Московская финансово-промышленная академия

Минашкин В.Г.
Шмойлова Р.А.
Садовникова Н.А.
Моисейкина Л.Г.
Рыбакова Е.С.

Теория статистики

Москва, 2004

УДК 311
ББК 60.6
М 613

Минашкин В.Г., Шмойлова Р.А., Садовникова Н.А., Моисейкина Л.Г., Рыбакова Е.С. Теория статистики / Московская финансово-промышленная академия, М., - 2004 г., 198 с.

Авторы:
проф. Минашкин В.Г. – гл. 1, 5, 7, 10, 11.
проф. Шмойлова Р.А. – гл. 4, 9.
проф. Садовникова Н.А. – гл. 3, 8.
доц. Моисейкина Л.Г. – гл. 6.
доц. Рыбакова Е.С. – гл. 2.

Содержание

Введение

Полная и достоверная статистическая информация является тем необходимым основанием, на котором базируется процесс управления экономикой. Принятие управленческих решений на всех уровнях - от общегосударственного или регионального и до уровня отдельной корпорации или частной фирмы - невозможно без должного статистического обеспечения.

Именно статистические данные позволяют определить объемы валового внутреннего продукта и национального дохода, выявить основные тенденции развития отраслей экономики, оценить уровень инфляции, проанализировать состояние финансовых и товарных рынков, исследовать уровень жизни населения и другие социально-экономические явления и процессы.

Статистика - это наука, изучающая количественную сторону массовых явлений и процессов в неразрывной связи с их качественной стороной, количественное выражение закономерностей общественного развития в конкретных условиях места и времени.

Для получения статистической информации органы государственной и ведомственной статистики, а также коммерческие структуры проводят различного рода статистические исследования. Процесс статистического исследования включает три основные стадии: сбор данных, их сводка и группировка, анализ и расчет обобщающих показателей.

От того, как собран первичный статистический материал, как он обработан и сгруппирован в значительной степени зависят результаты и качество всей последующей работы. Недостаточная проработка программно-методологических и организационных аспектов статистического наблюдения, отсутствие логического и арифметического контроля собранных данных, несоблюдение принципов формирования групп в конечном итоге могут привести к абсолютно ошибочным выводам.

Не менее сложной, трудоемкой и ответственной является и заключительная, аналитическая стадия исследования. На этой стадии рассчитываются средние показатели и показатели распределения, анализируется структура совокупности, исследуется динамика и взаимосвязи между изучаемыми явлениями и процессами.

Используемые на всех стадиях исследования приемы и методы сбора, обработки и анализа данных являются предметом изучения *общей теории статистики*, которая является базовой отраслью статистической науки. Разработанная ею методология применяется в макроэкономической статистике, отраслевых статистиках (промышленности, сельского хозяйства, торговли и прочих), статистике населения, социальной статистике и в других статистических отраслях.

Данное учебное пособие подготовлено преподавателями кафедры Прикладной статистики МФПА: проф. Минашкиным В.Г. (ред.), проф. Шмойловой Р.А., проф. Садовниковой Н.А., доц. Моисейкиной Л.Г., доц. Рыбаковой Е.С. Пособие предназначено для студентов, изучающих курсы «Элементарные методы анализа статистических данных», «Теория статистики» или «Общая теория статистики».

Глава 1. Предмет, метод и организация статистики

1.1. Статистика как наука и отрасль практической деятельности

Термин *статистика* имеет несколько значений. Во-первых, под статистикой понимают отрасль практической деятельности по сбору, обработке, анализу и публикации статистической информации как в целом по стране, так и по отдельным ее регионам. Такая деятельность, с определенными различиями в используемой методологии, осуществляется во всех странах. В России эта работа выполняется Федеральной службой государственной статистики (старое название организации -Государственный комитет Российской Федерации по статистике).

Статистикой также часто называют и сам результат статистической деятельности, т.е. массив статистических данных или обобщающие показатели, характеризующие состояние массовых явлений и процессов по той или иной совокупности за определенный период. Потребителями статистической информации являются органы государственного управления, научные организации, информационные агентства, аналитические службы компаний и банков, физические лица. В последние годы стремительно повышается значение статистической информации в маркетинговых исследованиях.

Статистика как наука начала формироваться в VII веке в ответ на потребность государства располагать достоверными статистическими данными об имеющихся ресурсах для эффективного управления, организации производства, торговли, налогообложения и т.п. В настоящее время ***статистика*** - это наука, включающая разветвленную систему научных дисциплин, изучающих количественную сторону массовых явлений и процессов в неразрывной связи с их качественной стороной.

Изучаемые статистикой явления и процессы многообразны. В первую очередь, статистика изучает все, что связано с экономической деятельностью общества - производство и реализация промышленной и сельскохозяйственной продукции, строительство новых объектов и реконструкция действующих основных фондов, работа транспорта и связи, формирование и движение финансовых потоков. Статистические методы широко используются в анализе социальных процессов и явлений - занятости и безработицы, доходов населения, изучении общественного мнения и т.д. Большую роль играет статистика в технике и производственной деятельности, например, в организации контроля качества продукции. Методы статистики применяются в экономическом анализе, менеджменте, маркетинге, бизнес-планировании, логистике, оценке недвижимости, антикризисном управлении и в других областях научной и практической деятельности.

Рассмотрим отраслевую структуру статистики как науки.

Теория статистики (общая теория статистики) – отрасль статистической науки, рассматривающая ее общие понятия, категории, принципы и методы сбора, обработки и анализа данных. Теорией статистики разрабатываются общие показатели и методы изучения структуры, взаимосвязи и динамики изучаемых процессов и явлений. Использование этих показателей и методов в отдельных областях научной и практической деятельности наполняет их качественным содержанием, а в ряде случаев – придает им определенную специфику.

Экономическая (макроэкономическая) статистика изучает количественные закономерности происходящих в экономике явлений и процессов, выявление основных пропорций и тенденций экономического развития на макроуровне, т.е. на уровне крупного региона или страны в целом. Экономическая статистика изучает как сам процесс воспроизводства материальных благ и услуг, так и его результаты, а также их воздействие на уровень жизни населения. К основным показателям экономической статистики относятся валовой внутренний продукт, валовой региональный продукт, такие элементы национального богатства как основные фонды, материальные и оборотные средства, домашнее имущество.

В соответствии с классификацией отраслей экономики в статистической науке и практике также выделяется отраслевой уровень. К отраслевым статистикам относятся:

- *статистика промышленности;*
- *статистика сельского хозяйства;*
- *статистика капитального строительства;*
- *статистика услуг, транспорта и связи;*
- *статистика торговли.*

Статистика населения изучает численный и национальный состав, а также возрастно-половую структуру населения, его размещение и воспроизводство как по стране в целом, так и в разрезе территориальных единиц. Одной из основных задач статистики населения является построение краткосрочных и долгосрочных демографических прогнозов.

Социальная статистика изучает социальную структуру населения, его уровень жизни и, в частности, доходы, а также уровень образования и культуры, состояния здоровья и медицинского обслуживания, использование свободного времени, общественное мнение, уровень преступности и другие социальные аспекты жизнедеятельности общества.

Для того, чтобы получить общее представление о статистической методологии, необходимо рассмотреть сам *процесс статистического исследования*, который включает четыре основных этапа:

Процесс статистического исследования начинается с этапа сбора первичного статистического материала, проверки его полноты и достоверности. С этой целью применяются методы сплошного и несплошного *статистического наблюдения*. От качества полученных исходных статистических данных во многом зависят окончательные результаты всего статистического исследования.

На втором этапе производится предварительная обработка данных, подсчет групповых и общих итогов, расчет некоторых относительных показателей. Основной метод, используемый на данном этапе - *метод группировок*. В результате его реализации от больших массивов статистических данных осуществляется переход к компактным и удобным для анализа статистическим таблицам.

Третий этап – расчет и интерпретация обобщающих статистических показателей. На данном этапе рассчитываются показатели *среднего уровня* и *вариации, структуры, взаимосвязи* и *динамики* изучаемых процессов и явлений. Полученные результаты подвергаются анализу.

В процессе реализации четвертого этапа осуществляется *моделирование взаимосвязей* между социально-экономическими процессами и явлениями, строятся уравнения регрессии, а также *трендовые модели*, отражающие основные тенденции динамики изучаемых показателей.

Используемые в процессе реализации всех этапов статистические приемы и методы в целом составляют статистическую методологию исследования.

1.2. Статистическая деятельность в Российской Федерации

Для сбора, обработки и анализа статистической информации в настоящее время в нашей стране функционирует единая централизованная система государственной статистики. Центральным органом этой системы является Федеральная служба государственной статистики. В субъектах Российской Федерации - республиках, краях, областях и районах - статистическая работа осуществляется территориальными органами государственной статистики, комитетами или отделами.

Непосредственная обработка поступающих из регионов статистических данных осуществляется в Главном межрегиональном центре обработки и распространения статистической информации, который обладает необходимыми для этих целей мощными вычислительными ресурсами.

