}{\sigma} \qquad (10.7)$$

Показатель Пирсона зависит от степени асимметричности в средней части ряда распределения, а показатель асимметрии, основанный на моменте третьего порядка (формула (10.6)), – от крайних значений признака.

С помощью момента четвертого порядка характеризуется еще более сложное свойство рядов распределения, чем асимметрия, называемое **эксцессом**.

Показатель эксцесса рассчитывается по формуле

$$E_X = \frac{m_4}{\sigma^4} \tag{10.8}$$

Часто эксцесс интерпретируется как «крутизна» распределения, но это неточно и неполно. График распределения может выглядеть сколь угодно крутым в зависимости от силы вариации признака: чем слабее вариация, тем круче кривая распределения при данном масштабе. Не говоря уже о том, что, изменяя масштабы по оси абсцисс и по оси ординат, любое распределение можно искусственно сделать «крутым» и «пологим». Чтобы показать, в чем состоит эксцесс распределения, и правильно его интерпретировать, нужно сравнить ряды с одинаковой силой вариации (одной и той же величиной σ) и разными показателями эксцесса. Чтобы не смешать эксцесс с асимметрией, все сравниваемые ряды должны быть симметричными. Такое сравнение изображено на рис. 10.2.

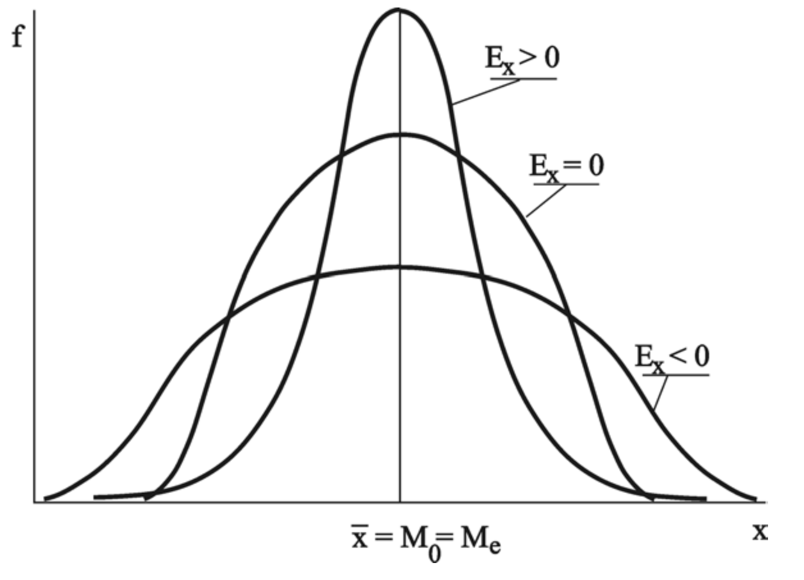


Рис. 10.2

Для вариационного ряда с нормальным распределением значений признака показатель эксцесса, рассчитанный по формуле (10.8), равен трем.

Однако такой показатель не следует называть термином «эксцесс», что в переводе означает «излишество». Термин «эксцесс» следует применять не к самому отношению по формуле (10.8), а к сравнению такого отношения для изучаемого распределения с величиной данного отношения нормального распределения, т.е. с величиной 3. Отсюда окончательные формулы показателя эксцесса, т.е. излишества в сравнении с нормальным распределением при той же силе вариации, имеют вид:

$$E_X = \frac{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^4 f_i}{\sigma^4 \sum_{i=1}^{n} f_i} - 3 \tag{10.10}$$

Все рассмотренные нами показатели характеризуют отдельные свойства совокупности. Общую характеристику ряда распределения можно представить аналитически, в виде функции, характеризующей зависимость между изменениями варьирующего признака и частотами. Каждому ряду свойственна определенная закономерность, выражением которой является кривая распределения, представляющая собой функцию распределения. Если имеется эмпирический ряд распределения, то необходимо найти функцию распределения, т.е. подобрать такую теоретическую кривую распределения, которая наиболее полно отражала бы закономерность распределения. Нахождение функции кривой распределения называется **моделированием**. Моделирование имеет большое познавательное значение, функции кривой распределения дают в компактной форме характеристику изучаемой совокупности, ее закономерности, сглаживают различные «неправильности» эмпирического ряда, возникшие вследствие случайных обстоятельств, дают возможность находить частоты интервалов, которые не встречались в эмпирическом распределении. Уравнения кривых распределения позволяют при помощи двух-трех, а иногда и одной сводной характеристики получить представление о характере распределения, они весьма важны для прогнозирования будущих распределений.

Для аппроксимации эмпирических кривых распределения и сопоставления их с теоретическими в статистике часто пользуются **нормальным распределением, распределением Пуассона** и **биноминальным распределением**. Мы рассмотрим только нормальное распределение.

Если случайная непрерывная величина имеет плотность распределения:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{\frac{\left(x_i - \bar{x}\right)^2}{2\sigma^2}}$$

то она подчиняется закону нормального распределения. Нормальное распределение описывается интегральной функцией распределения

$$F_{(x)} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{\frac{t^2}{2}} dt \tag{10.11}$$

где $t = \dfrac{x_i - \bar{x}}{\sigma}$ – нормированное отклонение.

Особенности кривой нормального распределения:

1. Кривая симметрична относительно максимальной ординаты, которая соответствует $\overline{x} = M_0 = M_e$, ее величина равна $\frac{1}{\sqrt{2\pi\sigma}}$;

2. Кривая асимптотически приближается к оси абсцисс, продолжаясь в обе стороны до бесконечности. При этом, чем больше значения отклоняются от $\overline{x}$, тем реже они встречаются;

3. Равновероятны одинаковые по абсолютному значению, но противоположные по знаку отклонения значений переменной x от $\overline{x}$;

4. Кривая имеет две точки перегиба, находящиеся на расстоянии $\pm\sigma$ от $\overline{x}$;

5. При $\overline{x}$ = const с увеличением σ кривая становится более пологой. При σ = const с изменением $\overline{x}$ кривая не меняет свою форму, а лишь сдвигается вправо или влево по оси абсцисс;

6. В промежутке $\overline{x} \pm \sigma$ (при t = l) находится 68,3% всех значений признака;

в промежутке $\overline{x} \pm 2\sigma$ (при t = 2) находится 95,4% всех значений признака;

в промежутке $\overline{x} \pm 3\sigma$ (при t = 3) – 99,7% всех значений признака.

В уравнении (10.11) $F_{(x)}$ рассматривается как функция переменной t: каждому значению t соответствует определенное значение $F_{(x)}$ (табулированное). Таким образом, кривую нормального распределения можно построить по двум параметрам $\overline{x}$ и σ. Вычислив по ним нормированное отклонение t, можно по специальной таблице получить готовые значения $F_{(x)}$ при условии, что t > 0. Например, если x = 115, $\overline{x}$ = 100, σ = 12, то $t = \frac{115-100}{12}$ = 1,25. По таблице данному значению *t* соответствует F (+1,25) = 0,894. Так как кривая нормального распределения симметрична, то $F_{(-t)} = 1 - F_{(t)}$ и F(–l,25) = 1 – 0,894 = 0,106. Следовательно вероятность того, что признак примет значение в интервале от x_i до x_{i+1}, равна $F(x_{i+1}) - F(x_i)$. Важное значение в анализе имеет проверка того, насколько фактическое распределение признака соответствует теоретическому, например нормальному распределению. Задача состоит в том, чтобы по фактическим данным вычислить теоретические частоты кривой распределения и сравнить их с эмпирическим рядом. Теоретические частоты (вес) f^1 для нормального распределения рассчитываются по формуле

$$f^1 = n\left(F_{(x_i+1)} - F_{x_i}\right) \tag{10.12}$$

где $n = \sum_{i=1}^{n} f_i$ – объем совокупности.

Пример. 10.1.

В табл. 10.2 показано распределение ткачих по степени выполнения норм выработки. На этом примере покажем расчет теоретических частот.

Таблица 10.2

Распределение ткачих по степени выполнения норм выработки

№ п/п	Группы ткачих, выполн. норму, %(x)	Число ткачих (f)	Норм. отклоне-ние,	Функция распределе-ния	Вероят-ность	Вес (f^1)
1	до 100	2	–1,67	0,047	0,047	5
2	100–110	15	–0,94	0,174	0,127	13
3	110–120	20	–0,22	0,413	0,230	24
4	120–130	32	+0,51	0,095	0,282	28
5	130–140	12	+1,23	0,891	0,196	20
6	140–150	9	+1,96	0,975	0,084	8
7	свыше 150	4	–	1,000	0,025	2
Итого	–	100	–	–	–	100

Среднее выполнение норм $\overline{x}$ = 123%, σ = 13,8.

Расчет производится следующим образом. По верхним границам интервалов определяется нормированное отклонение t. Так, для первой группы $t_1 = \frac{100-123}{13.8}$ = –1,67, для второй $t_2 = \frac{110-123}{13.8}$ = –0,94 и т.д. Затем по таблице (**см. приложение 1**) находим значение функции распределения F(t), потом вероятность $F(t_i+1) - F(t_i)$ и теоретические частоты f^1. Значению t_1 = –1,67 соответствует по таблице $F_{(t)}$ = 0,953. Но из симметричности нормального распределения следует, что $F_{(-t)} = 1 - F_{(t)}$, т.е. $F_{(-1.67)}$ = 1 – 0,953 = 0,047. Для t_2 = – 0,94 $F_{(t)}$ = 0,826, откуда $F_{(-0.94)}$ = 0,174 и т.д. Вероятность того, что признак имеет значение в интервале от x_i до x_{i+1}, будет равна для первой группы 0,047, для второй – $F(x_2) - F(x_1)$ = 0,174 – 0,047 = 0,127 и т.д. Теперь определяем теоретические частоты: $f^1 = n(\ F(x_{i+1}) - F(x_i)\ )$ = 100 · 0,047 ≅ 5,0; 100 · 0,127 ≅ 13 и т.д.

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

Эмпирические и теоретические распределения ткачих по степени выполнения норм показаны на рис. 10.3.

Коротко (для информации) рассмотрим **распределение Пуассона** и **биномиальное распределение**.

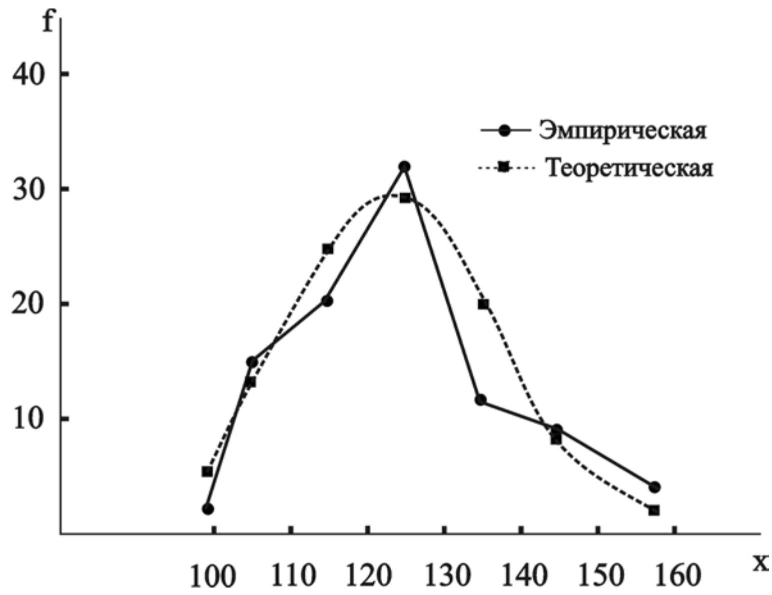


Рис. 10.3.

При рассмотрении маловероятных событий, имеющих место в большой серии независимых испытаний некоторое (конечное) число раз, вероятности появления этих событий подчиняются **закону Пуассона** или закону редких событий

$$P_m = \frac{\lambda \cdot e^{-\lambda}}{m!}$$

где λ равна среднему числу появления событий в одинаковых независимых испытаниях, т.е. $\lambda = \mathbf{n} \cdot \mathbf{p}$, где **p** – вероятность события при одном испытании, m – частота данного события.

Закон Пуассона можно применять для совокупностей, достаточно больших по объему ($n \geq 100$) и имеющих достаточно малую долю единиц, обладающих данным признаком ($p \leq 0,1$).

При этом распределение Пуассона можно применить, когда на только не известно значение n – общего числа возможных результатов, но и когда не известно конечное число, которое n может представлять.

Биномиальное распределение – это распределение именуемое так из-за его отношения к разложению двучлена $(p + q)^n$.

Биномиальное распределение есть распределение вероятности исходов события, которые могут быть классифицированы как положительные или отрицательные, т.е. оно связано с обстоятельствами, в которых какое-либо специфическое событие может или случиться, или не случиться. Здесь нет места для полумер, и не принимается в расчет степень интенсивности события. Общая вероятность события, случающегося или не случающегося, равна 1. Поэтому если вероятность того, что оно случится, равна **p**, то вероятность того, что оно не случится, равна q = 1 – p, p + q = 1.

Члены **p** и **q** относятся к вероятности наступления или не наступления только одного события. Вероятность наступления двух отдельных событий по закону .умножения равна для независимых явлений: $\mathbf{p \cdot p = p^2}$, т.е. 0,01 при **p** = 0,1. Вероятность не наступления каждого из двух событий составляет $\mathbf{q^2}$ или 0,81. Но возникает новая серия возможностей, поскольку в первом случае возможность того, что событие случится, связана с тем, что оно не случится во втором случае; вероятность этого равна $\mathbf{p \cdot q}$ или 0,09. Аналогичным образом вероятность наступления события во втором случае, связанная с не наступлением в первом случае, также равна 0,09. Поэтому все вероятности в целом таковы:

- наступление события в обоих случаях – $\mathbf{p^2}$ = 0,01;
- не наступление в обоих случаях – $\mathbf{q^2}$ = 0,81;
- наступление в первом и не наступление во втором – pq = 0,09;
- наступление во втором и не наступление в первом – qp = 0,09;

Общая вероятность всех исходов $p^2 + 2pq + q^2 = 1,0$.

Общая вероятность, таким образом, может быть алгебраически представлена как $p^2 + 2pq + q^2$ или $(p + q)^2$. Соответственно там, где имеются три события, вероятности будут составлять:

- наступление события в трех случаях – p^3 – 0,001;
- не наступление события в трех случаях – q^3 = 0,729;
- наступление в двух случаях и не наступление в одном – $3p^2q$ = 0,027;
- наступление в одном случае и не наступление в двух – $3pq^2$ = 0,243.

Общая вероятность всех исходов $(p + q)^3 = 1,0$.

Биномиальное, Пуассоново и нормальное распределения являются главными из тех форм распределения, которыми пользуются для выполнения значительной части статистической работы.

10.4. Критерий согласия

Для того, чтобы установить, подчиняется ли эмпирическое распределение закону нормального распределения, необходимо сравнить его с теоретическим распределением. Если теоретическое и эмпирическое распределения частот близки между собой, т.е. разности $f - f^1$ незначительны, то эмпирическое распределение подчиняется закону нормального распределения. Поэтому важно установить, являются ли разности между эмпирическими и теоретическими частотами результатом действия случайных причин или эта разница существенна и обусловлена неправильно подобранной функцией.