На Федеральную службу государственной статистики возложено как методологическое, так и практическое руководство всеми работами по сбору, обработке и анализу статистических данных на государственном уровне. Для реализации этих задач в структуре Федеральной службы государственной статистики выделены следующие Управления:

- организации статистического наблюдения и контроля;
- национальных счетов;
- статистики предприятий и других хозяйствующих субъектов;
- сводных статистических работ, общественных и международных связей;
- статистики цен и финансов;
- статистики торговли и услуг;
- статистики труда, образования, науки и культуры;
- статистики уровня жизни и обследований домашних хозяйств;
- статистики населения;
- административное;
- финансово-хозяйственного, информационного и производственно-технологического обеспечения.

Федеральная служба государственной статистики ежегодно разрабатывает и утверждает Федеральную программу статистических работ на календарный год, которая согласовывается на заседании Правительства Российской Федерации.

Работа по сбору статистической информации проводится не только Федеральной службой государственной статистики. В соответствии с Федеральной программой отдельные виды статистических работ осуществляются другими органами государственного управления - Банком России, Минфином России, Минздравом России, МВД России и др. Получаемые Федеральной службой государственной статистики статистические данные прежде всего предоставляются органам федеральной власти, а также публикуются для широкого использования в аналитических целях научными и практическими работниками, руководителями и специалистами предприятий и организаций всех форм собственности. Назовем основные статистические ежегодные издания:

- Российский статистический ежегодник;
- Россия в цифрах;
- Регионы России;
- Промышленность России;
- Строительство в России;
- Сельское хозяйство в России;
- Малое предпринимательство в России;
- Жилищное хозяйство в России;
- Финансы России;
- Цены в России;
- Транспорт в России;
- Женщины и мужчины России;
- Россия и страны мира.

К периодическим – ежемесячным и ежеквартальным статистическим изданиям – относятся:

- «Статистическое обозрение» (ежеквартальный журнал на русском языке);
- «Статистическое обозрение» (ежеквартальный журнал на английском языке);
- «Информация о социально-экономическом положении России» (ежемесячный краткий доклад);
- «Социально-экономическое положение России» (ежемесячный доклад);
- «Вопросы статистики» (ежемесячный научно-информационный журнал).

С важнейшими социально-экономическими показателями Российской Федерации можно познакомиться через сеть INTERNET на сайте Федеральной службы государственной статистики - http://www.gks.ru.

1.3. Основные категории статистики

Перед тем, как приступить к изучению основных статистических показателей, приемов и методов статистического исследования, необходимо познакомиться с используемой в статистике терминологией, с основными категориями статистики.

Важнейшей категорией статистической науки является категория признака. Именно значения различных признаков наблюдаются и регистрируются на первой стадии статистического исследования - стадии статистического наблюдения. *Признак* - это объективная характеристика единицы статистической совокупности, характерная черта или свойство, которое может быть определено или измерено. Признаками, характеризующими промышленное предприятие, является выручка от реализации продукции, прибыль, стоимость основных фондов, численность персонала и др. Признаками человека являются возраст, пол, место жительства, профессия, среднемесячный доход и пр. Для любых окружающих нас объектов и явлений можно выделить достаточно большое число признаков, которые наблюдаются или потенциально могут наблюдаться в процессе статистического исследования.

Возможное значение, которое может принимать признак, называется *вариантом*. Например, существуют всего четыре варианта значений признака “экзаменационная оценка”: “2”, “3”, “4”, “5”. Если же учитывать оценки, проставляемые в зачетную книжку бакалавра или магистра, то таких вариантов остается три, так как неудовлетворительная оценка в зачетку не проставляется. У отдельно взятого учащегося в зачетке могут быть и десять, и двадцать, и более значений признака “экзаменационная оценка”, но вариантов будет по-прежнему три, а возможно, два или один, если, например, студент или слушатель учится без троек и четверок.

Признаки подразделяются на количественные и качественные, а последние, в свою очередь, на альтернативные, атрибутивные и порядковые.

Количественным является признак, отдельные варианты которого имеют числовое выражение и отражают размеры, масштабы изучаемого объекта или явления. К количественным признакам, например, относятся доход домохозяйства, площадь жилого помещения, цена товара, стаж работы. Количественные признаки в статистике преобладают над другими видами признаков, они наиболее информативны, аналитичны, именно на работу с данными признаками нацелена большая часть многообразного статистического инструментария.

Альтернативным называется признак, имеющий только два варианта значений. Например, продукция предприятия может соответствовать предъявляемым требованиям или быть бракованной, пол человека может быть мужским или женским, население страны или региона обычно делится на городское и сельское. Альтернативный признак может иметь и числовое выражение. Предположим, при анкетировании потребителей вопрос о доходах в анкете предполагал всего два варианта: “до 5 тыс. рублей в месяц” и “5 тыс. рублей в месяц и более”. В этом случае количественный признак был преобразован в альтернативный.

В отличие от альтернативного *атрибутивный* признак имеет более двух вариантов, которые при этом выражаются в виде понятий или наименований. К атрибутивным признакам относятся район проживания, вид продукции, специальность работника, цвет товара. Такие признаки имеют место в различных областях исследования, но в большей степени они характерны для информации, с которой работают маркетологи, социологи, психологи.

Порядковые признаки отличаются от атрибутивных тем, что они имеют несколько ранжированных, т.е. упорядоченных по возрастанию или убыванию, качественных вариантов. Примерами таких признаков являются уровень образования (начальное, общее среднее и т.д.), уровень квалификации, воинское звание, различного рода рейтинги. Отдельные варианты порядкового признака трудно соизмерить количественно. Например, понятно, что высшее образование лучше, чем среднее специальное, но при этом нельзя утверждать, что оно лучше на 20% или на 30%. Водительская категория “Е” выше, чем водительская категория “В”, но количественных пропорций между ними не существует.

Следует отметить, что порядковый признак может иметь числовое выражение. В качестве примеров можно привести такие порядковые признаки как разряд рабочего, тарифный разряд служащего, рейтинговые оценки, экзаменационные оценки. Школьник, получивший четверку, не обязательно продемонстрировал ровно в два раза больше знаний по сравнению со школьником, получившим двойку. Рабочий 6-го разряда не обязательно в два раза больше вырабатывает продукции и в два раза больше зарабатывает по сравнению с рабочим 3-го разряда. В обозначении вариантов этих признаков цифры можно заменить буквами алфавита без какого-либо снижения их информативности.

Приведенные выше примеры показывают, что изучаемые статистикой признаки как правило подвержены вариации. *Вариация* – это колеблемость, изменение величины признака в статистической совокупности, т.е. принятие единицами совокупности или их группами разных значений признака.

Статистической совокупностью называется множество подвергающихся статистическому исследованию объектов или явлений, объединенных общими признаками, из которых один или несколько признаков не варьируют. Статистика имеет дело с совокупностями промышленных, сельскохозяйственных, строительных и торговых предприятий, с совокупностью коммерческих банков, с совокупностью населения страны или отдельного ее региона. Так, например, всех жителей г. Москвы можно рассматривать как статистическую совокупность, так как один признак - город проживания - будет неварьирующий. По остальным же признакам - полу, возрасту, социальному положению - население будет варьировать.

Индивидуальный составной элемент статистической совокупности, являющийся носителем изучаемых признаков, называется *единицей совокупности*. Для отрасли единицей совокупности будет являться отдельное предприятие, для банковской системы - отдельный банк. В некоторых случаях для одной и той же совокупности можно выделить разные группы единиц. Например, при изучении половозрастной структуры населения единицей является отдельный человек, при изучении же доходов, обеспеченности жильем, предметами длительного пользования (телевизоры, холодильники и т.п.) единицей будет являться домохозяйство.

Общее число единиц, образующих статистическую совокупность, называется *объемом совокупности*.

Объем совокупности следует отличать от *объема признака,* т.е. суммарного значения признака по всем единицам изучаемой совокупности. Так, например, число предприятий в отрасли - это объем совокупности, а общий выпуск продукции на всех предприятиях отрасли - это объем признака. В некоторых случаях объем признака не имеет реального экономического смысла, например, трудно интерпретировать суммарный рост всех студентов одной группы. Но для расчета отдельных статистических показателей, в частности - средних, такое суммирование необходимо.

Одной из важнейших характеристик статистической совокупности является ее однородность. *Однородной* является совокупность, единицы которой близки между собой по значениям признаков, существенных для данного исследования, или же они относятся к одному и тому же типу. Многие методы и приемы статистического исследования применимы лишь к однородным совокупностям.

Большую роль в статистическом исследовании играет *закон больших чисел* – общий принцип, в силу которого количественные закономерности, присущие массовым явлениям, отчетливо проявляются лишь при достаточно большом числе наблюдений. Единичные явления в большей степени подвержены действию случайных и несущественных факторов, чем масса в целом. При большом числе наблюдений случайные отклонения в ту или иную сторону от общей закономерности развития взаимно погашаются. В результате взаимопогашения случайных отклонений обобщающие показатели, исчисленные для величин одного и того же вида, становятся типичными, отражающими действие постоянных и существенных факторов в данных условиях места и времени.