Для оценки близости эмпирического отношения к теоретическому нормальному пользуются специальными **показателями**, которые называются **согласия**. Они разработаны Пирсоном, Колмогоровым, Романовским и Ястремским. Наиболее простым и доступным является критерий Пирсона:

$$\chi^2 = \frac{\left(f_1 - f_1^1\right)^2}{f_1^2} + \frac{\left(f_2 - f_2^2\right)^2}{f_2^2} + \ldots + \frac{\left(f_n - f_n^2\right)^2}{f_n^2} = \sum_{i=1}^{n} \frac{\left(f_i - f_i^1\right)^2}{f_i^2} \quad (10.13)$$

Чем меньше отклонение между эмпирическими и теоретическими частотами, тем меньше значение χ^2, а значит, теоретическое распределение лучше воспроизводит эмпирическое, и наоборот. Если эмпирические частоты совпадают с теоретическими, $\boldsymbol{\chi}^2 = 0$.

Вычисление $\boldsymbol{\chi}^2$ Пирсона связано с показателем, который называется **числом степеней свободы.** Под числом степеней свободы **К** понимают количество независимых величин, которые могут принимать независимые значения, не изменяющие заданные характеристики.

Так, если средняя выработка рабочих участка равна

$$\bar{x} = \frac{230 + 300 + 275 + 325 + 300 + 320}{6} = 300$$

деталей, то пять из шести могут быть произвольными, а шестое должно быть единственно возможным, при котором средняя выработка 300 деталей останется без изменений. В данном случае задан один параметр ($\bar{x}$), поэтому число степеней свободы будет равно: 6 – 1 = 5.

В нашем примере распределения ткачих по степени выполнения норм выделено семь групп, функция нормального распределения имеет два параметра: $\bar{x}$ и σ . Кроме того, исчисление критерия χ^2 связано с ограничительным условием: $\sum_{i=1}^{n} f_i = \sum_{i=1}^{n} f_i^1$. Следовательно, число степеней свободы будет равно: К = 7 – 2 – 1 = 4. Обозначим число групп **m**, число параметров r и получим **К = m – r – 1** . По специальной таблице (**см. приложение 2**) находим значение χ^2, соответствующее данному числу степеней свободы и заданной вероятности. По этой же таблице для заданного критерия χ^2 при разных значениях степеней свободы можно определить вероятность того, что расхождение между теоретическими и эмпирическими частотами изучаемого ряда является случайным. Если фактическое значение критерия χ^2 меньше табличного, то отклонение между эмпирическими и теоретическими частотами являются случайными, несущественными, и можно сделать вывод о том, что теоретическое распределение хорошо воспроизводит эмпирическое, и, наоборот, если фактическое значение больше табличного, то отклонения являются существенными и эмпирический ряд распределения не подчиняется закону нормального распределения.

Контрольные вопросы

1. Виды рядов распределения.
2. Моменты распределения.
3. Виды кривых распределения.
4. Понятие эксцесса кривой распределения.
5. Аппроксимация эмпирической кривой распределения.
6. Использования функции нормального распределения при аппроксимации эмпирических кривых.
7. Критерии согласия при аппроксимации эмпирических кривых.

ГЛАВА 11. РЯДЫ ДИНАМИКИ

11.1. Ряды Динамики. Установление вида ряда динамики

Основная цель статистического изучения динамики коммерческой деятельности состоит в выявлении и измерении закономерностей их развития во времени. Это достигается посредством построения и анализа статистических рядов динамики.

Рядами динамики называются статистические данные, отображающие развитие изучаемого явления во времени. В каждом ряду динамики имеются два основных элемента: **показатель времени t**; соответствующие им **уровни** развития **изучаемого явления y**. В качестве показаний времени в рядах динамики выступают либо определенные даты (моменты) времени, либо отдельные периоды (годы, кварталы, месяцы, сутки).

Уровни рядов динамики отображают количественную оценку (меру) развития во времени изучаемого явления. Они могут выражаться абсолютными, относительными или средними величинами.

В зависимости от характера изучаемого явления уровни рядов динамики могут относиться или к определенным датам (моментам) времени, или к отдельным периодам. В соответствии с этим, ряды динамики подразделяются на **моментные и интервальные**.

Моментные ряды динамики отображают состояние изучаемых явлений на определенные даты (моменты) времени.

Примером моментного ряда динамики является следующая информация о списочной численности работников фирмы N в 1994 г.:

Таблица 11.1

Списочной численности работников фирмы N

Дата	1.01	1.04	1.07	1.10	1.01
Год	1994	1994	1994	1994	1995
Число работников, чел.	192	190	195	198	200

Особенностью моментного ряда динамики является то, что в его уровни могут входить одни и те же единицы изучаемой совокупности. Так, основная часть персонала фирмы N, составляющая списочную численность на 1.01.1994 г., продолжающая работать в течение данного года, отображена в уровнях последующих периодов. Поэтому при суммировании уровней моментного ряда динамики может возникнуть повторный счет.

Интервальные ряды динамики отображают итоги развития (функционирования) изучаемых явлений за отдельные периоды (интервалы) времени.

Примером интервального ряда динамики могут служить данные о розничном товарообороте магазина в 1990–1994 гг.:

Таблица 11.2

Розничном товарообороте магазина в 1990–1994 гг.

Год	1990	1991	1992	1993	1994
Объем розничного товарооборота, тыс. руб.	885,7	932,6	980,1	1028,7	1088,4

Особенностью интервального ряда динамики является то, что каждый его уровень складывается из данных за более короткие интервалы времени. Например, суммируя товарооборот за первые три месяца года, получают его объем за I квартал, а сумма товарооборота четырех кварталов дает объем товарооборота за год и т.д.

Ряды динамики могут быть **полными и неполными**.

Полный ряд – ряд динамики, в котором одноименные моменты времени или периоды времени строго следуют один за другим в календарном порядке или равноотстоят друг от друга.

Неполный ряд динамики – ряд, в котором уровни зафиксированы в не равноотстоящие моменты или периоды времени.

Пример. 11.1.

Численность населения СССР характеризуется данными переписей, млн. чел.:

1939	1959	1970	1979	неполный моментный ряд
170,6	208,8	241,7	262,4	абсолютных величин

Пример.11.2.

Производство электроэнергии характеризуется следующими данными, млрд. кВт-ч.:

1930	1940	1950	1960	полный интервальный ряд
48,6	91,2	292,3	740,9	абсолютных величин

11.2. Приведение рядов динамики в сопоставимый вид

Ряды динамики, изучающие изменение статистического показателя, могут охватывать значительный период времени, на протяжении которого могут происходить события, нарушающие сопоставимость отдельных уровней ряда динамики (изменение методологии учета, изменение цен и т.д.).

Для того чтобы анализ ряда был объективен, необходимо учитывать события, приводящие к несопоставимости уровней ряда и использовать приемы обработки рядов для приведения их в сопоставимый вид.

Наиболее характерные случаи несопоставимости уровней ряда динамики:

Территориальные изменения объекта исследования, к которому относится изучаемый показатель (изменение границ городского района, пересмотр административного деления области и т.д.).

Разновеликие интервалы времени, к которым относится показатель. Так, например, в феврале – 28 дней, в марте – 31 день, анализируя изменения показателя по месяцам, необходимо учитывать разницу в количестве дней.

Изменение даты учета. Например, численность поголовья скота в разные годы могла определяться по состоянию на 1 января или на 1 октября, что в данном случае приводит к несопоставимости.

Изменение методологии учета или расчета показателя.

Изменение цен.

Изменение единиц измерения.

Пример. 11.3.

Динамика изменения численности населения района области по состоянию на 1 января (в тыс. человек) представлена рядом динамики:

1982	1983	1984
22,0	22,3	22,8 – в старых границах района.

В 1984 году произошло изменение административного деления области, и площадь района увеличилась, соответственно увеличилась и численность населения района:

1985	1986	1987
34,2	34,3	34,4 – в новых границах района.

Для приведения ряда в сопоставимый вид необходимо для 1984 года знать численность населения в старых и новых границах района для определения коэффициента пересчета:

$$K = \frac{34{,}2}{22{,}8} = 1{,}5$$

Все уровни ряда, предшествующие 1984 году, умножаются на коэффициент К и ряд принимает вид:

1982	1983	1984	1985	1986	1987
33,0	33,3	34,2	34,2	34,3	34,4

После этого преобразования ряда динамики возможен дальнейший анализ ряда (определение темпов роста и др.).

11.3. Определение среднего уровня ряда динамики

В качестве обобщенной характеристики уровней ряда динамики служит средний уровень ряда динамики $\overline{y}$. В зависимости от типа ряда динамики используются различные расчетные формулы.

Интервальный ряд абсолютных величин с равными периодами (интервалами времени):

$$\overline{y} = \frac{\sum_{i=1}^{n} y_i}{n} \tag{11.1}$$

Пример 11.4.

Таблица 11.3

Выпуск блюд комбинатом общественного питания, тыс. блюд

	Годы					
	1995	1996	1997	1998	1999	2000
Произведено блюд	2437,2	2657,0	2907,9	3144,4	3295,2	3477,9

Рассчитаем, сколько в среднем за год было выпущено блюд за период с 1995 по 2000 г.

$$\overline{y} = \frac{\sum_{i=1}^{n} y_i}{n} = \frac{2437.2 + 2657.0 + 2907.9 + 3144.4 + 3295.2 + 3477.9}{6} =$$

= 2986,6 тыс. блюд в год.

Моментный ряд с равными интервалами между датами (средняя хронологическая):

$$\overline{y} = \frac{\frac{1}{2} y_1 + y_2 + \ldots + y_{n-1} + \frac{1}{2} y_n}{n-1} \tag{11.2}$$

y_1 – начальный уровень ряда; y_n – конечный уровень ряда; n – число дат.

Пример 11.5.

Таблица 11.4

Товарный запас крупного торгового предприятия на конец года, тыс. руб.

	Годы				
	1996	1997	1998	1999	2000
Товарный запас	26528	27567	29073	31561	35253

Среднегодовой запас товаров предприятия за пятилетний период составил следующую сумму:

$$\overline{y}_{xp} = \frac{\frac{26528}{2} + 27567 + 29073 + 31561 + \frac{35253}{2}}{5-1=4} =$$

= 29772,9 тыс. руб.

Моментный ряд с неравными интервалами между датами:

$$\overline{y} = \frac{\sum_{i=1}^{n} \frac{(y_i + y_{i+1})}{2} t_i}{\sum_{i=1}^{n} t_i} \qquad (11.3)$$

где y_i, y_{i+1} – уровни ряда, сохраняющиеся без изменения на протяжении интервала времени t_i.

Пример 11.6.

Таблица 11.5

Средняя численность работников предприятий розничной торговли РФ, тыс. чел.

Годы	Численность работников
1970	2203
1980	2802
1990	2768
1995	3136
2000	3109

Определим средний уровень численности работников предприятий розничной торговли:

$$\overline{y} = \frac{\sum_{i=1}^{n} \frac{(y_i + y_{i+1})}{2} t_i}{\sum_{i=1}^{n} t_i} =$$

$$= \frac{\frac{2203+2802}{2}10 + \frac{2802+2768}{2}10 + \frac{2768+3136}{2}5 + \frac{3136+3109}{2}5}{30} = 2603$$

$\overline{y}$ = 2603 тыс. чел.

11.4. Показатели изменения уровней ряда динамики

Одним из важнейших направлений анализа рядов динамики является изучение особенностей развития явления за отдельные периоды времени.

С этой целью для динамических рядов рассчитывают ряд показателей:

К – темпы роста;

Δу – абсолютные приросты;

ΔК – темпы прироста.

Темп роста – относительный показатель, получающийся в результате деления двух уровней одного ряда друг на друга. Темпы роста могут рассчитываться как **цепные** (с переменной базой), когда каждый уровень ряда сопоставляется с предшествующим ему уровнем:

$$K_{\text{iц}} = \frac{y_i}{y_{i-1}} \tag{11.4}$$

либо как **базисные** (с постоянной базой), когда все уровни ряда сопоставляются с одним и тем же уровнем y_0, выбранным за базу сравнения:

$$K_{\text{iб}} = \frac{y_i}{y_0} \tag{11.5}$$

Темпы роста могут быть представлены в виде коэффициентов либо в виде процентов.

Абсолютный прирост – разность между двумя уровнями ряда динамики, имеет ту же размерность, что и уровни самого ряда динамики. Абсолютные приросты могут быть цепными и базисными, в зависимости от способа выбора базы для сравнения:

цепной абсолютный прирост: $$\Delta y_{i\text{ц}} = y_i - y_{i-1} \tag{11.6}$$

базисный абсолютный прирост: $$\Delta y_{\text{iб}} = y_i - y_0 \tag{11.7}$$

Для относительной оценки абсолютных приростов рассчитываются показатели темпов прироста.

Темп прироста – относительный показатель, показывающий на сколько процентов один уровень ряда динамики больше (или меньше) другого, принимаемого за базу для сравнения.

Базисные темпы прироста: $$\Delta K_{\text{iб}} = \frac{\Delta y_{i\text{б}}}{y_0} \tag{11.8}$$

Цепные темпы прироста: $$\Delta K_{\text{iц}} = \frac{\Delta y_{i\text{ц}}}{y_{i-1}} \tag{11.9}$$

$\Delta y_{\text{iб}}$ и $\Delta y_{\text{iц}}$ – абсолютный базисный или цепной прирост; y_0 – уровень ряда динамики, выбранный за базу для определения базисных абсолютных приростов; y_{i-1} – уровень ряда динамики, выбранный за базу для определения i-го цепного абсолютного прироста.

Существует связь между темпами роста и прироста:

ΔК = К – 1 или ΔК = К – 100% (если темпы роста определены в процентах).

Если разделить абсолютный прирост (цепной) на темп прироста (цепной) за соответствующий период, получим показатель, называемый – абсолютное значение одного процента прироста:

$$A = \frac{\Delta y_{ц}}{\Delta K_{ц}} \tag{11.10}$$

Пример 11.7.

Выпуск продукции предприятия за 1985–1990 гг. характеризуются следующими данными (в сопоставимых ценах), млн. руб.:

Таблица 11.6

Выпуск продукции предприятия за 1985–1990 гг.

1985	1986	1987	1988	1989	1990
23,3	24,9	26,6	27,6	29,0	32,2

Требуется произвести анализ динамики выпуска продукции предприятием за пять лет.

1. Определяем цепные и базисные темпы роста (К).

Цепные:	Базисные:
$К_{1986} = 24,9/23,3 = 1,069$	$К_{1986} = 24,9/23,3 = 1,069$
$К_{1987} = 26,6/24,9 = 1,068$	$К_{1987} = 26,6/23,3 = 1,142$
$К_{1988} = 27,6/26,6 = 1,038$	$К_{1988} = 27,6/23,3 = 1,185$
$К_{1989} = 29,0/27,6 = 1,051$	$К_{1989} = 29,0/23,3 = 1,245$
$К_{1990} = 32,2/29,0 = 1,110$	$К_{1990} = 32,2/23,3 = 1,386$

2. Определяем цепной и базисный абсолютный прирост (Δy).