Статистическое исследование независимо от его масштабов и целей всегда завершается расчетом и анализом различных по виду и форме выражения статистических показателей.

Статистический показатель представляет собой количественную характеристику социально-экономических явлений и процессов в условиях качественной определенности. Качественная определенность показателя заключается в том, что он непосредственно связан с внутренним содержанием изучаемого явления или процесса, его сущностью.

Как правило, изучаемые статистикой процессы и явления достаточно сложны, и их сущность не может быть отражена посредством одного отдельно взятого показателя. В таких случаях используется система статистических показателей.

Система статистических показателей - это совокупность взаимосвязанных показателей, имеющая одноуровневую или многоуровневую структуру и нацеленная на решение конкретной статистической задачи. Так, например, сущность промышленного предприятия заключается в производстве какой-либо продукции на базе эффективного взаимодействия средств производства и трудовых ресурсов. Следовательно, для полной экономической характеристики функционирования предприятия необходимо использовать систему, включающую прежде всего такие показатели, как прибыль, рентабельность, численность промышленно-производственного персонала, производительность труда, фондовооруженность и др.

В отличие от признака статистический показатель получается расчетным путем. Это могут быть простой подсчет единиц совокупности, суммирование их значений признака, сравнение двух или нескольких величин, а также более сложные расчеты.

Различают конкретный статистический показатель и показатель-категорию. *Конкретный статистический показатель* характеризует размер, величину изучаемого явления или процесса в данном месте и в данное время (под привязкой к месту понимается отношение показателя к какой-либо территории или объекту). Так, если мы называем конкретную величину стоимости промышленно-производственных фондов, то обязательно должны указать, к какому предприятию или отрасли и какому моменту времени она относится. Однако в теоретических работах и на этапе проектирования статистического наблюдения (при построении системы статистических показателей, обосновании методики их расчета) также оперируют и абстрактными показателями или показателями-категориями.

Показатель-категория отражает сущность, общие отличительные свойства конкретных статистических показателей одного и того же вида без указания места, времени и числового значения. Например, показатели розничного товарооборота предприятий торговли и общественного питания в Москве и Санкт-Петербурге в 2000 и 2003 гг. отличаются местом, временем и конкретными числовыми значениями, но имеют одну и ту же сущность (продажа товаров через розничную торговую сеть и сеть предприятий общественного питания), которая отражена в показателе-категории «Розничный товарооборот предприятий торговли и общественного питания».

Все статистические показатели по охвату единиц совокупности разделяются на индивидуальные и сводные, а по форме выражения - на абсолютные, относительные и средние.

Индивидуальные показатели характеризуют отдельный объект или отдельную единицу совокупности - предприятие, фирму, банк, домохозяйство и т. п. Примером индивидуальных абсолютных показателей может служить численность промышленно-производственного персонала предприятия, оборот торговой фирмы, совокупный доход домохозяйства.

На основе соотнесения двух индивидуальных абсолютных показателей, характеризующих один и тот же объект или единицу, получают индивидуальный относительный показатель. В статистике рассчитываются и индивидуальные средние показатели, но только во временном измерении (среднегодовая численность промышленно-производственного персонала предприятия).

Сводные показатели в отличие от индивидуальных характеризуют группу единиц, представляющую собой часть статистической совокупности или всю совокупность в целом. Эти показатели, в свою очередь, подразделяются на объемные и расчетные

Объемные показатели получают путем сложения значений признака отдельных единиц совокупности. Полученная величина, называемая объемом признака, может выступать в качестве объемного абсолютного показателя (например, стоимость основных фондов предприятий отрасли), а может сравниваться с другой объемной абсолютной величиной (например, с численностью промышленно-производственного персонала этих предприятий) или объемом совокупности (в данном примере- с числом предприятии). В последних двух случаях получают объемный относительный и объемный средний показатели (в наших примерах - фондовооруженность и средняя стоимость основных фондов).

Расчетные показатели, вычисляемые по различным формулам, служат для решения отдельных статистических задач анализа - измерения вариации, характеристики структурных сдвигов, оценки взаимосвязи и т. д. Они также делятся на абсолютные, относительные или средние. В эту группу входят индексы, коэффициенты тесноты связи, ошибки выборки и прочие показатели, подробно рассмотренные в соответствующих главах.

Охват единиц совокупности и форма выражения являются основными, но не единственными классификационными признаками статистических показателей. Важным классификационным признаком является также временной фактор. Социально-экономические процессы и явления находят свое отражение в статистических показателях либо по состоянию на определенный момент времени, как правило, на определенную дату, начало или конец месяца, года (численность населения, стоимость основных фондов, дебиторская задолженность), либо за определенный период - день, неделю, месяц, квартал, год (производство продукции, число заключенных браков, сумма страховых выплат). В первом случае показатели являются *моментными*, во втором - *интервальными*.

В зависимости от принадлежности к одному или двум объектам изучения различают *однообъектные и межобъектные показатели*. Если первые характеризуют только один объект, то вторые получают в результате сопоставления двух величин, относящихся к разным объектам (соотношение численности населения городов Тулы и Рязани, соотношение численности детей дошкольного возраста и числа мест в детских дошкольных учреждениях и т. п.). Межобъектные показатели выражаются в форме относительных или средних величин.

С точки зрения пространственной определенности статистические показатели подразделяются на *общетерриториальные*, характеризующие изучаемый объект или явление в целом по стране, *региональные и местные (локальные)*, относящиеся к какой-либо части территории или отдельному объекту.

Глава 2. Статистическое наблюдение

2.1. Сущность и виды статистического наблюдения

Статистическое исследование начинается со сбора статистической информации, характеризующей изучаемые социально-экономические явления и процессы. Данный этап называется статистическим наблюдением.

Статистическое наблюдение – это массовое, планомерное, научно организованное наблюдение за социально-экономическими явлениями и процессами, заключающееся в регистрации необходимых признаков у каждой единицы изучаемой совокупности. Например, при переписи населения по каждому жителю страны регистрируются сведения о поле, возрасте, семейном положении, образовании и др.

Статистическое наблюдение, как правило, носит массовый характер. Это проявляется в том, что при проведении наблюдения необходимо получить данные от максимально возможного числа изучаемых единиц совокупности. Массовый охват совокупности позволяет получать наиболее точные данные, характеризующие изучаемое социально-экономическое явление, выявить имеющиеся закономерности и взаимосвязи.

Планомерность проведения статистического наблюдения заключается в том, что любое исследование проводится по заранее разработанному плану, который включает в себя ряд вопросов, касающихся подготовительных работ, непосредственного сбора необходимой информации и обработки полученных данных.

Принцип научной организации лежит в основе любого этапа статистического исследования и заключается в комплексном применении статистической методологии сбора и обработки данных.

Основная цель статистического наблюдения – это сбор статистической информации о социально-экономических явлениях и процессах для получения обобщающих характеристик.

На современном этапе в статистике существует две основные формы статистического наблюдения:

- отчетность;
- специально организованное статистическое наблюдение;
- регистры.

Отчетность – это способ получения статистической информации от юридических лиц. Отчетность представляет собой специально разработанные формы, включающие в себя те признаки, которые подлежат регистрации. Формы статистической отчетности разрабатываются и утвержденные органами государственной статистики РФ. Одна из форм статистической отчетности представлена в приложении 1. Любое юридическое лицо, являющееся субъектом экономики РФ, обязано предоставлять отчетность органам государственной статистики по месту своей регистрации по установленным отчетным формам и в установленные сроки.

В период формирования рыночной экономики особое место в системе сбора статистической информации стали занимать **специально организованные статистические наблюдения**, которые проводятся для получения каких – либо данных, не содержащихся в предоставляемой отчетности или которые необходимы для проверки или уточнения данных, содержащихся в отчетах.

Особо следует выделить такой вид специально организованного наблюдения, как перепись.

Перепись – это специально проводимые широкомасштабные работы по сбору необходимой статистической информации об изучаемых объектах в границах отрасли, региона или страны в целом. Так, например, ранее упоминались переписи населения, которые проводятся примерно один раз в 10 лет и направлены на получение необходимой информации о населении страны. Примером также могут служить переписи крупного рогатого скота, которые проводятся в конце календарного года и позволяют получить информацию о численности и структуре поголовья крупного рогатого скота у сельхозпроизводителей. Органами статистики также проводятся переписи многолетних насаждений, жилого фонда, незавершенного строительства и пр.

Кроме переписей, к специально организованному наблюдению также относятся и другие единовременные работы по сбору необходимой статистической информации, в частности, в рамках социологических или маркетинговых исследований.