Цепные:	Базисные:
$\Delta y_{1986} = 24,9 – 23,3 = 1,6$	$\Delta y_{1986} = 24,9 – 23,3 = 1,6$
$\Delta y_{1987} = 26,6 – 24,9 = 1,7$	$\Delta y_{1987} = 26,6 – 23,3 = 3,3$
$\Delta y_{1988} = 27,6 – 26,6 = 1,0$	$\Delta y_{1988} = 27,6 – 23,3 = 4,3$
$\Delta y_{1989} = 29,0 – 27,6 = 1,4$	$\Delta y_{1989} = 29,0 – 23,3 = 5,7$
$\Delta y_{1990} = 32,3 – 29,0 = 3,3$	$\Delta y_{1990} = 32,3 – 23,3 = 9,0$

3. Определяем цепные и базисные темпы прироста (ΔК).

Цепные:	Базисные:
$\Delta K_{1986} = 1{,}6/23{,}3 = 0{,}069$	$\Delta K_{1986} = 1{,}6/23{,}3 = 0{,}069$
$\Delta K_{1987} = 1{,}7/24{,}9 = 0{,}068$	$\Delta K_{1987} = 3{,}3/23{,}3 = 0{,}142$
$\Delta K_{1988} = 1{,}0/26{,}6 = 0{,}038$	$\Delta K_{1988} = 4{,}3/23{,}3 = 0{,}185$
$\Delta K_{1989} = 1{,}4/27{,}6 = 0{,}051$	$\Delta K_{1989} = 5{,}7/23{,}3 = 0{,}245$
$\Delta K_{1990} = 3{,}3/29{,}0 = 0{,}114$	$\Delta K_{1990} = 9{,}0/23{,}3 = 0{,}386$

Проверим связь между темпами роста и прироста.

Цепные темпы прироста:

$\Delta K = K - 1$

$\Delta K_{1986} = K_{1986} - 1$

$\Delta K_{1986} = 1{,}069 - 1 = 0{,}069$

$\Delta K_{1987} = 1{,}068 - 1 = 0{,}068$ и т.д.

Видим, что получаем такие же результаты.

11.5. Определение среднего абсолютного прироста, средних темпов роста и прироста

По показателям изменения уровней ряда динамики (абсолютные приросты, темпы роста и прироста), полученным в результате анализа исходного ряда, могут быть рассчитаны обобщающие показатели в виде средних величин – средний абсолютный прирост, средний темп роста, средний темп прироста.

Средний абсолютный прирост может быть получен по одной из формул:

$$\Delta\overline{y} = \frac{\sum_{i=1}^{n} \Delta y_{iц}}{n} \qquad (11.11)$$

или

$$\Delta\overline{y} = \frac{y_n - y_1}{n-1} \qquad (11.12)$$

где n – число уровней ряда динамики; y_1 – первый уровень ряда динамики; y_n – последний уровень ряда динамики; $\Delta y_{iц}$ – цепные абсолютные приросты.

Средний темп роста можно определить, пользуясь формулами:

$$\overline{K} = \sqrt[n]{K_{1ц} \cdot K_{2ц} \cdot \ldots \cdot K_{nц}} = \sqrt[n]{\prod_{i=1}^{n} K_{iц}} \qquad (11.13)$$

$$\overline{K} = \sqrt[n-1]{\frac{y_n}{y_0}} \tag{11.14}$$

$$\overline{K} = \sqrt[n]{\frac{K_{nб}}{K_{1б}}} \tag{11.15}$$

где n – число рассчитанных цепных или базисных темпов роста; y_0 – уровень ряда, принятый за базу для сравнения; y_n – последний уровень ряда; $K_{iц}$ – цепные темпы роста (в коэффициентах); $K_{iб}$ – первый базисный темп роста; $K_{nб}$ – последний базисный темп роста.

Между темпами прироста ΔK и темпами роста K существует соотношение ΔK = K – 1, аналогичное соотношение верно и для средних величин.

Пример 11.8.

Имеются данные о продаже мебели по региону (в млн. руб.): за 1998 г. – 95,3, за 1999 г. – 103,0, за 2000 г. – 109,5. Определить, как изменилась в среднем ежегодно за весь период (1998–2000 гг.) продажа мебели (в %). Для решения воспользуемся формулой (11.14):

$$\overline{K} = \sqrt[n-1]{\frac{y_n}{y_0}} = \sqrt[3-1]{\frac{y_{2000}}{y_{1998}}} = \sqrt{\frac{109.5}{95.3}} = \sqrt{1.149} = 1.072 \quad \text{или } 107{,}2\%$$

Следовательно, среднегодовой темп продаж мебели по региону составил 7,2%.

Пример 11.9.

Коэффициенты роста продаж швейных изделий по региону на человека каждого последующего года по сравнению с предыдущим составили: 1,065, 1,055, 1,042, 1,036, 1,005. Определить средний темп роста. Для решения воспользуемся формулой (11.13):

$$\overline{K} = \sqrt[n]{\prod_{i=1}^{n} K_{iц}} = \sqrt[5]{1.065 \cdot 1.055 \cdot 1.042 \cdot 1.036 \cdot 1.005} = 1.05 \quad \text{или } 105\%.$$

Средний темп роста – 5%.

11.6. Определение в рядах динамики общей тенденции развития

Определение уровней ряда динамики на протяжении длительного периода времени обусловлено действием ряда факторов, которые неоднородны по силе и направлению воздействия, оказываемого на изучаемое явление.

Рассматривая динамические ряды, пытаются разделить эти факторы на **постоянно действующие** и оказывающие **определяющее воздействие** на уровни ряда, формирующие основную тенденцию развития, и **случайные факторы**, приводящие к кратковременным изменениям уровней ряда динамики. Наиболее важна при анализе ряда динамики его основная тенденция развития, но часто по одному лишь внешнему виду ряда динамики ее установить невозможно, поэтому используют специальные методы обработки, позволяющие показать **основную тенденцию ряда**. Методы обработки используются как простые, так и достаточно сложные. Простейший способ обработки ряда динамики, применяемый с целью установления закономерностей развития – **метод укрупнения интервалов**.

Суть метода в том, чтобы от интервалов, или периодов времени, для которых определены исходные уровни ряда динамики, перейти к более продолжительным периодам времени и посмотреть, как уровни ряда изменяются в этом случае.

Пример 11.10.

Данные о реализации молочной продукции в магазинах города по месяцам представлены таблицей 11.7 (в тоннах)

Исходные уровни ряда динамики подвержены сезонным изменениям. Для определения общей тенденции развития переходят от ежемесячных уровней к годовым уровням:

1987г. – 96,7 тонн

1988г. – 98,4 тонн

1989г. – 101 тонна

Эти цифры, полученные в результате перехода к годовым уровням ряда динамики, показывают общую тенденцию роста реализации молочной продукции.

Таблица 11.7

Данные о реализации молочной продукции в магазинах города

месяц	1987	1988	1989
январь	5,3	5,3	5,4
февраль	5,3	5,1	5,2
март	7,9	8,3	8,2
апрель	8,2	9,0	9,3
май	9,8	9,5	10,1
июнь	12,5	13,0	13,1
июль	11,8	12,2	12,5
август	10,3	10,4	10,8
сентябрь	8,2	8,0	8,3
октябрь	6,5	6,6	6,8
ноябрь	5,4	5,5	5,7
декабрь	5,5	5,5	5,6
итого за год	96,7	98,4	101

Другой способ определения тенденции в ряду динамики – **метод скользящих средних**. Суть метода заключается в том, что фактические уровни ряда заменяются средними уровнями, вычисленными по определённому правилу, например:

$y_1, y_2, y_3, y_4, y_5, ..., y_n$ – исходные или фактические уровни ряда динамики заменяются средними уровнями:

$$y_1 = \frac{y_1 + y_2 + y_3 + y_4 + y_5}{5}$$

$$y_2 = \frac{y_2 + y_3 + y_4 + y_5 + y_6}{5}$$

$$y_3 = \frac{y_3 + y_4 + y_5 + y_6 + y_7}{5}$$

...

$$y_{n-2} = \frac{y_{n-4} + y_{n-3} + y_{n-2} + y_{n-1} + y_n}{5}$$

В результате получается сглаженный ряд, состоящий из скользящих пятизвенных средних уровней y_1, y_2, y_3, y_4, y_5, … , y_{n-1}. Между расположением уровней y_i и $\overline{y}_i$ устанавливается соответствие:

$$y_1, y_2, y_3, y_4, y_5, ..., y_{n-3}, y_{n-2}, y_{n-1}, y_n$$

$$- \; - \; \overline{y}_1, \overline{y}_2, \overline{y}_3, ..., \overline{y}_{n-3}, \overline{y}_{n-2} \; - \; -,$$

сглаженный ряд короче исходного на число уровней $k - 1$, где k – число уровней, выбранных для определения средних уровней ряда.

Сглаживание методом скользящих средних можно производить по четырём, пяти или другому числу уровней ряда, используя соответствующие формулы для усреднения исходных уровней.

Полученные при этом средние уровни называются четырехзвенными скользящими средними, пятизвенными скользящими средними и т.д.

При сглаживании ряда динамики по четному числу уровней выполняется дополнительная операция, называемая **центрированием**, поскольку, при вычислении скользящего среднего, например по четырем уровням, $\overline{y}_1 = \dfrac{y_1 + y_2 + y_3 + y_4}{4}$ относится к временной точке между моментами времени, когда были зафиксированы фактические уровни y_2 и y_3. Схема вычислений и расположений уровней сглаженного ряда становится сложнее:

$$y_1, y_2, y_3, y_4, y_5, y_6 \cdots$$

$$- \; - \; \overline{y}_1, \overline{y}_2, \overline{y}_3 \cdots$$

$$- \; - \; \overline{y}_{1ц}, \overline{y}_{2ц} \cdots$$

$$\overline{y}_{1ц} = \frac{\overline{y}_1 + \overline{y}_2}{2}, \; \overline{y}_{2ц} = \frac{\overline{y}_2 + \overline{y}_3}{2}.$$

Метод скользящих средних не позволяет получить численные оценки для выражения основной тенденции в ряду динамики, давая лишь наглядное графическое представление.

Пример 11.11.

Таблица 11.8

Данные о валовом сборе хлопка-сырца в стране

Годы	Валовый сбор хлопка-сырца, млн. т.	Скользящая средняя по 5 уровням
1960	4,3	–
1961	4,5	–
1962	4,3	4,72
1963	5,2	5,00
1964	5,3	5,30
1965	5,7	5,64
1966	6,0	5,78
1967	6,0	5,86
1968	5,9	6,10
1969	5,7	6,32
1970	6,9	6,58
1971	7,1	6,94
1972	7,3	7,48
1973	7,7	7,68
1974	8,4	7,92
1975	7,9	8,22
1976	8,3	8,38
1977	8,8	8,54
1978	8,5	8,94
1979	9,2	9,18
1980	9,9	9,30
1981	9,6	–
1982	9,3	–

На рис. 11.1 показан график, построенный по данным о валовом сборе хлопка-сырца в стране за ряд лет наблюдения и по расчетным данным, представленным в таблице 11.8.

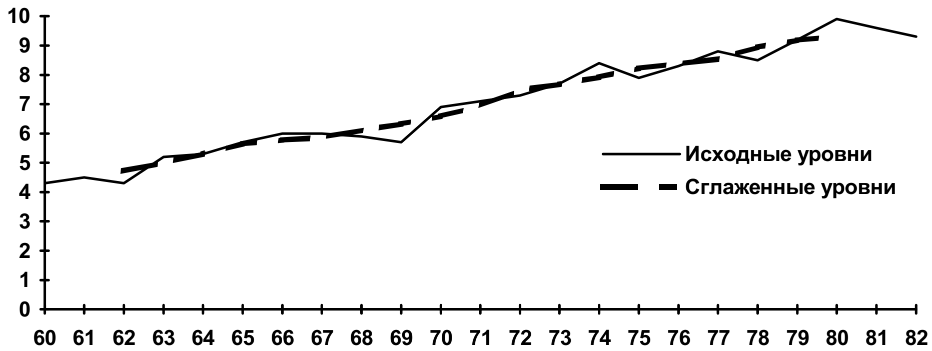


Рис. 11.1. Валовой сбор хлопка-сырца

Наиболее совершенным способом определения тенденции развития в ряду динамики является **метод аналитического выравнивания**. При этом методе исходные уровни ряда динамики y_i заменяются теоретическими или расчетными y_i', которые представляют из себя некоторую достаточно простую математическую функцию времени, выражающую общую тенденцию развития ряда динамики. Чаще всего в качестве такой функции выбирают **прямую, параболу, экспоненту** и др.

Например,

$$y_i' = a_0 + a_1 t_i$$

где $\mathbf{a_0}$, $\mathbf{a_1}$ – коэффициенты, определяемые в методе аналитического выравнивания; $\mathbf{t_i}$ – моменты времени, для которых были получены исходные и соответствующие теоретические уровни ряда динамики, образующие прямую, определяемую коэффициентами $\mathbf{a_0}$, $\mathbf{a_1}$.

Расчет коэффициентов $\mathbf{a_0}$, $\mathbf{a_1}$ ведется на основе метода наименьших квадратов (МНК):

$$\sum_{i=1}^{n}(y_i' - y_i)^2 \to \min$$

Если вместо y_i' подставить $a_0 + a_1 t_i$ (или соответствующее выражение для других математических функций), получим:

$$f(a_0, a_1) = \sum_{i=1}^{n}(a_0 + a_1 t_i - y_i)^2 \to \min$$

Эта функция двух переменных a_0, a_1 (все t_i и y_i известны), которая при определенных $\mathbf{a_0}$, $\mathbf{a_1}$ достигает минимума. Из этого выражения на основе понятий об экстремумах функций **n** переменных (частные производные функции приравниваются к нулю), получаем:

$$\frac{df}{da_0} = 2\sum_{i=1}^{n}(a_0 + a_1 t_i - y_i) = 0$$

$$\frac{df}{da_1} = 2\sum_{i=1}^{n}(a_0 + a_1 t_i - y_i)\cdot t_i = 0$$

Отсюда получаем систему нормальных уравнений МНК для прямой:

$$\begin{cases} na_0 + a_1 \sum\limits_{i=1}^{n} t_i = \sum\limits_{i=1}^{n} y_i \\ a_0 \sum\limits_{i=1}^{n} t_i + a_1 \sum\limits_{i=1}^{n} (x_i)^2 = \sum\limits_{i=1}^{n} t_i y_i \end{cases}$$

откуда значения коэффициентов $\mathbf{a_0}$, $\mathbf{a_1}$ будут равны:

$$a_0 = \frac{\sum\limits_{i=1}^{n} y_i \sum\limits_{i=1}^{n} t_i^2 - \sum\limits_{i=1}^{n} t_i y_i \sum\limits_{i=1}^{n} t_i}{n \sum\limits_{i=1}^{n} t_i^2 - \sum\limits_{i=1}^{n} t_i \sum\limits_{i=1}^{n} t_i}$$

$$a_1 = \frac{n \sum\limits_{i=1}^{n} y_i t_i - \sum\limits_{i=1}^{n} t_i \sum\limits_{i=1}^{n} y_i}{n \sum\limits_{i=1}^{n} t_i^2 - \sum\limits_{i=1}^{n} t_i \sum\limits_{i=1}^{n} t_i}$$

где n – число моментов времени, для которых были получены исходные уровни ряда y_i.