Регистровое наблюдение представляет собой постоянный мониторинг состояния и развития наблюдаемых единиц, заключающийся в первичном размещении и своевременной актуализации информации в ведущейся базе данных. В статистической практике ряда стран находят применение регистры населения, т.е. постоянно актуализируемые списки жителей страны с указанием их основных социально-демографических признаков, а также регистры предприятий, содержащие информацию организационно-правового и экономического характера.

Общая классификация видов статистического наблюдения представлена на рис.2.1.

По *охвату единиц совокупности* наблюдение бывает двух видов: сплошное и несплошное.

При **сплошном наблюдении** обследованию подвергаются все единицы изучаемой совокупности. При этом в силу действия ряда факторов возможен незначительный процент неохвата единиц изучаемой совокупности. Примером сплошного наблюдения могут служить переписи населения.

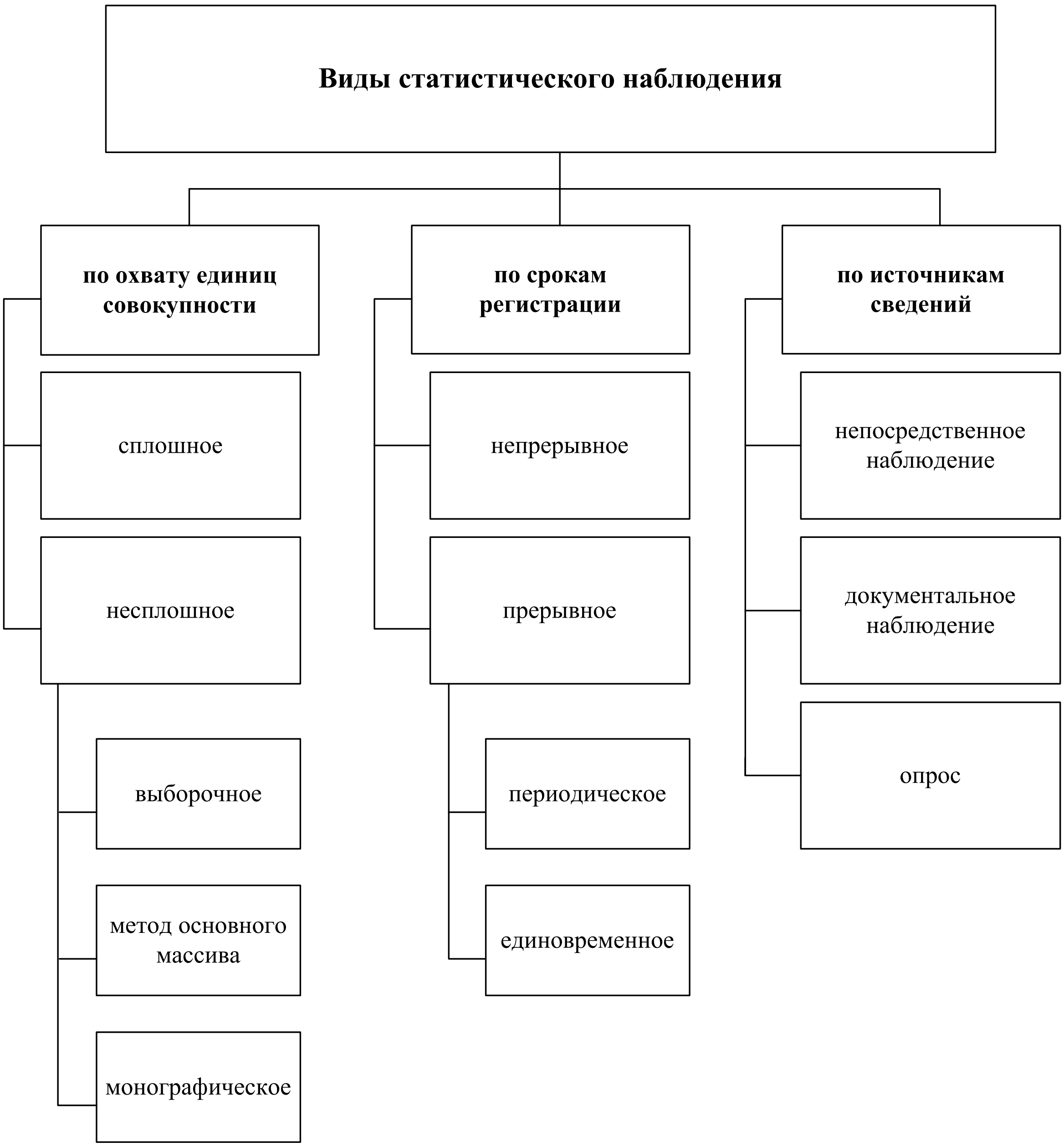

Рис.2.1. Виды статистического наблюдения

При **несплошном наблюдении** обследованию подвергается только лишь часть единиц изучаемой совокупности. При этом охватываемая наблюдением часть определяется заранее, т.е. неудавшееся сплошное наблюдение нельзя рассматривать как наблюдение несплошное. Принято выделять следующие виды несплошного статистического наблюдения: выборочное, метод основного массива, монографическое обследование.

Выборочным называют наблюдение, основанное на принципе случайного отбора тех единиц изучаемой совокупности, которые должны быть подвергнуты наблюдению. Выборочное наблюдение, при правильной его организации и проведении, дает достаточно достоверные данные для характеристики изучаемой совокупности в целом. Во многих случаях им вполне можно заменить сплошной учет. При этом обеспечивается значительная экономия средств, затрачиваемых на сбор и обработку данных.

Монографическое обследование представляет собой детальное, глубокое изучение и описание отдельных, характерных в каком-либо отношении единиц совокупности, как правило, по расширенной программе. Монографическое исследование проводится с целью выявления имеющихся или намечающихся тенденций в развитии явления, для выявления имеющихся резервов, оценки результатов экономических экспериментов.

Метод основного массива заключается в том, что обследованию подвергаются наиболее крупные единицы, которые вместе взятые имеют преобладающий удельный вес в совокупности по основному для данного исследования признаку. Например, в ряде отраслей подавляющий объем выпуска продукции приходится на крупные и средние предприятия, поэтому результаты деятельности малых предприятий в этих отраслях практически не отражаются на обобщающих статистических показателях.

По *срокам регистрации* наблюдение может быть непрерывным (текущим) и прерывным.

Непрерывным называют такое наблюдение, которое ведется постоянно, и регистрация фактов производится по мере их свершения. Так, например, осуществляется регистрация рождений, заключенных браков и т.п. в органах ЗАГС.

Прерывное наблюдение проводится не постоянно, время от времени. При этом прерывное наблюдение бывает двух видов: периодическое и единовременное. Периодическое — это наблюдение, которое повторяется через определенные, равные промежутки времени. В качестве примера можно выделить ежегодное предоставление отчетности в органы государственной статистики.

Единовременным называется такое наблюдение, которое проводится по мере необходимости, без соблюдения строгой периодичности или вообще проводится один раз и больше не повторяется. Таковым наблюдением являлась перепись многолетних насаждений, проведенная в прошлом веке.

По *источнику сведений* различают непосредственное наблюдение, документальное наблюдение и опрос.

Непосредственным называют такое наблюдение, при котором сами регистраторы путем непосредственного замера, взвешивания или подсчета устанавливают значение признака и на этом основании производят запись в формуляре наблюдения. Этим способом проводится инвентаризации основных средств на предприятиях.

Документальное наблюдение предполагает запись ответов на вопросы формуляра на основании соответствующих документов. Примером такого наблюдения является сбор данных об успеваемости студентов вуза на основе зачетно-экзаменационных ведомостей, заполнение форм статистической отчетности на основании данных бухгалтерского учета и т.п.

Опрос — это наблюдение, при котором ответы на вопросы формуляра записываются со слов опрашиваемого (респондента). Этим способом проводятся переписи населения, опросы общественного мнения.

В статистике применяются следующие *способы* сбора сведений:

- отчетный,
- экспедиционный,
- самоисчисление,
- анкетный,
- корреспондентский.

Сущность **отчетного способа** заключается, как уже отмечалось выше, в обязательном представлении хозяйствующими субъектами статистических отчетов о своей деятельности в установленной форме и в установленные сроки.

Экспедиционный способ наблюдения заключается в том, что специально привлеченные и обученные работники посещают каждую единицу наблюдения и сами заполняют формуляр наблюдения. Этим способом собираются сведения при переписях населения.

При способе **самоисчисления** формуляры заполняют сами опрашиваемые. Обязанность специально привлеченных для получения информации сотрудников состоит в раздаче формуляров опрашиваемым, инструктаже их, сборе заполненных формуляров и проверке правильности их заполнения.

Анкетный способ — это сбор статистических данных с помощью специальных вопросников, рассылаемых определенному кругу лиц или публикуемых в периодической печати. Как правило этим способом получения информации пользуются при проведении социологических опросов и также многие крупные производители бытовой техники, мебели и других предметов потребления. Анкеты вкладываются в упаковку товара с просьбой заполнить и вернуть производителю по указанному адресу. Анкета фирмы SONY представлена в приложении 2.