Если вместо абсолютного времени t_i выбрать условное время таким образом, чтобы $\sum t_i = 0$, то записанные выражения для определения a_0, a_1 упрощаются:

$$a_0 = \frac{\sum\limits_{i=1}^{n} y_i}{n} \qquad a_1 = \frac{\sum\limits_{i=1}^{n} t_i y_i}{\sum\limits_{i=1}^{n} t_i^2}$$

Пример 11.12.

Нечетное число уровня ряда.

1981	1982	1983	1984	1985	1986	1987	абсолютное время
–3	–2	–1	0	1	2	3	условное время

Четное число уровней ряда.

1981	1982	1983	1984	1985	1986	1987	1988	абсолютное время
–7	–5	–3	–1	1	3	5	7	условное время

В обоих случаях $\sum t_i = 0$.

Пример 11.13.

Выполняется аналитическое выравнивание ряда, отражающего производство стали в стране по годам (млн. т).

1985	1986	1987	1988	1989
141,3	144,8	146,7	151,5	149,0

В качестве математической функции, отражающей тенденцию развития, выбираем прямую $y'_i = a_0 + a_1 t_i$, определение a_0, a_1 производится для условного времени, в результате $a_0 = 146{,}66$, $a_1 = 2{,}21$.

Таблица 11.9

Расчетные значения производства стали

Год	Производство стали y_i	Условное время t_i	Теоретические уровни $y'_i = 146{,}66 + 2{,}21t_i$
1985	141,3	–2	142,2
1986	144,8	–1	144,4
1987	146,7	0	146,7
1988	151,5	1	148,9
1989	149,0	2	151,1

11.7. Определение в рядах внутригодовой (сезонной) динамики

Многие процессы хозяйственной деятельности, торговли, сельского хозяйства и других сфер человеческой деятельности подвержены сезонным изменениям, например, продажа мороженого, потребление электроэнергии, производство молока, сахара, продажа сельхозпродукции и др.

Для анализа рядов динамики, подверженных сезонным изменениям, используются специальные методы, позволяющие установить и описать особенности изменения уровней ряда. Прежде, чем использовать методы изучения сезонности, необходимо подготовить данные, приведенные в сопоставимый вид, за несколько лет наблюдения по месяцам или кварталам. Изменения сезонных колебаний производится с помощью **индексов сезонности**. В зависимости от существующих в ряду динамики тенденций используются различные правила построения индексов.

1. Ряд динамики не имеет общей тенденции развития, либо она не велика.

Индекс сезонности:

$$I_{s_i} = \frac{\overline{y}_i}{\overline{y}}$$

где $\overline{y}_i$ – средний уровень ряда, полученный в результате осреднения уровней ряда за одноименные периоды времени (например, средний уровень января за все годы наблюдения); $\overline{y}$ – общий средний уровень ряда за все время наблюдения.

Вывод о наличии или отсутствия в ряду динамики ярко выраженной тенденции может производиться, например, при помощи метода укрупнения интервалов.

Пример 11.14.

Имеются данные заключения брака в городе за ряд лет наблюдения:

Таблица 11.10

Данные о заключении брака в городе

Месяц	1986	1987	1988
январь	173	183	178
февраль	184	185	179
март	167	162	161
апрель	142	160	184
май	137	143	151
июнь	145	150	156
июль	153	167	177
август	171	173	181
сентябрь	143	150	157
октябрь	162	165	174
ноябрь	178	181	193
декабрь	185	189	197
Итого за год	1940	2008	2088

При переходе от месячных к годовым уровням можно установить, что тенденция роста очень незначительна.

Общий средний уровень ряда:

$\overline{y} = \dfrac{1940 + 2008 + 2088}{3 \cdot 365} = 5{,}51$ – среднее число браков, заключаемых за один день.

Средний уровень января:

$\overline{y}_1 = \dfrac{y_{1_{1986}} + y_{1_{1987}} + y_{1_{1989}}}{3 \cdot 31} = \dfrac{173 + 183 + 178}{93} = 5{,}74$ – среднее число браков за один день января.

Аналогично рассчитывается средние уровни февраля, марта и т.д. Результаты расчётов сведены в таблицу 11.11:

Таблица 11.11

Сводная таблица результатов расчета

Месяц	$\overline{y_i}$	$I_{s_i} \cdot 100\%$
январь	5,74	104,2
февраль	6,45	117,1
март	5,27	95,6
апрель	5,4	88,0
май	4,63	84,0
июнь	5,01	91,0
июль	5,34	96,9
август	5,64	102,4
сентябрь	5,0	90,7
октябрь	5,39	97,8
ноябрь	6,13	111,3
декабрь	6,14	111,4

Полученные индексы сезонности дают оценку того, как в отдельные месяцы года количество заключенных браков отклоняется от среднего значения. Построенный по полученным индексам сезонности линейный график наглядно покажет сезонность рассматриваемого процесса.

2. Ряд динамики имеет общую тенденцию, и она определена либо методом скользящего среднего, либо методом аналитического выравнивания.

Индекс сезонности, $$\bar{I}_{s_i} = \frac{1}{n}\left(\sum_{i=1}^{n} \frac{y_i}{y_i'}\right) \qquad (11.16)$$

где y_i – исходные уровни ряда; y_i' – уровни ряда, полученные в результате определения скользящих средних для тех же периодов времени, что и исходные уровни; I – номер месяца или квартала, для которого определяется индекс сезонности; n – число лет наблюдения за процессом.

В случае, если тенденция развития определялась методом аналитического выравнивания, расчетная формула получения индексов сезонности совершенно аналогична предыдущей, но вместо y_i' – уровней, полученных методом скользящих средних, используются y_i' – полученные методом аналитического выравнивания.

Пример 11.15.

На основе исходных данных о реализации сахара в продовольственных магазинах города в 1990–1992 гг. (т) (табл. 11.12), определены скользящие средние по трем уровням ряда:

Таблица 11.12

Данные о реализации сахара
в продовольственных магазинах города в 1990–1992 гг. (т)

Месяц	1990		1991		1992	
	Исходные уровни y_i	Сглажен. уровни y'_i	Исходные уровни y_i	Сглажен. уровни y'_i	Исходные уровни y_i	Сглажен. уровни y'_i
январь	78,9	–	108,6	106,2	129,1	131,3
февраль	78,1	81,0	107,9	107,8	128,6	129,5
март	86,0	87,2	106,8	115,4	130,7	137,4
апрель	97,5	88,9	132,1	117,3	152,8	141,1
май	83,3	88,9	113,0	119,0	139,8	146,7
июнь	86,0	86,6	111,8	116,4	147,4	150,3
июль	90,6	87,6	124,4	116,8	163,8	152,5
август	86,1	86,0	114,1	115,6	146,3	149,3
сентябрь	81,3	90,8	108,4	115,6	137,8	145,4
октябрь	105,1	94,5	124,0	117,0	152,2	144,4
ноябрь	97,2	101,5	118,0	126,2	143,2	150,6
декабрь	102,1	102,6	136,3	128,0	156,5	–

На основе исходных и сглаженных уровней ряда строятся индексы сезонности (формула (11.16)):

Так для января:

$$\bar{I}_{s_i} = [\frac{y_{1_{1991}}}{y'_{1_{1991}}} + \frac{y_{1_{1992}}}{y'_{1_{1992}}}] : 2 = [\frac{108,6}{106,2} + \frac{129,1}{131,3}] : 2 = 1,0$$

Для февраля:

$$\bar{I}_{s_i} = [\frac{y_{2_{1990}}}{y'_{2_{1990}}} + \frac{y_{2_{1991}}}{y'_{2_{1991}}} + \frac{y_{2_{1992}}}{y'_{2_{1992}}}] : 3 = [\frac{78,1}{81,0} + \frac{107,9}{107,8} + \frac{128,6}{129,5}] : 3 = 0,98$$

и т.д.

Индексы сезонности по месяцам сведены в таблицу:

Месяц	1	2	3	4	5	6	7	8	9	10	11	12
	100	98	96	110	95	98	106	96	93	107	95	103

Построив линейный график, можно увидеть закономерности изменения объема продаж сахара по месяцам года.

Контрольные вопросы.

1. Понятие о рядах динамики в статистике.
2. Установление вида ряда динамики.
3. Приведение рядов динамики в сопоставимый вид.
4. Определение среднего уровня ряда динамики.
5. Показатели изменения уровня ряда динамики. Темпы роста, абсолютный прирост и темпы прироста.
6. Определение среднего абсолютного прироста, средних темпов роста и прироста.
7. Определение в рядах динамики общей тенденции развития.
8. Определение в рядах сезонной динамики.

ГЛАВА 12. ИНДЕКСЫ

12.1 Индексный метод. Статистические индексы

Важное значение в статистических исследованиях коммерческой деятельности имеет индексный метод. Полученные на основе этого метода показатели используются для характеристики развития анализируемых показателей во времени, по территории, изучения структуры и взаимосвязей, выявления роли факторов в изменении сложных явлений.

Индексы широко применяются в экономических разработках государственной и ведомственной статистики.

Статистический индекс – это относительная величина сравнения сложных совокупностей и отдельных их единиц. При этом под сложной понимается такая статистическая совокупность, отдельные элементы которой непосредственно не подлежат суммированию.

Например, ассортимент продовольственных товаров состоит из товарных разновидностей, первичный учет которых на производстве и в оптовой торговле ведется в натуральных единицах измерения: молоко – в литрах, мясо – в центнерах, яйцо – в штуках, консервы – в условных банках и т.д. Для определения общего объема производства и реализации продовольственных товаров суммировать данные учета разнородных товарных масс в натуральных измерителях нельзя. Не подлежат непосредственному суммированию и данные о количестве произведенных и реализованных различных видов непродовольственных товаров. Было бы, например, бессмысленно для получения общего объема реализации суммировать данные о продаже тканей (в метрах), костюмов (в штуках), обуви (в парах) и т.д.

В этих сложных статистических совокупностях единицами наблюдения являются товары с различными потребительскими свойствами. Данные о натурально-вещественной форме реализации отдельных товарных разновидностей непосредственному суммированию не подлежат. Для получения в сложных статистических совокупностях обобщающих (суммарных) величин прибегают к индексному методу.

Основой индексного метода при определении изменений в производстве и обращении товаров является переход от натурально – вещественной формы выражения товарных масс к стоимостным (денежным) измерителям. Именно посредством денежного выражения стоимости отдельных товаров устраняется их несравнимость как потребительских стоимостей и достигается единство.

12.2. Индивидуальные и общие индексы

В зависимости от степени охвата подвергнутых обобщению единиц изучаемой совокупности индексы подразделяются на индивидуальные (элементарные) и общие.

Индивидуальные индексы характеризуют изменения отдельных единиц статистической совокупности. Так, например, если при изучении оптовой реализации продовольственных товаров определяются изменения в продаже отдельных товарных разновидностей, то получают индивидуальные (однотоварные) индексы.

Общие индексы выражают сводные (обобщающие) результаты совместного изменения всех единиц, образующих статистическую совокупность. Пример, показатель изменения объема реализации товарной массы продуктов питания по отдельным периодам будет общим индексом физического объема товарооборота.

Важной особенностью общих индексов является то, что они обладают синтетическими и аналитическими свойствами.

Синтетические свойства индексов состоят в том, что посредством индексного метода производится соединение (агрегирование) в целом разнородных единиц статистической совокупности.

Аналитические свойства индексов состоят в том, что посредством индексного метода определяется влияние факторов на изменение изучаемого показателя.

Для определения индекса надо произвести сопоставление не менее двух величин. При изучении динамики социально-экономических явлений сравниваемая величина (числитель индексного отношения) принимается за текущий (или отчетный) период, а величина, с которой производится сравнение – за базисный период.

Основным элементом индексного отношения является индексируемая величина. Под индексируемой величиной понимается значение признака статистической совокупности, изменение которой является объектом изучения. Так, при изучении изменения цен индексируемой величиной является цена единицы товара p. При изучении изменения физического объема товарной массы в качестве индексируемой величины выступают данные о количестве товаров в натуральных измерителях q. Стоимость продукции обозначается через s.

Индивидуальные индексы принято обозначать v, а общие индексы – I.

Знак внизу справа означает период:

$_0$ – базисный,

$_1$ – отчетный.

Пример 12.1.

В текущем, отчётном году предприятие произвело 120 тыс. т. продукции вместо 100 тыс. т. в прошлом базисном, году. Цены за каждую тонну этой продукции снизились с 20 до 18 рублей; а ее общая стоимость возросла с 2 000 до 2 160 тыс. руб.

В данном примере можно вычислить три индекса:

индекс объёма продукции: $v_q = \frac{q_1}{q_0} = \frac{120000}{100000} = 1{,}2$ или 120%;

индекс цен: $v_p = \frac{p_1}{p_0} = \frac{18}{20} = 0{,}9$ или 90%;

индекс стоимости продукции: $v_s = \frac{s_1}{s_0} = \frac{2160000}{2000000} = 1{,}08$ или 108%.

Полученные индексы показывают, что объём продукции и ее стоимость возросла в отчетном году по сравнению с базисным в 1,2 и 1,08 раза, а цены, наоборот, снизились до 1,9 их базисного уровня. Все три индекса образуют систему показателей-сомножителей: $\nu_q \cdot \nu_p = \nu_s$ или $1{,}2 \times 0{,}9 = 1{,}08$.

12.3. Агрегатные индексы

Основной формой общих индексов являются **агрегатные индексы**.

Достижение в сложных статистических совокупностях сопоставимости разнородных единиц осуществляется введением в индексные отношения специальных сомножителей индексируемых величин. Такие сомножители называются **соизмерителями**. Они необходимы для перехода от натуральных измерителей разнородных единиц статистической совокупности к однородным показателям. При этом в числителе и знаменателе общего индекса изменяется лишь значение индексируемой величины, а их соизмерители являются постоянными величинами.

В качестве соизмерителей индексируемых величин выступают тесно связанные с ними экономические показатели: цены, количество и др.

Произведение каждой индексируемой величины на соизмеритель образует в индексном отношении определенные экономические категории.

Рассмотрим агрегатный индекс **цен**.

Пример 12.2.

Таблица 12.1

Данные о реализация товаров

Товар	Ед. изм.	I период		II период		Индивидуальные индексы	
		цена за ед. товара, руб. p_{0i}	кол-во q_{0i}	цена за ед. товара, руб. p_{1i}	кол-во q_{1i}	цен $i_p = \frac{p_{1i}}{p_{0i}}$	физич. объёма $i_q = \frac{q_{1i}}{q_{0i}}$
А	т	20	7 500	25	9 500	1,25	1,27
Б	м	30	2 000	30	2 500	1,0	1,25
В	шт.	15	1 000	10	1 500	0,67	1,5

При определении по данным таблицы статистических индексов первый период принимается за базисный, в котором цена единицы товара принимается p_{0i}, а количество – q_{0i}.

Второй период принимается за текущий (или отчетный), в котором цена единицы товара обозначается p_{1i}, а количество – q_{1i}.

Индивидуальные индексы показывают, что в текущем периоде по сравнению с базисным цена на товар А повысилась на 25%, на товар Б осталась без изменения, а на товар В снизилась на 33%. Количество реализации товара А возросло на 27%, товара Б – на 25%, а товара В – на 50%.

При определении общего индекса цен в агрегатной форме I_p в качестве соизмерителя индексируемых величин p_{1i} и p_{0i} могут приниматься данные о количестве реализации товаров в текущем периоде q_{1i}. При умножении q_{1i} на индексируемые величины в числителе индексного отношения образуется значение $\sum_{i=1}^{n} q_{1i} p_{1i}$, сумма стоимости продажи товаров в текущем периоде по ценам того же текущего периода. В знаменателе индексного отношения образуется значение $\sum_{i=1}^{n} q_{1i} p_{0i}$, т.е. сумма стоимости продажи товаров в текущем периоде по ценам базисного периода (n – число видов товаров).

Агрегатная формула такого общего индекса цен имеет следующий вид

$$I_p = \frac{\sum_{i=1}^{n} q_{1i} p_{1i}}{\sum_{i=1}^{n} q_{1i} p_{0i}} \tag{12.1}$$

Расчет агрегатного индекса цен по данной формуле предложил немецкий экономист Г. Пааше, поэтому он называется индексом Пааше.

Применяем формулу для расчета агрегатного индекса цен по данным табл.12.1:

числитель индексного отношения

$$\sum_{i=1}^{n} q_{1i} p_{1i} = 25 \times 9\,500 + 30 \times 2\,500 + 10 \times 1\,500 = 327\,500 \text{ руб.}$$

знаменатель индексного отношения

$$\sum_{i=1}^{n} q_{1i} p_{0i} = 20 \times 9\,500 + 30 \times 2\,500 + 15 \times 1\,500 = 287\,500 \text{ руб.}$$

Полученные значения подставляем в формулу (12.1):

$$I_p = \frac{327500}{287500} = 1{,}139 \text{ или } 113{,}9\%$$

Применение формулы 1 показывает, что по данному ассортименту товаров в целом цены повысились в среднем на 13,9%.

При другом способе определения агрегатного индекса цен в качестве соизмерителя индексируемых величин p_{1i} и p_{0i} могут применяться данные о количестве реализации товаров в базисном периоде q_{0i}. При этом умножение q_{0i} на индексируемые величины в числителе индексного отношения образует значение $\sum_{i=1}^{n} q_{0i} p_{1i}$, т.е. сумму стоимости продажи товаров в базисном периоде по ценам текущего периода.

В знаменателе индексного отношения образуется значение $\sum_{i=1}^{n} q_{0i} p_{0i}$, т.е. сумма стоимости продажи товаров в базисном периоде по ценам того же базисного периода.

Агрегатная формула такого общего индекса имеет вид:

$$I_p = \frac{\sum_{i=1}^{n} q_{0i} p_{1i}}{\sum_{i=1}^{n} q_{0i} p_{0i}} \qquad (12.2)$$

Расчет общего индекса цен по данной формуле предложил немецкий экономист Э. Ласпейрес, и получил название индекса Ласпейреса.

Применяем формулу для расчета агрегатного индекса цен по данным табл.12.1:

числитель индексного отношения

$$\sum_{i=1}^{n} q_{0i} p_{1i} = 25 \times 7\,500 + 30 \times 2\,000 + 10 \times 1000 = 257\,500 \text{ руб.}$$

знаменатель индексного отношения

$$\sum_{i=1}^{n} q_{0i} p_{0i} = 20 \times 7\,500 + 30 \times 2\,000 + 15 \times 1\,000 = 225\,000 \text{ руб.}$$

Полученные значения подставляем в формулу 2:

$$I_p = \frac{257500}{225000} = 1{,}144 \text{ или } 114{,}4\%.$$

Применение формулы (12.2) показывает, что по данному ассортименту товаров в целом цены повысились в среднем на 14,4%.

Таким образом, выполненные по формулам (12.1) и (12.2) расчеты имеют разные показания индексов цен. Это объясняется тем, что индексы Пааше и Ласпейреса характеризуют различные качественные особенности изменения цен.

Индекс Пааше характеризует влияние изменения цен на стоимость товаров, реализованных в отчетном периоде. Индекс Ласпейреса показывает влияние изменения цен на стоимость количества товаров, реализованных в базисном периоде.

Другим важным видом общих индексов, которые широко применяются в статистике, являются агрегатные индексы **физического объема товарной массы.**

При определении агрегатного индекса физического объема товарной массы I_q в качестве соизмерителей индексируемых величин q_{1i} и q_{0i} могут применяться неизменные цены базисного периода p_{0i}. При умножении p_0 на индексируемые величины в числителе индексного отношения образуются значение $\sum_{i=1}^{n} q_{1i} p_{0i}$, т.е. сумма стоимости товарной массы текущего периода в базисных ценах. В знаменателе – $\sum_{i=1}^{n} q_{0i} p_{0i}$, то есть сумма стоимости товарной массы базисного периода в ценах того же базисного периода.

Агрегатная форма общего индекса имеет следующий вид:

$$I_q = \frac{\sum_{i=1}^{n} q_{1i} p_{0i}}{\sum_{i=1}^{n} q_{0i} p_{0i}} \qquad (12.3)$$

Поскольку, в числителе формулы (12.3) содержится сумма стоимости реализации товаров в текущем периоде по неизменным (базисным) ценам, а в знаменателе – сумма фактической стоимости товаров, реализованных в базисном периоде в тех же неизменных (базисных) ценах, то данный индекс является агрегатным индексом товарооборота в сопоставимых (базисных) ценах.

Используем формулу (12.3) для расчёта агрегатного индекса физического объёма реализации товаров по данным табл. 1:

числитель индексного отношения

$$\sum_{i=1}^{n} q_{1i} p_{0i} = 9\,500 \times 20 + 2\,500 \times 30 + 1\,500 \times 15 = 287\,500 \text{ руб.}$$

знаменатель индексного отношения

$$\sum_{i=1}^{n} q_{1i} p_{0i} = 7\,500 \times 20 + 2\,000 \times 30 + 1\,000 \times 15 = 225\,000 \text{ руб.}$$

Полученные значения подставляем в формулу (12.3):

$$I_p = \frac{287500}{225000} = 1{,}278 \text{ или } 127{,}8\%.$$

Применение формулы (12.3) показывает, что по данному ассортименту товаров в целом прирост физического объёма реализации в текущем периоде составил в среднем 27,8%.

Агрегатный индекс физического объёма товарооборота может определяться посредством использования в качестве соизмерителя индексируемых величин q_{1i} и q_{0i} цен текущего периода p_{1i}.

Агрегатная формула общего индекса будет иметь вид:

$$I_q = \frac{\sum_{i=1}^{n} q_{1i} p_{1i}}{\sum_{i=1}^{n} q_{0i} p_{1i}} \qquad (12.4)$$

числитель индексного отношения

$$\sum_{i=1}^{n} q_{1i} p_{1i} = 9\,500 \times 25 + 2\,500 \times 30 + 1\,500 \times 10 = 327\,500 \text{ руб.}$$

знаменатель индексного отношения

$$\sum_{i=1}^{n} q_{0i} p_{1i} = 7\,500 \times 25 + 2\,000 \times 30 + 1\,000 \times 10 = 257\,500 \text{ руб.}$$

Полученные значения подставляем в формулу (12.4):

$$I_p = \frac{327500}{257500} = 1{,}272 \text{ или } 127{,}2\%.$$

Применение формулы (12.4) показывает, что по данному ассортименту товаров в целом прирост физического объёма реализации в текущем периоде составил в среднем 27,2%.

Аналогичным образом производится расчёт индекса себестоимости, при этом сравниваются суммы затрат в производстве в отчётном периоде ($\sum\limits_{i=1}^{n} q_{1i} z_{1i}$ – числитель индекса) с суммой затрат в производстве на продукцию отчётного периода по себестоимости базисного периода ($\sum\limits_{i=1}^{n} q_{1i} z_{0i}$ – знаменатель) (z – затраты).

12.4. Индексы с постоянными и переменными весами

При изучении динамики коммерческой деятельности приходится производить индексные сопоставления более чем за два периода.

Поэтому индексные величины могут определяться как на постоянной, так и на переменной базах сравнения. При этом, если задача анализа состоит в получении характеристик изменения изучаемого явления во всех последующих периодах по сравнению с начальным, то вычисляются базисные индексы. Например, сопоставление объёма розничного товарооборота II, III и IV кварталов с I кварталом.

Но если требуется охарактеризовать последовательно изменения изучаемого явления из периода в период, то вычисляются цепные индексы. Например, при изучении объёма розничного товарооборота по кварталам года сопоставляют товарооборот II квартала с I, III – со II и IV – с III кварталом.

В зависимости от задачи исследования и характера исходной информации базисные и цепные индексы исчисляются как индивидуальные, так и общие.

Способы расчета индивидуальных базисных и цепных индексов аналогичны расчету относительных величин динамики. Общие индексы в зависимости от их вида вычисляются с переменными и постоянными весами – соизмерителями.

Используя индексный ряд за несколько периодов, можно получить динамику стоимости продукции и динамику товарооборота в неизменных ценах, т.е. в ценах какого-то одного прошлого периода. Такие индексные ряды называются индексами с постоянными весами. Для них действует правило: произведение цепных индексов дает индекс базисный.

Пример12.3.

По заводу имеются данные об объеме производства и стоимости продукции (табл. 12.2).

Требуется рассчитать индексы физического объёма продукции с постоянными весами.

Индексы с постоянной базой (базисные):

Таблица 12.2

Данные по заводу об объеме производства и стоимости продукции

Вид. прод.	Ед. изм.	Произведено продукции			Цены в 1985 г., тыс. руб.	Стоимость продукции в неизменных ценах 1985 г., тыс. руб.		
		1988	1989	1990		1988	1989	1990
А	тыс. т.	60	64	69	5 000	300	320	345
Б	млн. шт.	5,5	6,2	7,0	2 000	11 000	12 400	14 000
	всего	–	–	–	–	11 300	12 720	14 345

$$I_{\frac{1989}{1988}} = \frac{12720}{11300} = 1{,}126 \qquad I_{\frac{1990}{1988}} = \frac{14350}{11300} = 1{,}27$$

Индексы с переменной базой (цепные):

$$I_{\frac{1989}{1988}} = \frac{12720}{11300} = 1{,}126 \qquad I_{\frac{1990}{1989}} = \frac{14350}{12720} = 1{,}128$$

Убедимся, что произведение цепных индексов равно базисному:

$$1{,}126 \times 1{,}128 = 1{,}27$$

Если индексы цен, себестоимости и производительности труда имеют в качестве весов количество продукции отчётного периода, то эти индексы образуют индексные ряды с переменными весами, поскольку в каждом отдельном индексе отчётный период изменяется. Индексы с переменными весами не подчиняются правилу, согласно которому произведение цепных индексов равно базисному.

Пример 12.4.

Имеются данные об объёме производства и себестоимости продукции:

Таблица 12.3

Данные об объеме производства и себестоимости продукции

Вид продукции	Единица измерения	Выработано продукции за квартал			Себестоимость единицы продукции в квартал, руб.		
		I	II	III	I	II	III
А	шт.	100	120	150	10	9,9	9,6
Б	шт.	300	310	320	35	35	34
В	кг.	7 800	8 200	8 500	0,5	0,48	0,45

Рассчитать индексы себестоимости с переменными весами.

$$I_{\frac{IIKB}{IKB}} = \frac{9.9 \cdot 120 + 35 \cdot 310 + 0.48 \cdot 8200}{10 \cdot 120 + 35 \cdot 310 + 0.5 \cdot 8200} = 0.989$$

$$I_{\frac{IIIKB}{IIKB}} = \frac{9.6 \cdot 150 + 34 \cdot 320 + 0.45 \cdot 8500}{9.9 \cdot 150 + 35 \cdot 320 + 0.48 \cdot 8500} = 0.963$$

Перемножив цепные индексы, получим:

0,989 · 0,963 = 0,9524

Рассчитаем базисный индекс III квартала:

$$I_{\frac{IIIKB}{IKB}} = \frac{9.6 \cdot 150 + 34 \cdot 320 + 0.45 \cdot 8500}{10 \cdot 150 + 35 \cdot 320 + 0.5 \cdot 8500} = 0.9525$$

Как видим, расхождение есть, но оно проявляется только в четвёртом знаке после запятой. Величина расхождения не многим более 0,01%.

12.5. Средние индексы

Всякий агрегатный индекс может быть преобразован в средний арифметический из индивидуальных индексов. Для этого индексируемая величина отчётного периода, стоящая в числителе агрегатного индекса, заменяется произведением индивидуального индекса на индексируемую величину базисного периода.

Так, индивидуальный индекс цен равен $v_i = \dfrac{p_{1i}}{p_{0i}}$, откуда $p_{1i} = v_i \cdot p_{0i}$

Следовательно, преобразование агрегатного индекса цен в средний арифметический имеет вид:

$$I_p = \frac{\sum_{i=1}^{n} q_{1i} p_{1i}}{\sum_{i=1}^{n} q_{1i} p_{0i}} = \frac{\sum_{i=1}^{n} q_{1i} p_{0i} v_i}{\sum_{i=1}^{n} q_{1i} p_{0i}}$$

Аналогично индекс себестоимости равен $v_i = \frac{z_{1i}}{z_{0i}}$, откуда $z_{1i} = v_i\, z_{0i}$, следовательно,

$$I_z = \frac{\sum_{i=1}^{n} q_{1i} z_{1i}}{\sum_{i=1}^{n} q_{1i} z_{0i}} = \frac{\sum_{i=1}^{n} q_{1i} z_{0i} v_i}{\sum_{i=1}^{n} q_{1i} z_{0i}}$$

Аналогично индекс физического объёма продукции (товарооборота) равен $v_i = \frac{q_{1i}}{q_{0i}}$, откуда $q_{1i} = v_i\, q_{0i}$, следовательно,

$$I_p = \frac{\sum_{i=1}^{n} q_{1i} p_{0i}}{\sum_{i=1}^{n} q_{0i} p_{0i}} = \frac{\sum_{i=1}^{n} q_{0i} p_{0i} v_i}{\sum_{i=1}^{n} q_{0i} p_{0i}}$$

Пример 12.5.

Определить средний арифметический индекс физического объема продукции.

Таблица 12.4

Средний арифметический индекс физического объема продукции

Отрасль производства	Стоимость прод. в базисном году, млн. руб.	Индексы физич. объёма прод. в отчёт. году (базис. год = 1)
Сахарная	20	1,47
Мукомольная	30	1,55
Мясная	25	1,71
Рыбная	15	2,1
ИТОГО	90	–

$$I_p = \frac{\sum_{i=1}^{n} q_{0i} p_{0i} \cdot v_i}{\sum_{i=1}^{n} q_{0i} p_{0i}} = \frac{1.47 \cdot 20 + 1.55 \cdot 30 + 1.71 \cdot 25 + 2.1 \cdot 15}{20 + 30 + 25 + 15} = 1.667 \text{ или } 166{,}7\%$$

Физический объём продукции 4 отраслей увеличился на 66,7%.