Сущность **корреспондентского способа** наблюдения заключается в том, что статистические органы договариваются с определенными лицами, которые берут на себя обязательство вести наблюдение за какими-либо явлениями, процессами и в установленные сроки сообщать результаты наблюдений статистическим органам. Таким способом изучаются бюджеты отдельных домохозяйств, цель которых – получение статистической информации о доходах и расходах населения.

2.2. План статистического наблюдения

Как уже отмечалось выше, планомерность является основой статистического наблюдения, поэтому его проведение должно основываться на детально разработанном плане.

План статистического наблюдения состоит из двух частей, первая включает программно-методологические вопросы, а вторая организационные вопросы.

Программно-методологическая часть плана включает:

— определение объекта наблюдения;

— определение единицы объекта наблюдения;

— составление программы статического наблюдения;

— составление программы разработки материалов наблюдения;

— проектирование формуляра наблюдения;

— определение времени проведения статистического наблюдения и его критического момента;

— составление инструкции.

При планировании статистического наблюдения необходимо, прежде всего, определить его объект и единицу.

Объектом статистического наблюдения называется та совокупность, о которой должны быть собраны нужные сведения. Объектами наблюдения могут быть, например, коммерческие банки, сельхозпроизводители, промышленные предприятия, студены, население и т.п.

Единицей наблюдения называют составной элемент объекта наблюдения, являющийся носителем признаков, подлежащих регистрации. Единицей наблюдения может быть человек, фермерское хозяйство, коммерческий банк.

Программа наблюдения – это перечень признаков, подлежащих регистрации при проведении статистического наблюдения. К программе наблюдения предъявляется ряд требований, которым она должна удовлетворять, а именно:

а) программа должна включать только существенные признаки, характеризующие изучаемый объект;

б) в программу не следует включать второстепенные вопросы, которые могут затруднить работу по сбору информации, а в дальнейшем ее обработку и анализ;

в) разрабатывая программу, необходимо стремиться к полноте собираемых сведений;

г) в программу наблюдения должны включаться только такие вопросы, на которые действительно можно получить объективные и достаточно точные ответы;

д) в программу иногда следует включать вопросы контрольного характера, служащие целям проверки и уточнения собираемых сведений.

Вопросы программы могут дополняться статистическим подсказом, т.е. вариантами ответов. Подсказ может быть закрытым или открытым. Закрытый подсказ подразумевает ряд ответов, из которых респондент должен выбрать один или несколько. При открытом подсказе респондент может выбрать один или несколько ответов из предлагаемого перечня или сформулировать на специально выделенном поле формуляра свой собственный ответ.

При планировании обследования как правило составляют и **программу разработки** собранных материалов, которая конкретизирует задачи статистического наблюдения, показывает, какие данные необходимо собирать и в каком виде оформлять результаты их обработки.

Для записи ответов на вопросы программы конструируется **формуляр** наблюдения. Формуляр наблюдения разрабатывается специально для записи ответов на вопросы программы и представляет собой особым образом разграфленный лист (листы) бумаги, в котором содержится перечень вопросов программы, свободные места для записи ответов на них, а также для записи шифров (кодов) ответов. Особое внимание при разработке формуляра следует уделить формулировке вопросов. Они должны быть сформулированы кратко и четко, не должны вызывать разночтения. Помимо вопросов программы формуляр включает в себя титульную и адресную части. В титульной части содержится наименование статистического наблюдения, указывается наименование органа, проводящего наблюдение, кем и когда утвержден этот формуляр, иногда и номер, присвоенный ему в общей системе формуляров наблюдений, осуществляемых данным органом статистики. В адресной части предусматривается запись точного адреса единицы или совокупности единиц наблюдения и ряд других сведений.

Однако, насколько четко не был бы составлен формуляр, к нему обычно составляется **инструкция**, которая включает совокупность разъяснений и указаний, главным образом по программе статистического наблюдения. Инструкция может быть представлена в виде отдельного документа (часто — брошюры) или, изложена на формуляре наблюдения. Инструкцию следует писать кратко, просто, пояснения и указания должны быть ясными и четкими.

При организации статистического наблюдения необходимо решить вопрос о времени проведения данного наблюдения, включая выбор сезона, установления срока (периода) наблюдения, а в некоторых случаях и так называемого критического момента.

Период наблюдения — это время, в течение которого осуществляется регистрация признаков у единиц наблюдения по установленной программе. Продолжительность периода наблюдения зависит от многих факторов, среди которых можно выделить: размер и состояние объекта наблюдения, объем и сложность программы наблюдения.

Для наиболее подвижных объектов изучения, таких как население, например, устанавливается критический момент статистического наблюдения. **Критическим моментом** называется момент времени, по состоянию на который производится регистрация собираемых сведений. На практике критический момент назначается на начало периода наблюдения.

В целях успешного проведения наблюдения разрабатываются организационные вопросы плана статистического наблюдения, которые фиксируются в **организационном плане**. Организационный план предполагает решение следующих вопросов:

– объект наблюдения (дается его определение, описание, указываются отличительные признаки);

– цели и задачи наблюдения;

– орган наблюдения, осуществляющий подготовку и проведение наблюдения;

– место и сроки наблюдения;

– подготовительные работы к наблюдению включающие в себя подбор и обучение кадров, составление списков единиц изучаемой совокупности, в некоторых случаях эти работы включают рекламную компанию проводимого наблюдения и т.д.;

– порядок проведения наблюдения;

– порядок приема и сдачи материалов наблюдения и представления предварительных и окончательных итогов наблюдения;

– финансирование и материально-техническое обеспечение работ.

2.3. Точность статистического наблюдения

Под точностью в статистике понимают степень соответствия данных, полученных в результате статистического наблюдения, реальным их значениям. Возникающие расхождения между данными статистического наблюдения и реальными значениями признака называются ошибками. Ошибки определяются как разность или как отношение между этими значениями. Как правило, ошибки возникаю в результате следующих причин: ошибки при регистрации, ошибки при измерении. Следует отметить, что ошибки наблюдения наиболее опасны, поскольку их достаточно тяжело исправить, и они оказывают огромное влияние на дальнейшие расчеты.

В статистике выделяют ошибки регистрации и ошибки репрезентативности.

Ошибки регистрации возникают вследствие неправильного установления фактов в процессе наблюдения, или ошибочной их записи, или того и другого вместе. Ошибки регистрации могут иметь место как при сплошном, так и при несплошном наблюдении. При несплошном наблюдении возникаю так называемые **ошибки репрезентативности**, или как их еще называют ошибки представительности. Они заключаются в том, что значения признаков по отобранной выборочной совокупности не отражают реально существующей картины.

В зависимости от характера ошибки наблюдения бывают случайными и систематическими.

Случайные ошибки возникают случайным образом, в результате опечаток, описок, оговорок и т.п. Например, при регистрации регистратор при записи даты рождения вместо 15 июня написал 15 июля. При достаточно большом числе наблюдений благодаря действию закона больших чисел эти ошибки более или менее взаимно погашаются.

Систематические ошибки наиболее опасны, поскольку действуют только в одном направлении и приводят к сильному искажению данных. Наиболее показательной систематической ошибкой являются ошибки при переписи населения, которые заключаются в том, что населению свойственно округлять свой возраст на цифры оканчивающиеся на 5 или 0. К этому же виду ошибок можно отнести сокрытие реальных размеров финансовых результатов производственно-хозяйственной деятельности экономическими субъектами, стремление респондентов указать заниженное значение своего возраста и т.п..

С целью выявления ошибок проводят контроль полученных материалов. С этой целью после проведения наблюдения весь собранный материал проверяют на полноту охвата объекта наблюдением и на качество заполнения формуляров и других документов наблюдения. В последнем случае используют два вида контроля: логический и арифметический.

При контроле полноты охвата объекта наблюдения устанавливается, от всех ли единиц совокупности, подлежащих наблюдению, получены данные. Если выявлена неполнота охвата объекта наблюдением, дальнейшие действия зависят от того, представляется возможным восполнение пробелов или нет.

Логический контроль состоит в сопоставлении между собой ответов на вопросы формуляра наблюдения и выяснения их логической совместимости. При обнаружении несовместимых ответов пытаются путем дальнейших сопоставлений с ответами на другие вопросы или каким-либо иным путем установить, какой из ответов является неправильным.

Арифметический контроль состоит в проверке различных расчетов, результаты которых проведены в формуляре наблюдения, в частности, итогов, вычисления процентов, расчетов средних величин и т.п.

Глава 3. Статистическая сводка и группировка

3.1. Задачи сводки и ее содержание

Важнейшим этапом исследования социально-экономических явлений и процессов является систематизация первичных данных и получение на этой основе сводной характеристики всего объекта при помощи обобщающих показателей, что достигается путем сводки и группировки первичного статистического материала.