12.6. Расчеты недостающих индексов с помощью индексных систем

Многие экономические индексы тесно связаны между собой и образуют индексные системы. Так, индекс цен связан с индексом физического объема товарооборота или физического объема продукции, образуя следующую индексную систему:

$$\frac{\sum_{i=1}^{n} p_{1i}q_{1i}}{\sum_{i=1}^{n} p_{0i}q_{1i}} \times \frac{\sum_{i=1}^{n} q_{1i}p_{0i}}{\sum_{i=1}^{n} q_{0i}p_{0i}} = \frac{\sum_{i=1}^{n} p_{1i}q_{1i}}{\sum_{i=1}^{n} p_{0i}q_{0i}} \text{ или } I_p \cdot I_q = I_{pq}$$

Произведение индекса цен на индекс физического объема товарооборота или продукции дает индекс физического объема товарооборота в фактических ценах, или индекс стоимости продукции.

Индекс себестоимости промышленной продукции связан с индексом физического объема продукции по себестоимости, образуя следующую индексную систему:

$$\frac{\sum_{i=1}^{n} z_{1i}q_{1i}}{\sum_{i=1}^{n} z_{0i}q_{1i}} \times \frac{\sum_{i=1}^{n} q_{1i}z_{0i}}{\sum_{i=1}^{n} q_{0i}z_{0i}} = \frac{\sum_{i=1}^{n} z_{1i}q_{1i}}{\sum_{i=1}^{n} z_{0i}q_{0i}} \text{ или } I_z \cdot I_q = I_{zq}$$

Произведение индекса себестоимости продукции на индекс физического объема дает индекс затрат в производстве.

Используя индексы системы, можно по двум известным индексам найти третий, неизвестный.

Пример 12.6.

Имеются следующие данные о продаже товаров в магазинах А:

Таблица 12.5

Данные о продаже товаров в магазинах А

Товар	Продано, кг		Цена 1 кг, руб.	
	базисный период (q_{0i})	отчетный период (q_{0i})	базисный период (p_{0i})	отчетный период (p_{0i})
Яблоки	5000	6000	12	10
Бананы	2000	2500	25	24
Апельсины	4000	3800	16	14

Необходимо исчислить индексы цен, физического объема товарооборота в фактических ценах по трем товарам вместе.

Рассчитаем индекс цен:

$$I_p = \frac{\sum_{i=1}^{n} q_{1i} p_{1i}}{\sum_{i=1}^{n} q_{1i} p_{0i}} = \frac{10 \cdot 6000 + 24 \cdot 2500 + 14 \cdot 3800}{12 \cdot 6000 + 25 \cdot 2500 + 16 \cdot 3800} = \frac{173200}{195300} = 0.8867$$

Цены снизились на 11,33%, и покупатель имел экономию, равную 22100 руб. (19530 – 173200).

Определим индекс физического объема товарооборота:

$$I_q = \frac{\sum_{i=1}^{n} q_{1i} p_{0i}}{\sum_{i=1}^{n} q_{0i} p_{0i}} = \frac{6000 \cdot 12 + 2500 \cdot 25 + 3800 \cdot 16}{5000 \cdot 12 + 2000 \cdot 25 + 4000 \cdot 16} = \frac{195300}{174000} = 1.1223$$

Товарооборот в неизменных ценах вырос на 12,23%, прирост товарооборота в неизменных ценах составил 21300 руб. (195300 – 174000).

Рассчитаем индекс товарооборота в фактических ценах:

$$I_{pq} = \frac{\sum_{i=1}^{n} p_{1i} q_{1i}}{\sum_{i=1}^{n} p_{0i} q_{0i}} = \frac{173200}{174000} = 0{,}995$$

Товарооборот в фактических ценах снизился на 0,5%, что в абсолютном выражении составляет 800 руб. (174000 – 173200). Произведение первых двух индексов дает третий индекс

$$I_p \cdot I_q = I_{pq} \text{ или } 0{,}8867 \cdot 1{,}1223 = 0{,}995$$

В определенной связи находятся и разности между знаменателем и числителем индексов: населению по ценам базисного периода было продано товаров на 21300 руб. больше, но в силу того, что население имело экономию от снижения цен на товары в сумме 22100 руб., оно за эти товары в отчетном периоде по фактическим ценам уплатило на 800 руб. меньше.

Контрольные вопросы

1. Понятие индексного метода в статистике.
2. Индивидуальные и общие индексы.
3. Агрегатный индекс цен.
4. Агрегатные индексы физического объема товарной массы.
5. Индексы с постоянным и переменным весом.
6. Средние индексы.
7. Расчет недостающих индексов с помощью индексных систем.

ГЛАВА 13. СИСТЕМА ПОКАЗАТЕЛЕЙ ХОЗЯЙСТВЕННОЙ ДЕЯТЕЛЬНОСТИ ПРЕДПРИЯТИЙ

13.1 Статистика кадров предприятия

Инвестиционный характер производства, его высокая наукоемкость, приоритетность вопросов качества продукции изменили требования к работнику, повысили значимость творческого отношения к труду и высокого профессионализма. Это привело к существенным изменениям в принципах, методах и социально-психологических вопросах управления персоналом на предприятии.

Хорошо подобранный **трудовой коллектив** – одна из основных задач предпринимателя. Это должна быть команда единомышленников и партнеров, способных осознавать, понимать и реализовывать замыслы руководства предприятия. Только она служит залогом успеха предпринимательской деятельности, выражения и процветания предприятия.

Трудовые отношения – едва ли не самый сложный аспект работы предприятия. Гораздо легче справиться с техническими и технологическими неполадками, чем разрешить конфликтные ситуации, возникающие в коллективе, где нужно учитывать индивидуальные склонности, личностные установки, психологические предпочтения.

Основными аспектами влияния человеческого фактора на повышение эффективности работы предприятия являются:

- отбор и продвижение кадров;
- подготовка кадров и их непрерывное обучение;
- стабильность и гибкость состава работников;
- совершенствование материальной и моральной оценки труда работников.

Для решения вопросов, связанных с персоналом на современных предприятиях создаются специальные службы управления персоналом (СУП), отдельное подразделение профессиональных специалистов – менеджеров, главной целью которых является повышение производственной, творческой отдачи и активности персонала, разработка и реализация программ развития кадром.

Независимо от сферы приложения труда весь персонал предприятия подразделяется на категории. В настоящее время принято выделять следующие категории персонала: **рабочие, служащие, специалисты и руководители.**

К рабочим относят работников предприятия, непосредственно занятых созданием материальных ценностей или оказанием производственных и транспортных услуг. Рабочие подразделяются на **основных**

и вспомогательных. Их соотношение – **аналитический** показатель работы предприятия.

Коэффициент численности основных рабочих определяется по формуле:

$$K_{OP} = 1 - \frac{N_{BP}}{N_O} \qquad (13.1)$$

где N_{BP} – среднесписочная численность вспомогательных рабочих на предприятии, в цехах, на участке, чел.; N_0 – среднесписочная численность всех рабочих на предприятии в цехе, на участке,.

Специалисты и руководители осуществляют организацию производственного процесса и руководство им (директора, мастера, главные специалисты и др.).

К служащим относятся работники, осуществляющие финансово-расчетные, снабженческо-сбытовые и другие функции (агенты, кассиры, делопроизводители, секретари, статистики и др.).

Квалификация работ определяется уровнем специальных знаний и практических навыков и характеризует степень сложности выполняемого им конкретного вида работы. Соответствие его способностей, физических и психических качеств той или иной профессии означает профессиональную пригодность работника.

Структура кадров предприятия, цеха, участка характеризуется соотношением различных категорий работников в их общей численности. В целях анализа структуры кадров определяется и сравнивается удельный вес каждой категории работников $\mathbf{dp_i}$ в общей среднесписочной численности персонала предприятия $\overline{N}$.

$$dp_i = \frac{\overline{N_i}}{\overline{\overline{N}}}, \text{ или } dp_i = \frac{\overline{N_i}}{\overline{\overline{N}}} \times 100 \qquad (13.2)$$

где N_i – среднесписочная численность работников i-ой категории.

Структура кадров определяется и анализируется по каждому подразделению, а также может рассматриваться по таким признакам, как возраст, пол, уровень образования, стаж работы, квалификация, степень выполнения норм и т.д.

Трудовой коллектив по численному составу, уровню квалификации не является постоянной величиной, он все время изменяется: увольняются одни работники, принимаются другие. Изменения такого рода характеризуются **текучестью кадров**.

Состояние кадров на предприятии может быть определено с помощью следующих коэффициентов.

Коэффициент выбытия кадров $К_{ВК}$ (%) определяется отношением количества работников, уволенных по всем причинам за данный период $N_{УВ}$ к среднесписочной численности работников за тот же период $\overline{N}$:

$$K_{ВК} = \frac{N_{УВ}}{\overline{N}} \times 100 \qquad (13.3)$$

Коэффициент приема кадров $К_{ПК}$ (%) определяется отношением количества работников, принятых на работу за данный период $N_{П}$ к среднесписочной численности работников за тот же период $\overline{N}$:

$$K_{ПК} = \frac{N_{П}}{\overline{N}} \times 100 \qquad (13.4)$$

Коэффициент стабильности кадров $К_{СК}$ рекомендуется использовать при оценке уровня организации управления производством как на предприятии в целом, так и в отдельных подразделениях:

$$K_{СК} = 1 - \frac{N_{УВ}}{\overline{N}} + N_{П} \qquad (13.5)$$

где $N_{УВ}$ – численность работников, уволившихся с предприятия по собственному желанию и из-за нарушения трудовой дисциплины за отчетный период.

Коэффициент текучести кадров $К_{ТК}$ (%) определяется делением численности работников предприятия (цеха, участка), выбывших или уволенных за данный период $N_{УВ}$, на среднесписочную численность за тот же период $\overline{N}$:

$$K_{ТК} = \frac{N_{УВ}}{\overline{N}} \times 100 \qquad (13.6)$$

13.2 Статистика рабочей силы и рабочего времени

Статистика рабочей силы изучает главным образом состав и численность рабочей силы. В отраслях сферы материального производства рабочая сила делится на персонал, занятый в **основной деятельности** предприятия, и персонал **не основной деятельности.**

Основной категорией персонала являются рабочие. Рабочие группируются по профессиям, по степени механизации труда (рабочие механизированного и рабочие ручного труда) и по квалификации. Основным показателем квалификации служит тарифный разряд или тарифный коэффициент. Средний уровень квалификации определяется средним тарифным разрядом, исчисляемым как средняя арифметическая разрядов, взвешенная по численности или по проценту рабочих

$$\overline{P} = \frac{\sum_{i=1}^{n} P_i N_i}{\sum_{i=1}^{n} N_i} \qquad (13.7)$$

где P_i – тарифные разряды, а N_i – численность рабочих с данным разрядом.

Кроме этого, все работники группируются по полу, возрасту, стажу работы и образованию.

Основными категориями численности рабочих и служащих в народном хозяйстве являются списочная и явочная численность, число фактически работавших. В списочную численность, включаются все работники предприятия, принятые на срок один и более дней. Явочное число включает работников, явившихся на работу, а также находящихся в командировках и занятых на других предприятиях по нарядам своей организации. Численность фактически работавших может быть меньше явочной численности за счет работников, имевших целодневный простой по независящим от них причинам .

Все эти категории численности определяются на конкретную дату (день), однако для многих экономических расчетов (определение производительности труда, средней зарплаты и др.) требуется знать среднюю численность работников – **среднесписочную, среднеявочную и среднюю фактически работавших.**

В зависимости от исходных данных **среднесписочную** численность находят следующими способами:

1. Если известна списочная численность на начало и конец периода, среднесписочная численность определяется как полусумма этих величин.

2. Среднесписочная численность за **квартал, полугодие и год** может быть найдена как средняя арифметическая из среднемесячных чисел:

$$\overline{N} = \frac{\sum_{i=1}^{n} \overline{N}_i}{n} \qquad (13.8)$$

где $\overline{N}_i$ – среднемесячная численность, n – число месяцев периода

3. Если известна списочная численность на **даты** через одинаковые интервалы времени, например на начало или конец каждого месяца, то среднесписочная численность за квартал, полугодие или год находится по формуле средней хронологической:

$$\overline{N} = \frac{\frac{N_1}{2} + N_2 + \ldots + \frac{N_i}{2}}{i-1} \qquad (13.9)$$

где i – число показателей (дат); N_1 – численность на первую дату, N_2, …N_i – на другие даты.

4. Наиболее точные результаты дает следующий способ:

$$\overline{N} = \frac{\sum_{j=1}^{n} N_j}{n} \qquad (13.10)$$

где N_j – списочная численность работников за календарный день, n – число календарных дней.

Среднеявочная численность работников определяется по формуле:

$$\overline{N}_m = \frac{\sum_{j=1}^{n} N_{mj}}{n} \qquad (13.11)$$

где N_{mj} – человеко-дни явок, n – число рабочих дней.

Средняя численность фактически работавших вычисляется по формуле:

$$\overline{N}_ф = \frac{\sum_{j=1}^{n} N_{фj}}{n} \qquad (13.12)$$

где $N_{фj}$ – число отработанных человеко-дней, n – число рабочих дней.

При определении **средней списочной численности** следует учитывать, что если предприятие начало работать не с начала года, эта численность рассчитывается на весь календарный период. Например, завод начал работать с I июня, и в июне средняя списочная численность работников составила 300 чел. Тогда средняя численность за II квартал составит 300:3 = 100 чел., за 1 полугодие – 300:6 = 50 чел.

Рабочее время измеряется в человеко-днях и человеко-часах. Статистика выделяет следующие фонды рабочего времени (в человеко-днях):

Календарный фонд ($Т_{КФ}$) включает все время отчетного периода и равен произведению числа календарных дней в периоде на списочную численность работников.

Табельный фонд ($Т_{ТФ}$) меньше календарного на число праздничных и выходных человеко-дней.

Максимально возможный фонд ($Т_{МВФ}$) меньше табельного фонда за счет времени очередных отпусков.

Фактически отработанный фонд времени меньше максимально возможного за счет различных потерь рабочего времени.

Использование фондов времени измеряется следующими коэффициентами:

$$K_{КФ} = \frac{T}{T_{КФ}}, \qquad K_{ТФ} = \frac{T}{T_{ТФ}}, \qquad K_{МВФ} = \frac{T}{T_{МВФ}} \qquad (13.13)$$

где $К_{КФ}$ – коэффициент использования календарного фонда, $К_{ТФ}$ – коэффициент использования табельного фонда, $К_{МВФ}$ – коэффициент использования максимально возможного фонда.

Статистика также анализирует использование сменного времени. Для этого применяют следующие показатели:

$$K_C = \frac{\sum_{i=1}^{n} N_{Ci}}{N_{C\max}}, \; K_{CP} = \frac{K_C}{n}, \; K_H = \frac{N_{C\max}}{n}, \; K_{CУ} = K_H \times K_{CP} \quad (13.14)$$

где $К_С$ – коэффициент сменности, N_{Ci} – число рабочих, занятых во всех сменах, N_{Cmax} – число рабочих в наибольшей смене, $К_{СР}$ – коэффициент использования сменного режима, n – число смен, $К_Н$ – коэффициент непрерывности.

13.3 Характеристика производительности труда

Под производительностью труда понимается результативность конкретного живого труда, эффективность целесообразной производительной деятельности по созданию продукта в течение определенного промежутка времени. Перед статистикой производительности труда стоят задачи:

1) совершенствования методики расчета производительности труда;

2) выявления факторов роста производительности труда;

3) определения влияния производительности труда на изменение объема продукции.

В экономической практике уровень производительности труда характеризуется через показатели **выработки и трудоемкости**. Выработка (W) продукции в единицу времени измеряется соотношением объема произведенной продукции (V) и затратами (T) рабочего времени (среднесписочная численность): W = V/T. Это прямой показатель производительности труда. Обратным показателем является трудоемкость: t = T/V, откуда W = 1/t.

Система статистических показателей производительности труда определяется единицей измерения объема произведенной продукции. Эти единицы могут быть натуральными, условно-натуральными, трудовыми и стоимостными. Соответственно применяют **натуральный, условно-натуральный, трудовой**

и стоимостный методы измерения уровня и динамики производительности труда.

В зависимости от того, чем измеряются затраты труда, различают следующие уровни его производительности.

Средняя часовая выработка ($\overline{W}_Ч$):

$$\overline{W}_Ч = \frac{V}{N_T} \qquad (13.15)$$

где N_T – число человеко-часов, отработанных в течение данного времени. Она показывает среднюю выработку рабочего за один час фактической работы (исключая время внутрисменных простоев и перерывов, но с учетом сверхурочной работы).

Средняя дневная выработка ($\overline{W}_Д$):

$$\overline{W}_Д = \frac{V}{N_{T\Sigma}} \qquad (13.16)$$

где $N_{T\Sigma}$ – число человеко-часов, отработанных всеми рабочими предприятия. Она характеризует степень производственного использования рабочего дня.

Средняя месячная выработка ($\overline{W}_M$):

$$\overline{W}_M = \frac{V}{\overline{N}_i} \qquad (13.17)$$

где $\overline{N}_i$ – среднесписочная численность рабочих. В этом случае в знаменателе отражаются не затраты, а резервы труда.

Средняя квартальная выработка рассчитывается аналогично среднемесячной. В настоящее время средняя выработка характеризуется через соотношение товарной продукции (объема продукции, работ, услуг) и среднесписочной численности промышленно-производственного персонала (ППП).

Между вышеперечисленными средними показателями существует взаимосвязь:

$$W_{1ППП} = W_Ч \times T_{РД} \times T_{РП} \times d_{РППП} \qquad (13.18)$$

где $W_{1ППП}$ – выработка на одного работника; $W_Ч$ – среднечасовая выработка; $T_{РД}$ – продолжительность рабочего дня; $T_{РП}$ – продолжительность рабочего периода; $d_{РППП}$ – доля рабочих в общей численности промышленно-производственного персонала.

Динамика производительности труда в зависимости от метода измерения ее уровня и анализируется при помощи статистических индексов (см. главу 12): **натуральных** (1), **трудовых** (2) и **стоимостных** (3):

$$1.\ I_w = \frac{\sum_{i=1}^{n} t_{1i} \sum_{i=1}^{n} q_{1i}}{\sum_{i=1}^{n} t_{0i} \sum_{i=1}^{n} q_{1i}}, \qquad 2.\ I_w = \frac{\sum_{i=1}^{n} t_{1i} q_{1i}}{\sum_{i=1}^{n} t_{0i} q_{1i}}, \qquad 3.\ I_w = \frac{\sum_{i=1}^{n} q_{1i} p_{0i}}{\sum_{i=1}^{n} T_{1i}} : \frac{\sum_{i=1}^{n} q_{0i} p_{0i}}{\sum_{i=1}^{n} T_{0i}},$$

где $\sum_{i=1}^{n} t_{1i} q_{1i}$ – общие затраты времени на единицу продукции текущего (отчетного) года;

$\sum_{i=1}^{n} t_{0i} q_{1i}$ – общие затраты времени на единицу продукции по базисным нормам времени;

$\frac{\sum_{i=1}^{n} q_{1i} p_{0i}}{\sum_{i=1}^{n} T_{1i}}$ – уровень производительности труда одного работника в текущем (отчетном) периоде;

$\frac{\sum_{i=1}^{n} q_{0i} p_{0i}}{\sum_{i=1}^{n} T_{0i}}$ – средняя выработка одного работника в прошлом (базисном) периоде.

13.4 Анализ основного капитала предприятия

Капитал является одним их факторов производства. Внешне капитал представлен в конкретных формах – средства производства (производственный капитал), деньги (денежный), товары (товарный).

Часть производственного капитала (здания, сооружения, машины и оборудование) носит название **основного капитала**.

Другая часть производственного капитала (сырье, материалы, энергетические ресурсы и др.) представляют собой **оборотный капитал**. В практике учета применяются термины «основные средства», «основные фонды».

Основные производственные фонды – та часть производственных фондов, которая участвует в процессе производства длительное время, сохраняя при этом свою натуральную форму, а их стоимость переносится на изготовляемый продукт постепенно, по частям, по мере использования. Пополняются они за счет капитальных вложений.

Наряду с производственными существуют **непроизводственные основные фонды** – жилые дома, детские и спортивные учреждения и другие объекты культурно-бытового обслуживания трудящихся, которые находятся на балансе предприятий. В отличие от производственных фондов они не участвуют в процессе производства и не переносят своей стоимости на продукт, ибо он не производится. Стоимость их исчезает в потреблении. Фонд возмещения не создается. Воспроизводятся они за счет национального дохода.

Несмотря на то, что непроизводственные основные фонды не оказывают непосредственного влияния на объем производства, рост производительности труда, постоянное увеличение этих фондов неразрывно связано с улучшением благосостояния работников предприятия и повышением материального и культурного уровня их жизни, что, в конечном счете, сказывается на результатах деятельности предприятия.

Основные производственные фонды – материально-техническая база общественного производства. От их объема зависят производственная мощность предприятия и в значительной мере уровень технической вооруженности труда. Накопление основных фондов и повышение технической вооруженности труда обогащают процесс труда, придают труду творческий характер, повышают культурно-технический уровень общества.

Основные производственные фонды промышленности – это огромное количество средств труда, которые несмотря на свою экономическую однородность отличаются целевым назначением, сроком службы. Отсюда возникает необходимость классификации основных фондов по определенным группам, учитывающим специфику производственного назначения различных видов фондов.

По действующей видовой классификации основные производственные фонды промышленных предприятий делятся на следующие группы:

- здания, сооружения;
- передаточные устройства;
- машины и оборудование, в том числе силовые машины и оборудование, рабочие машины и оборудование, измерительные и регулирующие приборы и устройства и лабораторное оборудование, вычислительная техника, прочие машины и оборудование;
- инструменты и приспособления, служащие более года и стоящие более 1 млн руб. за штуку. Инструменты и инвентарь, служащие менее года или стоящие дешевле 1 млн руб. за штуку, относятся к оборотным средствам как малоценные и быстроизнашивающиеся;
- производственный и хозяйственный инвентарь.

Соотношение отдельных групп основных фондов в их общем объеме представляет **видовую (производственную) структуру основных фондов.** Обществу не безразлично, в какую из групп основных фондов вкладываются средства. Оно заинтересовано в оптимальном повышении удельного веса машин, оборудования – активной части основных фондов, которые обслуживают решающие участки производства и характеризуют производственные возможности предприятия по выпуску тех или иных изделий

Здания, сооружения, инвентарь, обеспечивающие нормальное функционирование активных элементов основных фондов, относятся к **пассивной части основных фондов.**

Чем выше доля оборудования в стоимости основных производственных фондов, тем при прочих равных условиях больше выпуск продукции, выше показатель фондоотдачи. Поэтому улучшение структуры основных производственных фондов рассматривается как условие роста производства и показателя фондоотдачи, снижения себестоимости, увеличения денежных накоплений предприятий.

Важнейшими факторами, влияющими на структуру основных производственных фондов, являются: характер выпускаемой продукции, объем выпуска продукции, уровень автоматизации и механизации, уровень специализации и кооперирования, климатические и географические условия расположения предприятий.

Влияние первого фактора сказывается на величине и стоимости зданий, доле транспортных средств и передаточных устройств. Чем больше объем выпуска продукции, тем выше удельный вес специальных прогрессивных рабочих машин и оборудования. Такая же картина характерна и в отношении влияния на структуру фондов третьего и четвертого факторов. От климатических условий зависит доля зданий, сооружений.

Улучшить структуру основных производственных фондов позволяют:

- обновление и модернизация оборудования;
- совершенствование структуры оборудования за счет увеличения доли прогрессивных видов станков и машин, особенно станков для выполнения финишных операций, автоматических и полуавтоматических станков, универсальных агрегатных станков, автоматических линий, станков с числовым программным управлением;
- лучшее использование зданий и сооружений, установка дополнительного оборудования на свободных площадках;
- правильная разработка проектов строительства и высококачественное выполнение планов строительства предприятий;
- ликвидация лишнего и малоиспользуемого оборудования и установка оборудования, обеспечивающего более правильные пропорции между его отдельными группами.

Учет и планирование основных фондов ведутся в **натуральной и денежной формах.** При оценке основных фондов в натуральной форме устанавливаются число машин, их производительность, мощность, размер производственных площадей и другие количественные величины. Эти данные используются для расчета производственной мощности предприятий и отраслей, планирования производственной программы, резервов повышения выработки на оборудовании, составления баланса оборудования. С этой целью ведутся инвентаризация и паспортизация оборудования, учет его выбытия и прибытия.

Денежная, или стоимостная, оценка основных фондов необходима для планирования расширенного воспроизводства основных фондов, определения степени износа и размера амортизационных отчислений, объема приватизации.

Существует несколько видов оценок основных фондов, связанных с длительным участием их и постепенным снашиванием в процессе производства, изменением за этот период условий воспроизводства: по первоначальной, восстановительной и остаточной стоимости.

Первоначальная стоимость основных производственных фондов – это сумма затрат на изготовление или приобретение фондов, их доставку и монтаж.

Восстановительная стоимость – затраты на воспроизводство основных фондов в современных условиях; как правило, она устанавливается во время переоценки фондов. Последняя переоценка основных фондов была произведена по состоянию на 1 января 1995 г.

Остаточная стоимость (ОС) представляет собой разность между первоначальной или восстановительной стоимостью основных фондов и суммой их износа.

Основные производственные фонды в процессе функционирования изнашиваются, перенося свою стоимость на произведенную продукцию.

Амортизация – это денежное выражение стоимости износа основных фондов, перенесенной на продукцию. Она включается в себестоимость продукции.

Годовая сумма амортизационных отчислений определяется по формуле:

$$A = \frac{B - Л}{T},$$

где **В** – полная первоначальная стоимость основных фондов;

Л – ликвидационная стоимость основных фондов за вычетом расходов на их демонтаж;

Т – нормативный срок службы основных фондов;

М – предполагаемая стоимость модернизации в течение всего эксплуатационного периода.

Годовые нормы амортизации определяются по формулам:

$$N_a = \frac{A}{B} \cdot 100 \text{ , или } \quad N_a = \frac{B + M + K - Л}{B \cdot T}$$

Для характеристики изменения объема и движения основных фондов, их воспроизводства, как правило, за год составляются балансы основных фондов.

На основе балансов основных фондов можно проанализировать процессы их воспроизводства, а также изучить динамику, исчислить показатели обновления, выбытия и состояния основных фондов.

Источником поступления основных фондов является:

- ввод в действие новых основных фондов;
- покупка основных фондов у юридических лиц и физических лиц;
- безвозмездное получение основных фондов других юридических лиц и физических лиц;
- аренда основных фондов.

Выбытие происходит по следующим причинам:

- ликвидация из-за ветхости и износа;
- продажа основных фондов другим юридическим и физическим лицам;
- безвозмездная передача;
- передача основных фондов в долговременную аренду.

На основе данных балансов как по балансовой стоимости, так и по стоимости за вычетом износа можно рассчитать целый ряд показателей, которые характеризуют состояние и воспроизводство основных фондов:

фондоотдача: $$\varPhi_0 = \frac{S_{ПП}}{S_{ОФ}} ,$$

фондоемкость: $$\varPhi_e = \frac{1}{\varPhi_0} ,$$

фондовооруженность: $$\varPhi_B = \frac{\overline{S}_{ОФГ}}{\overline{N}} ,$$

где $S_{ПП}$ – стоимость произведенной продукции, $S_{ОФ}$ – средняя величина стоимости ОФ за период, $\overline{S}_{ОФГ}$ – среднегодовая стоимость ОФ, $\overline{N}$ – среднесписочная численность.

13.5. Анализ оборотного капитала предприятия

Источниками образования основного капитала являются долгосрочные финансовые вложения, а отличительным признаком – достаточно продолжительный период использования средств, вложенных в основной капитал в целях извлечения прибыли; **оборотный капитал** – это финансовые ресурсы, вложенные в объекты, использование которых осуществляется предприятием либо в рамках относительно **короткого календарного периода** времени (как правило, не более одного года).

Таблица 13.1.

<table>
<tr><td rowspan="13">ВЕСЬ ОБОРОТ ПРОИЗВОДСВА</td><td rowspan="5">КАПИТАЛ В ПРОИЗВ.</td><td>Производственные запасы</td><td rowspan="8">НОРМИРУЕМЫЕ ОБОРОТ. СРЕДСТВА</td></tr>
<tr><td>Незавершенное производство и полуфабрикаты</td></tr>
<tr><td>Незавершенное сельскохозяйственное производ.</td></tr>
<tr><td>Корма и фураж</td></tr>
<tr><td>Расходы будущих отчетных периодов</td></tr>
<tr><td rowspan="8">КАПИТАЛ В ОБРАЩЕНИИ</td><td>Готовая продукция</td></tr>
<tr><td>Товары</td></tr>
<tr><td>Прочие товарно-материальные ценности</td></tr>
<tr><td>Товары отгруженные</td><td rowspan="5">НЕНОРМИР. ОБОРОТ. СР.</td></tr>
<tr><td>Денежные средства</td></tr>
<tr><td>Дебиторы</td></tr>
<tr><td>Краткосрочные финансовые вложения</td></tr>
<tr><td>Прочие оборотные средства</td></tr>
</table>

Состав оборотного капитала фирмы по более или менее укрупненным позициям можно представить в виде, показанном в таблиц 13.1.

Деление оборотного капитала на нормируемый и ненормируемый необходимо крупным и средним по размерам предприятиям, потому что расчеты потребности в оборотном капитале, требуемом для обеспечения нормального хода воспроизводственного процесса, могут делаться как на основе более или менее подробных технико-экономических расчетов (потребность в сырье и материалах, запасы готовых изделий), так и по данным учета, статистики и экспертных оценок (в частности дебиторской задолженности, например).