Сводка - это научная обработка первичных данных с целью получения обобщенных характеристик изучаемого социально-экономического явления по ряду существенных для него признаков с целью выявления типичных черт и закономерностей, присущих изучаемому явлению в целом.

По глубине и точности обработки материала различают сводку простую и сложную.

Простая сводка - это операция по подсчету общих итогов по совокупности единиц наблюдения и оформление этого материала в статистических таблицах.

Сложная сводка - это комплекс последовательных операций, включающих группировку полученных при наблюдении материалов, составление системы показателей для характеристики типичных групп и подгрупп изучаемой совокупности явлений, подсчет числа единиц и итогов по каждой группе и подгруппе, и по всему объекту и представление результатов в виде статистических таблиц.

Проведение сводки включает следующие этапы:

– выбор группировочного признака;

– определение порядка формирования групп;

– разработка системы статистических показателей для характеристики групп и объекта в целом;

– разработка макетов статистических таблиц для представления результатов сводки.

По форме обработки материала сводка бывает:

– централизованная, когда весь первичный материал поступает в одну организацию, подвергается в ней обработке от начала до конца;

– децентрализованная, когда отчеты предприятий сводятся статистическими органами субъектов РФ, а полученные итоги поступают в Госкомстат РФ и там определяются итоговые показатели в целом по народному хозяйству страны.

3.2. Виды статистических группировок

Группировкой называется разбиение общей совокупности единиц объекта наблюдения по одному или нескольким существенным признакам на однородные группы, различающиеся между собой в количественном и качественном отношении и позволяющие выделить социально-экономические типы, изучить структуру совокупности и проанализировать связи между отдельными признаками. Группировки являются важнейшим статистическим методом обобщения статистических данных, основой для правильного исчисления статистических показателей.

С помощью метода группировок решаются следующие задачи:

- выделение социально-экономических типов явлений;
- изучение структуры явления и структурных сдвигов, происходящих в нем;
- выявление взаимосвязи и взаимозависимости между явлениями.

В соответствии с познавательными задачами, решаемыми в ходе построения статистических группировок, различают следующие их виды: типологические, структурные, аналитические.

Типологическая группировка - это разбиение разнородной совокупности единиц наблюдения на отдельные качественно однородные группы и выявление на этой основе социально-экономических типов явлений. При построении группировки этого вида основное внимание должно быть уделено идентификации типов и выбору группировочного признака. Решение вопроса об основании группировки должно осуществляться на основе анализа сущности изучаемого социально-экономического явления.

Структурной называется группировка, которая предназначена для изучения состава однородной совокупности по какому-либо варьирующему признаку, а также структуры и структурных сдвигов, происходящих в нем.

Группировка, выявляющая взаимосвязи между изучаемыми явлениями и признаками, их характеризующими, называется **аналитической** группировкой.

В статистике при изучении связей социально-экономических явлений признаки необходимо делить на факторные и результативные.

Факторными называются признаки, под воздействием которых изменяются другие **результативные** признаки. Взаимосвязь проявляется в том, что с возрастанием или убыванием значения факторного признака систематически возрастает или убывает значение признака результативного и наоборот.

Особенностями построения аналитической группировки являются:

- единицы статистической совокупности группируются по факторному признаку;

- каждая выделенная группа характеризуется средними величинами результативного признака.

По способу построения группировки бывают простые и комбинационные.

Простой называется группировка, в которой группы образованы только по одному признаку.

Комбинационной называется группировка, в которой разбиение совокупности на группы производится по двум и более признакам, взятым в сочетании (комбинации).

Сначала группы формируются по одному признаку, затем группы делятся на подгруппы по другому признаку, а эти в свою очередь делятся по третьему и так далее. Таким образом, комбинационные группировки дают возможность изучить единицы совокупности одновременно по нескольким взаимосвязанным признакам.

При построении комбинационной группировки возникает вопрос о последовательности разбиения единиц объекта по признакам. Как правило, рекомендуется сначала производить группировку по атрибутивным признакам, значения которых имеют ярко выраженные качественные различия.

3.3. Принципы построения статистических группировок и классификаций

Построение статистических группировок осуществляется по следующим этапам:

1. Определение группировочного признака.
2. Определение числа групп.
3. Расчет ширины интервала группировки.
4. Определение признаков, которые в комбинации друг с другом будут характеризовать каждую выделенную группу.

Построение группировки начинается с определения группировочного признака.

Группировочным признаком называется признак, по которому проводится разбиение единиц совокупности на отдельные группы. От правильного выбора группировочного признака зависят выводы статистического исследования. В качестве основания группировки необходимо использовать существенные, теоретически обоснованные признаки.

В основание группировки могут быть положены как количественные, так и качественные признаки. **Количественные признаки -** это признаки, которые имеют числовое выражение (объем выпускаемой продукции, возраст человека, доход сотрудника фирмы и т. д.). **Качественные признаки** отражают состояние единицы совокупности (пол, отраслевая принадлежность предприятия, форма собственности фирмы и т. д.).

После того, как определено основание группировки следует решить вопрос о количестве групп, на которые необходимо разбить исследуемую совокупность единиц наблюдения.

Число групп зависит от задач исследования и вида показателя, положенного в основание группировки, объема изучаемой совокупности и степени вариации признака. Вид показателя особенно существенен при анализе качественных признаков. Так, например, группировка сотрудников фирмы по полу учитывает только две градации: «мужской» и «женский».

В случае группировки единиц наблюдения по количественному признаку особое внимание необходимо обратить на число единиц исследуемого объекта, объем совокупности и степень колеблемости группировочного признака.

При небольшом объеме совокупности ($n<50$) не следует образовывать большого количества групп, так как группы будут включать недостаточное число единиц объекта. Показатели, рассчитанные для таких групп, не будут представительными и не позволят получить адекватную характеристику исследуемого явления.

Часто группировка по количественному признаку имеет задачу отразить распределение единиц совокупности по этому признаку. В этом случае количество групп зависит, в первую очередь, от степени колеблемости группировочного признака: чем больше его колеблемость, тем больше можно образовать групп. Поэтому при определении числа групп необходимо принять во внимание размах вариации признака (R), который позволяет оценить вариацию признака между крайними значениями признака – максимальным (Xmax) и минимальным (Xmin) и определяется по следующей формуле:

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

Чем больше размах вариации признака, положенного в основание группировки, тем, как правило, может быть образовано большее число групп. При этом может возникнуть проблема получения пустых групп, т.е. групп, не содержащих ни одной единицы наблюдения.

Построение большого числа групп позволит, с одной стороны, точнее воспроизвести характер исследуемого объекта. Однако, с другой стороны, слишком большое число групп затрудняет выявление закономерностей при исследовании социально-экономических явлений и процессов. Поэтому в каждом конкретном случае при определении числа групп следует исходить не только из степени колеблемости признака, но и из особенностей объекта и показателей, его характеризующих а также цели исследования.

Определение числа групп можно осуществить несколькими способами. Формально-математический способ предполагает использование формулы Стерджесса:

$$n = 1 + 3{,}322 \times \lg N, \qquad (3.1)$$

где n - число групп

N - число единиц совокупности.

Согласно этой формуле выбор числа групп зависит только от объема изучаемой совокупности.

Применение данной формулы дает хорошие результаты в том случае, если совокупность состоит из большого числа единиц наблюдения ($n>50$).

Другой способ определения числа групп основан на применении показателя среднего квадратического отклонения (σ). Если величина интервала равна $0{,}5\sigma$, то совокупность разбивается на 12 групп, а когда величина интервала равна $2/3\sigma$ и σ, то совокупность делится, собственно, на 9 и 6 групп. Однако, при определении групп данными методами существует большая вероятность получения «пустых» или малочисленных групп, характеристики изучаемого явления на основе которых будут недостаточно типичными для выделенной группы и изучаемой совокупности в целом.

Когда определено число групп, то следует определить интервалы группировки.

Интервал - это значения варьирующего признака, лежащие в определенных границах. Каждый интервал имеет верхнюю и нижнюю границы или одну из них. **Нижней границей** интервала называется наименьшее значение признака в интервале. **Верхней границей интервала** называется наибольшее значение признака в интервале. Величина интервала представляет собой разность между верхней и нижней границами интервала.

Интервалы группировки бывают:

- равные и неравные;
- открытые и закрытые.

В зависимости от величины интервалы группировки бывают: равные и неравные. В свою очередь неравные интервалы подразделяются на прогрессивно возрастающие, прогрессивно убывающие, произвольные и специализированные.

Равные интервалы применяются в случае, если изменение количественного признака внутри изучаемой совокупности единиц наблюдения происходит равномерно и его вариация проявляется в сравнительно узких границах.

Ширина равного интервала определяется по следующей формуле:

$$h = \frac{R}{n} = \frac{x_{\max} - x_{\min}}{n} \qquad (3.2)$$

где x_{max}, x_{min} - максимальное и минимальное значения признака в совокупности;

n -число групп.

Если максимальные или минимальные значения сильно отличаются от смежных с ними значений вариантов в упорядоченном ряду значений группировочного признака, то для определения величины интервала следует использовать не максимальное или минимальное значения, а значения, несколько превышающие минимум, и несколько меньше, чем максимум.

Полученную по формуле (3.2) величину округляют и она будет являться шириной интервала.

Существуют следующие правила определения ширины интервала.

Если величина интервала, рассчитанная по формуле (3.2) представляет собой величину, которая имеет один знак до запятой (например: 0,67; 1,487; 3,82), то в этом случае полученные значения целесообразно округлить до десятых и их использовать в качестве ширины интервала. В приведенном выше примере это будут соответственно значения: 0,7; 1,5; 3,8.

Если рассчитанная величина интервала имеет две значащие цифры до запятой и несколько после запятой (например 14,876), то это значение необходимо округлить до целого числа (до 15).

В случае, когда рассчитанная величина интервала представляет собой трехзначное, четырехзначное и так далее число, то эту величину следует округлить до ближайшего числа, кратного 100 или 50. Например, 652 следует округлить до 650 или до 700.

Если размах вариации признака в совокупности велик и значения признака варьируют неравномерно, то надо использовать группировку с неравными интервалами. Неравные интервалы могут быть получены в процессе объединения пустых, не содержащих ни одной единицы совокупности, равных интервалов. Это происходит в том случае, если после построения равных интервалов по изучаемому признаку образуются группы, содержащие мало или не содержащие вообще ни одной единицы, т.е. группы, не отражающие определенных типов изучаемого явления по признаку. В этом случае возникает необходимость в увеличении интервалов группировки.

Также неравные интервалы могут быть прогрессивно возрастающие или прогрессивно убывающие в арифметической или геометрической прогрессии. Величина интервалов, изменяющихся в арифметической и геометрической прогрессии определяются следующим образом:

$$hi+1 = hi + a,$$

а в геометрической прогрессии:

$$hi+1 = hi \times q,$$

где a - константа: для прогрессивно-возрастающих интервалов имеет знак «+«, а при прогрессивно-убывающих - знак «-«.

q - константа: для прогрессивно – возрастающих – больше «1»; для прогрессивно-убывающих – меньше «1».

Применение неравных интервалов обусловлено тем, что в первых группах небольшая разница в показателях имеет большое значение, а в последних группах эта разница не существенна.

Например, при построении группировки строительных компаний города по показателю численности работающих, который варьирует от 500 человек до 3500 человек, нецелесообразно рассматривать равные интервалы, т. к. учитываются как малые, так и крупнейшие строительные фирмы города. Поэтому следует образовывать неравные интервалы: 500-1000, 1000-2000, 2000-3500, т. е. величина каждого последующего интервала больше предыдущего на 500 человек и увеличивается в арифметической прогрессии. Выбор исследователя в построении равных или неравных интервалов зависит от степени заполнения каждой выделенной группы, т.е. от числа единиц в них. Если величина интервала существенна и содержит большое число единиц совокупности, то эти интервалы необходимо дробить, а в противном случае – объединять.

Интервалы группировок могут быть закрытыми и открытыми.

Закрытыми называются интервалы, у которых имеются обе границы: верхняя и нижняя границы.

Открытые - это интервалы, у которых указана только одна граница: как правило, верхняя - у первого интервала и нижняя - у последнего. Например, группы страховых компаний по числу работающих в них сотрудников (чел.): до 50, 50-100, 100-150, 150 и более. Применение открытых интервалов целесообразно в тех случаях, когда в совокупности встречается незначительное число единиц наблюдения с очень малыми или очень большими значениями вариантов, которые резко, в несколько раз, отличаются от всех остальных значений изучаемого признака.

При группировке единиц совокупности по количественному признаку границы интервалов могут быть обозначены по-разному, в зависимости от того, непрерывный или дискретный признак положен в основание группировки.

Если основанием группировки служит непрерывный признак (например, группы строительных фирм по объему строительно-монтажных работ, выполненных собственными силами (тыс. руб.): 1200-1400, 1400-1600, 1600-1800, 1800-2000), то одно и то же значение признака выступает и верхней и нижней границами двух смежных интервалов. В данном случае объем работ 1400 тыс. руб. составляет верхнюю границу первого интервала и нижнюю границу второго, 1600 тыс. руб. - соответственно второго и третьего и т. д.,т. е. верхняя граница i - го интервала равна нижней границе (i+1) - го интервала.

При таком обозначении границ может возникнуть вопрос, в какую группу включать единицы наблюдения, значения признака у которых совпадают с границами интервалов. Например, во вторую или третью группу должна войти строительная фирма с объемом строительно-монтажных работ 1600 тыс. рублей? Если верхняя граница формируется по принципу «исключительно», то фирма должна быть отнесена к третьей группе, в противном случае - ко второй. Для того, чтобы правильно отнести к той или иной группе единицу совокупности, значение признака которой совпадает с границами интервалов, можно ориентироваться на открытые интервалы (по нашему примеру группы строительных фирм по объему строительно-монтажных работ преобразуются в следующие: до 1400, 1400-1600, 1600-1800, 1800 и более). В данном случае, вопрос отнесения отдельных единиц совокупности, значения которых являются граничными, к той или иной группе решается на основе анализа последнего открытого интервала. Возможны два случая обозначения последнего открытого интервала: 1).1800 тыс. руб. и более; 2). более 1800 тыс. руб. В первом случае, строительные фирмы с объемом строительно-монтажных работ 1600 тыс. руб. попадут в третью группу; во втором случае - во вторую группу.

Если в основании группировки лежит дискретный признак, то нижняя граница i-го интервала равна верхней границе i-1-го интервала, увеличенной на 1. Например, группы строительных фирм по числу занятого персонала (чел.) будут иметь вид: 100-150, 151-200, 201-300.

При определении границ интервалов статистических группировок иногда исходят из того, что изменение количественного признака приводит к появлению нового качества. В этом случае граница интервала устанавливается там, где происходит переход от одного качества к другому.

Строя такую группировку, следует дифференцированно устанавливать границы интервалов для разных отраслей народного хозяйства. Это достигается путем использования группировок со специализированными интервалами. **Специализированные интервалы -** это такие интервалы, которые применяются для выделения из совокупности одних и тех же типов по одному и тому же признаку для явлений, находящихся в различных условиях.

При изучении социально-экономических явлений на макроуровне часто применяют группировки, интервалы которых не будут ни прогрессивно возрастающими, ни прогрессивно убывающими. Такие интервалы называются **произвольными** и, как правило, используются при группировке предприятий, например, по уровню рентабельности.

Пример.

Произведем группировку совокупности, включающей 30 банков Российской Федерации (на 01.01.04 г.):

Номер банка	Капитал, млн. руб.	Рабочие активы, млн. руб.	Уставный фонд, млн. руб.
1	207,7	2,48	1,14
2	200,3	2,40	1,10
3	190,2	2,28	1,05
4	323,0	3,88	1,88
5	247,1	2,96	1,36
6	177,7	2,12	0,97
7	242,5	2,90	1,33
8	182,9	2,18	0,99
9	315,6	3,78	1,73
10	183,2	2,20	1,01
11	320,2	3,84	1,76
12	207,3	2,48	1,14
13	181,0	2,17	0,99
14	172,4	2,06	0,94
15	234,3	2,81	1,29
16	189,5	2,27	1,04
17	187,8	2,24	1,03
18	166,9	1,99	0,91
19	157,7	1,88	0,86
20	168,3	2,02	0,93
21	224,4	2,69	1,23
22	166,5	1,99	0,91
23	198,5	2,38	1,09
24	240,4	2,88	1,32
25	229,3	2,75	1,26
26	175,2	2,10	0,96
27	156,8	1,87	0,86
28	160,1	1,92	0,88
29	178,7	2,14	0,98
30	171,6	2,05	0,94

В качестве группировочного признака возьмем капитал банка. Образуем четыре группы банков с равными интервалами. Величину интервала определим по формуле:

$$h = \frac{X\max - X\min}{n} = \frac{323{,}0 - 156{,}0}{4} = 41{,}8$$

Обозначим границы групп:

1-я группа – 156,0-197,8

2-я группа – 197,8-239,6

3-я группа – 239,6-281,4

4-я группа – 281,4-323,2

После того как определен группировочный признак - капитал, задано число групп - 4 и образованы сами группы, необходимо отобрать показатели, которые характеризуют группы и определить их величины по каждой группе. Показатели, характеризующие банки, разносятся по четырем указанным группам и подсчитываются групповые итоги. Результаты группировки заносятся в таблицу и определяются общие итоги по совокупности единиц наблюдения по каждому показателю.

Таблица 3.1.