Источниками образования элементов оборотного капитала предприятия являются во всех случаях финансовые ресурсы. В их составе выделяют собственные средства (входящие в состав уставного капитала, специальных фондов и образуемые за счет прибыли) и привлеченные средства. К привлеченным относят полученные в коммерческих банках ссуды (кредиты), коммерческий кредит, кредиторскую задолженность поставщикам и привлеченные средства юридических и физических лиц (депозиты, реализованные сторонним лицам облигации, выданные векселя и пр.).

Таким образом, **оборотный капитал** состоит из активов, которые находятся в постоянном движении и превращаются в денежные средства.

Статистика изучает наличие и структуру оборотных фондов, показатели их оборачиваемости, материалоемкость продукции, удельные расходы и другие показатели использования материальных ресурсов.

Для характеристики использования оборотных фондов служат показатели скорости их обращения:

1. Коэффициент оборачиваемости характеризует число оборотов среднего остатка производственных оборотных фондов за отчетный период:

$$K_{обор} = \frac{P}{CO},$$

где Р – стоимость реализованной продукции за период; СО – средний остаток оборотных фондов, определяемый как средняя арифметическая из средних месячных (за квартал, полугодие, год) или как средняя хронологическая.

2. Коэффициент закрепления оборотных фондов – величина, обратная коэффициенту оборачиваемости, показывает, сколько надо иметь оборотных средств на 1 руб. стоимости реализованной продукции.

3. Средняя продолжительность одного оборота оборотных фондов в днях:

$$K_{закр} = \frac{CO}{P} = \frac{1}{K_{обор}} \text{ (руб.).}$$

4. Средняя продолжительность одного оборота оборотных фондов в днях:

$$П = CO \times \frac{D}{P} = \frac{D}{K_{обор}} = D \times K_{закр} \text{ (дни),}$$

где D – число дней в периоде (обычно принимают месяц – 30 дней, квартал – 90 дней, год – 360 дней).

Для совокупности предприятий рассчитывают средние показатели скорости обращения оборотных фондов. При этом коэффициент оборачиваемости и закрепления исчисляют как средние арифметические взвешенные:

$$\overline{K}_{обор} = \frac{\sum_i K_{обор.i} \times CO_i}{\sum_i CO_i} \qquad \overline{K}_{закр} = \frac{\sum_i K_{закр.i} \times P_i}{\sum_i P_i}$$

Средняя продолжительность одного оборота в днях определяется как средняя гармоническая взвешенная:

$$\overline{\Pi} = \frac{\sum_i CO_i}{\sum_i \frac{1}{\Pi_i} \times CO_i}$$

Эффект от ускорения оборачиваемости оборотных фондов выражается суммой фондов, условно высвобожденных из оборота вследствие ускорения их оборачиваемости. Эту сумму в рублях можно определить по нескольким формулам:

$$Э = (\Pi_0 - \Pi_1) \times \frac{P_1}{D} = \frac{P_1 \times \Pi_0}{D} - CO_1 = (K_{закр.1} - K_{закр.0}) \times P_1$$

где $П_0$, $П_1$ – продолжительность одного оборота в базисном и отчетном периодах, $Р_1$ – объем реализации продукции в отчетном периоде, $СО_1$ – средний остаток оборотных фондов в отчетном периоде, D – число дней в периоде, $К_{закр.1}$ и $К_{закр.0}$ – коэффициенты закрепления в базисном и отчетном периодах.

Сводным показателем использования предметов труда является **материалоемкость**, характеризующая в денежном выражении расход материальных ресурсов на единицу результата производства. **Показатель материалоемкости** вычисляется по формуле:

$$ME = \frac{МЗ}{Q}$$

где МЗ – материальные производственные затраты без амортизации основных фондов, Q – объем совокупного общественного продукта, национального дохода или продукции отдельных отраслей и предприятий.

13.6. Статистическое изучение финансов предприятия

Финансы предприятий (фирм) – представляют собой финансовые отношения, выраженные в денежной форме, возникающие при образовании, распределении и использовании денежных фондов и накоплений в процессе производства и реализации товаров, выполнения работ и оказание различных услуг.

Предметом изучения статистики финансов предприятий является количественная характеристика их финансово-денежных отношений с учетом их качественных особенностей, обусловленных образованием, распределением и использованием финансовых ресурсов, выполнением обязательств хозяйствующих субъектов друг перед другом, перед финансово-банковской системой и государством.

Основным задачами статистики финансов предприятия являются:

- изучение состояния и развития финансово-денежных отношений хозяйствующих субъектов;
- анализ объема и структуры источников формирования финансовых ресурсов;
- определение направлений использования денежных средств;
- анализ уровня и динамики прибыли, рентабельности (доходности) предприятия;
- оценка финансовой устойчивости и состояния платежеспособности;
- оценка выполнения хозяйствующими субъектами финансово-кредитных обязательств.

Финансовые ресурсы – это денежные средства (собственные и привлеченные) хозяйствующих субъектов, находящиеся в их распоряжении и предназначенные для выполнения финансовых обязательств и осуществления затрат для производства.

Объем и состав финансовых ресурсов непосредственно связан с уровнем развития предприятия и его эффективностью. Чем успешнее деятельность предприятия, тем выше размеры его денежных доходов.

Первоначальное формирование финансовых ресурсов происходит в момент учреждения предприятия, когда образуется уставный фонд. Его источниками выступают:

- акционерный капитал;
- паевые взносы членов кооперативов;
- долгосрочный кредит;
- бюджетные средства.

На действующих предприятиях в условиях рыночной экономики важнейшими источниками формирования финансовых ресурсов являются:

- прибыль от реализованной продукции, выполненных работ и оказанных услуг;
- амортизационные отчисления, поступления от продажи акций, ценных бумаг;
- краткосрочные и долгосрочные кредиты;
- доходы от продажи имущества и т.д.

Прибыль – экономическая категория, комплексно отражающая хозяйственную деятельность предприятия в форме денежных накоплений.

Прибыль, характеризующая конечные результаты торгово-производственной деятельности, является основным показателем финансового состояния предприятия.

Различают в статистике финансов предприятий:

- прибыль балансовая;
- прибыль от реализации продукции (работ, услуг);
- валовая прибыль;
- чистая прибыль.

Балансовая прибыль – это финансовые результаты от реализации продукции (работ, услуг) основных средств и другого имущества хозяйствующих субъектов, а также доходы за вычетом убытков от внереализационных операций.

Прибыль от реализаций продукции рассчитывается как разность между выручкой от ее продажи (за вычетом НДС) и затратами на производство и реализацию, включаемыми в себестоимость продукции.

Валовая прибыль в отличие от балансовой прибыли в составе внереализационых доходов и убытков учитывает уплаченные штрафы и пени (за исключением суммы штрафов и пени, перечисленных в бюджет и внебюджетные фонды).

Прибыль, оставшаяся в распоряжении предприятия после уплаты налогов и других платежей в бюджет – является **чистой прибылью**.

Предприятия определяют направления, объемы и характер использования чистой прибыли. Как правило, формируется фонд развития производства, фонд накопления, социального развития и фонд материального поощрения резервный фонд.

Из образуемых на предприятиях фондов выделяют средства на стимулирование заинтересованности работников в результатах труда, приобретение акций, текущие расходы, благотворительные нужды и т.д.

Экономическая деятельность предприятия оценивается при помощи показателя рентабельности и деловой активности.

Рентабельность – это прибыльность предприятия.

1. **Общая рентабельность**:

$$P_0 = \frac{П_б}{Ф} \times 100,$$

где $П_б$ – общая сумма балансовой прибыли, Ф – среднегодовая стоимость основных производственных фондов и нормируемых оборотных средств.

2. **Рентабельность реализованной продукции:**

$$P_{р.п.} = \frac{П_{р.п.}}{C} \times 100,$$

где $П_{р.п.}$ – прибыль от реализации продукции, С – полная себестоимость реализованной продукции.

3. **Деловая активность предприятия** определяется с помощью показателя общей оборачиваемости капитала:

$$O_к = \frac{B}{K},$$

где В – выручка от реализации продукции, К – основной капитал предприятия.

В условиях рыночной экономики важное значение имеет анализ финансовой устойчивости предприятия.

Под финансовой устойчивостью понимается способность хозяйствующего субъекта своевременно из собственных средств возмещать затраты, вложенные в основной и оборотный капитал, нематериальные активы, и расплачиваться по своим обязательствам, т.е. быть платежеспособным.

Для оценки устойчивости предприятия применяются коэффициенты.

1. **Коэффициент автономии**:

$$K_a = \frac{C_C}{S_C},$$

где C_C – собственные средства, S_C – сумма всех источников финансовых ресурсов.

2. **Коэффициент соотношения** заемных и собственных средств (коэффициент устойчивости):

$$K_{уст.} = \frac{K_З}{C_C},$$

где $К_З$ – кредиторская задолженность и другие заемные средства (без кредитов банков и займов).

Оптимальным считается вариант, если $К_{уст.} = 1$.

3. **Коэффициент маневренности**:

$$K_M = \frac{C_C + ДКЗ + O_{CB}}{C_C},$$

где ДКЗ – долгосрочные кредиты и займы, $О_{СВ}$ – основные средств и иные внеоборотные активы.

4. **Коэффициент ликвидности**:

$$K_{ликв.} = \frac{Д_{CA}}{K_З},$$

где $Д_{СА}$ – денежные средства, вложенные в ценные бумаги, запасы товарно-материальных ценностей, дебиторская задолженность, $К_З$ – краткосрочная задолженность.

Контрольные вопросы

1. Что входит в статистику кадров на предприятии?
2. Что входит в понятие статистики рабочей силы предприятия?
3. Статистика рабочего времени.
4. Основные характеристики производительности труда.
5. Дать краткий статистический анализ основного и оборотного капиталов предприятия.
6. Что входит в статистический анализ финансов предприятия?

ЛИТЕРАТУРА

1. Ефимова М.Р., Рябцев В.М. Общая теория статистики. М.: Финансы и статистика, 1991. 302 с.
2. Общая теория статистики / Под ред. А.А. Спириной. М.: Финансы и статистика, 1996. 296 с.
3. Харченко Л.П., Долженкова В.Г., Ионин В.Г. и др. Статистика: Курс лекций / Под ред. В.Г. Ионина. Новосибирск: изд. НГЭиУ, М.: ИНФРА-М, 1997. 310 с.
4. Ефимова М.Р., Петрова Е.В., Румянцев В.Н. Общая теория статистики: Учебник. М.: ИНФРА-М, 1997. 416 с.
5. Статистика рынка товаров и услуг / Под ред. И.К. Белявского. М.: Финансы и статистика, 1997. 432 с.
6. Годин А.М. Статистика: Учебник. М.: Дашков и К^{О}, 2002. 472 с.
7. Теория статистики / Под ред. Р.А. Шмойловой, М.: Финансы и статистика. 1999. 560 с.
8. Елисеева И.И., Юзбашев М.М. Общая теория статистики: Учебник / Под ред.чл.-корр. РАН И.И. Елисеевой. М.: Финансы и статистика, 2002. 480 с.

ПРИЛОЖЕНИЕ

Приложение 1

Функция нормального распределения

$$F_{(x)} = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{x} e^{-\frac{t}{2}} dt$$

t	0	1	2	3	4	5	6	7	8	9
1,0	0,841	844	846	849	851	853	855	858	860	862
1,1	864	867	869	871	873	857	877	879	881	883
1,2	885	887	889	891	893	894	896	898	900	901
1,3	903	905	907	908	910	911	913	915	916	918
1,4	913	921	922	924	925	926	928	929	931	932
1,5	933	934	936	837	938	939	941	942	943	944
1,6	945	946	947	948	950	951	952	953	954	954
1,7	955	906	957	938	959	960	961	962	962	963
1,8	964	965	966	967	967	969	969	969	970	971

Приложение 2

Критические значения критерия χ^2 Пирсона

P / K	0,90	0,95	0,975	0,99
1	2,71	3,84	5,02	6,64
2	4,61	5,99	7,38	9,21
3	6,25	7,82	9,39	11,35
4	7,78	9,49	11,14	13,28
5	9,24	11,07	12,03	15,09
6	10,65	12,59	14,45	16,81
7	12,02	14,07	16,01	18,48
8	13,36	15,51	17,54	10,09
9	14,68	16,92	19,02	21,67
10	15,99	18,31	20,48	23,21

Приложение 3

Критические значения корреляционного отношения η^2 и коэффициента детерминации R^2

K_2 \ K_1	1	2	3	4	5	6	8	10	20
а) уровень значимости а = 0,05									
3	0,771	865	903	924	938	947	959	967	983
4	658	776	832	865	887	905	924	937	967
5	569	699	764	806	835	854	885	904	948
6	500	632	704	751	785	811	847	871	928
7	444	575	651	702	739	768	810	839	908
8	399	527	604	657	697	729	775	807	887
9	362	488	563	618	659	692	742	777	867
10	332	451	527	582	524	659	711	749	847
б) уровень значимости а = 0,01									
3	0,919	954	967	975	979	982	987	989	994
4	841	900	926	941	951	958	967	973	986
5	765	842	879	901	916	928	943	953	974
6	696	785	830	859	879	894	915	929	961
7	636	732	784	818	842	860	887	904	946
8	585	684	740	778	806	827	858	879	931
9	540	641	700	741	741	795	829	854	914
10	501	602	663	706	738	764	802	829	898

Приложение 4

Критические значения F-критерия

R_1 / R_2	1	2	3	4	5	6	8	10	20
а) уровень значимости а = 0,05									
4	7,71	6,94	6,59	6,39	6,26	6,16	6,04	5,96	5,80
5	6,61	5,79	5,41	5,19	5,05	4,95	4,82	4,74	4,56
6	5,99	5,14	4,76	4,53	4,39	4,28	4,15	4,06	3,87
7	5,59	4,74	4,35	4,12	3,97	3,87	3,73	3,63	3,44
8	5,32	4,46	4,07	3,84	3,69	3,58	3,44	3,34	3,15
9	5,12	4,26	3,86	3,63	3,48	3,37	3,23	3,13	2,93
10	4,96	4,10	3,71	3,48	3,33	3,22	3,07	2,97	2,77
б) уровень значимости а = 0,01									
30	7,56	5,39	4,51	4,02	3,70	3,47	3,17	2,98	2,55
40	7,31	5,18	4,31	3,83	3,51	3,29	2,99	2,80	2,37
60	7,08	4,98	4,13	3,65	3,34	3,12	2,82	2,63	2,20
120	6,85	4,79	3,95	3,48	3,17	2,96	2,66	2,47	2,03

Николай Григорьевич Филонов

СТАТИСТИКА

Учебное пособие

Ответственный за выпуск: Л.В. Домбраускайте
Технический редактор: С.Н. Чуков

Сдано в печать: 05.04.2004 г.	Печать трафаретная
Подписано в печать: 20.02.2004 г.	Бумага офсетная
Тираж: 150 экз.	Уч. изд. л. 8,58
Формат: 60x84/16	Усл.-печ. л. 13,00

Заказ: 017/У

Центр учебно-методической литературы
Томского Государственного Педагогического Университета

Отпечатано в типографии ТГПУ:
г. Томск, ул. Герцена, 49. Тел. (3822) 52-12-93