Группировка коммерческих банков по величине капитала

Группы банков по величине капитала, млн. руб.	Число банков	Капитал, млн. руб.	Активы, млн. руб.	Работающие активы, млн. руб.
А	1	2	3	4
156,0-197,8	17	2966,5	35,48	16,25
197,8-239,6	7	1501,8	17,99	8,25
239,6-281,4	3	730,0	8,74	4,01
281,4-323,2	3	958,8	11,5	5,37
Итого	30	6157,1	73,71	33,88

Структурная группировка коммерческих банков на основе данных таблицы 3.1 будет иметь вид:

Таблица 3.2.

Группировка коммерческих банков по величине капитала (в %% к итогу)

Группы банков по величине капитала, млн. руб.	Число банков	Капитал	Активы	Работающие активы
156,0-197,8	56,7	48,2	48,1	48,0
197,8-239,6	23,3	24,4	24,4	24,3
239,6-281,4	10,0	11,9	11,9	11,8
281,4-323,2	10,0	15,5	15,6	15,9
Итого	100,0	100,0	100,0	100,0

Из таблицы 3.2 видно, что в основном преобладают малые банки – 56,7%, на долю которых приходится 48,2% всего капитала. Более конкретный анализ взаимосвязи показателей можно сделать на основе аналитической группировки.

Таблица 3.3.

Группировка коммерческих банков по величине капитала

Группы банков по величине капитала, млн. руб.	Число банков	Капитал, млн.руб.		Активы, млн. руб.		Работающие активы, млн.руб.	
		всего	в среднем на один банк	всего	в среднем на один банк	всего	в среднем на один банк
156,0-197,8	17	2966,5	174,5	35,48	2,09	16,25	0,96
197,8-239,6	7	1501,8	214,5	17,99	2,57	8,25	1,18
239,6-281,4	3	730,0	243,3	8,74	2,91	4,01	1,34
281,4-323,2	3	958,8	319,6	11,5	3,83	5,37	1,79
Итого	30	6157,1	205,2	73,71	2,46	33,88	1,13

Величина капитала, все активы банка и работающие активы прямо зависят между собой, и чем крупнее банк, тем эффективнее управление работающими активами.

Мы рассмотрели примеры группировок по одному признаку. Однако в ряде случаев для решения поставленных задач такая группировка является недостаточной. В этих случаях переходят к группировке исследуемой совокупности по двум и более существенным признакам во взаимосвязи (комбинационной группировке).

Произведем группировку данных коммерческих банков по двум признакам: величине капитала и работающим активам.

Каждую группу и подгруппу охарактеризуем следующими показателями: число коммерческих банков, капитал, работающие активы.

Таблица 3.4.

Группировка коммерческих банков по величине капитала и работающим активам

Номер группы	Группы банков по величине капитала, млн. руб.	Подгруппы по величине работающих активов, млн. руб.	Число бан-ков	Капи-тал, млн. руб.	Работа-ющие активы, млн. руб.
1	**2**	**3**	**4**	**5**	**6**
1	156,0-197,8	0,86-1,37 1,37-1,88	17 -	2966,5 -	16,25 -
	Итого		17	2966,5	16,25
2	197,8-239,6	0,86-1,37 1,37-1,88	7 -	1501,8 -	8,25 -
	Итого		7	1501,8	8,25
3	239,6-281,4	0,86-1,37 1,37-1,88	3 -	730,0 -	4,01 -
	Итого		3	730,0	4,01
4	281,4-323,2	0,86-1,37 1,37-1,88	- 3	- 958,8	- 5,37
	Итого		3	958,8	5,37
5	Всего по подгруппам	0,86-1,37 1,37-1,88	27 3	5198,3 958,8	28,51 5,37
	Всего		30	6157,1	33,88

От группировок следует отличать классификацию. **Классификацией** называется систематизированное распределение явлений и объектов на определенные группы, классы, разряды на основании их сходства и различия.

Отличительными чертами классификаций является:

- в их основе лежит качественный признак;
- классификации стандартны и устанавливаются органами государственной и международной статистики;
- классификации устойчивы, так как остаются неизменными в течение длительного периода времени.

Ряды распределения представляют собой простейшую группировку, в которой каждая выделенная группа характеризуется только частотой.

В зависимости от признака, положенного в основу образования ряда распределения, различают атрибутивные и вариационные ряды распределения.

Атрибутивными называют ряды распределения, построенные по качественным признакам, то есть признакам, характеризующим состояние изучаемого явления и не имеющим числового выражения.

Атрибутивные ряды распределения характеризуют состав совокупности по тем или иным существенным признакам. Взятые за несколько периодов, эти данные позволяют исследовать изменение структуры.

Вариационными рядами называют ряды распределения, построенные по количественному признаку, т.е. признаку, имеющему числовое выражение у отдельных единиц совокупности. Вариационный ряд состоит из двух элементов: вариантов и частот. **Вариантами** называются отдельные значения признака, которые он принимает в вариационном ряду, то есть конкретное значение варьирующего признака. **Частотами** называются численности отдельных вариант или каждой группы вариационного ряда. Частоты показывают, как часто встречаются те или иные значения признака в изучаемой совокупности. Сумма всех частот определяет численность всей совокупности, ее объем. Частостями называются частоты, выраженные в долях единицы или в процентах к итогу. Соответственно сумма частостей равна 1 или 100%.

В зависимости от характера вариации признака различают дискретные и интервальные вариационные ряды.

Дискретный вариационный ряд - это ряд распределения в котором группы составлены по признаку, изменяющемуся прерывно, т.е. через определенное число единиц и характеризуют распределение единиц совокупности по дискретному признаку, принимающему только целые значения. Например, группы студентов по баллу в сессию по предмету: 5,4,3,2.

Интервальный вариационный ряд распределения – это ряд распределения, в котором группировочный признак, составляющий основание группировки, может принимать в интервале любые значения, отличающиеся друг от друга на сколь угодную малую величину.

Построение **интервальных вариационных рядов** целесообразно прежде всего при непрерывной вариации признака, а также если дискретная вариация признака проявляется в широких пределах, то есть число вариантов дискретного признака достаточно велико.

Правила построения рядов распределения аналогичны правилам построения группировки.

Анализ рядов распределения наглядно можно проводить на основе их графического изображения. Для этой цели строят полигон, гистограмму, огиву и кумуляту распределения.

Полигон используется при изображении дискретных вариационных рядов. Для его построения в прямоугольной системе координат по оси абсцисс в одинаковом масштабе откладываются ранжированные значения варьирующего признака, а по оси ординат наносится шкала для выражения величины частот. Полученные на пересечении оси абсцисс (х) и оси ординат (у) точки соединяются прямыми линиями, в результате чего получают ломаную линию, называемую полигоном частот. Иногда для замыкания полигона предлагается крайние точки (слева и справа на ломаной линии) соединить с точками на оси абсцисс, в результате чего получается многоугольник.

Гистограмма применяется для изображения интервального вариационного ряда. При построении гистограммы на оси абсцисс откладываются величины интервалов, а частоты изображаются прямоугольниками, построенным на соответствующих интервалах. Высота столбиков должна быть пропорциональна частотам. В результате получается график, на котором ряд распределения изображен в виде смежных друг с другом столбиков.

Гистограмма может быть преобразована в полигон распределения, если середины верхних сторон прямоугольников соединить прямыми линиями.

При построении гистограммы распределения вариационного ряда с неравными интервалами по оси ординат наносят не частоты, а плотность распределения признака в соответствующих интервалах. Это необходимо сделать для устранения влияния величины интервала на распределение интервала и получения возможности сравнивать частоты. **Плотность распределения** - это частота, рассчитанная на единицу ширины интервала, то есть, сколько единиц в каждой группе приходится на единицу величины интервала.

Для графического изображения вариационных рядов может использоваться кумулятивная кривая. При помощи **кумуляты** изображается ряд накопленных частот. Накопленные частоты определяются путем последовательного суммирования частот по группам. Накопленные частоты показывают, сколько единиц совокупности имеют значения признака не больше, чем рассматриваемое значение.

При построении кумуляты интервального вариационного ряда по оси абсцисс (х) откладываются варианты ряда, а по оси ординат (у) накопленные частоты, которые наносят на поле графика в виде перпендикуляров к оси абсцисс в верхних границах интервалов. Затем эти перпендикуляры соединяют и получают ломаную линию, то есть кумуляту.

Если при графическом изображении вариационного ряда в виде кумуляты оси х и у поменять местами, то получим **огиву.**

3.4. Сравнимость статистических группировок. Вторичная группировка

Группировки, построенные за один и тот же период времени, но для разных объектов или, наоборот, для одного объекта, но за два разных периода времени могут оказаться несопоставимыми из-за различного числа выделенных групп или неодинаковости границ интервалов.

Вторичная группировка, или перегруппировка сгруппированных данных применяется для лучшей характеристики изучаемого явления (в случае, когда первоначальная группировка не позволяет четко выявить характер распределения единиц совокупности), либо для приведения к сопоставимому виду группировок с целью проведения сравнительного анализа.

Вторичная группировка - операция по образованию новых групп на основе ранее осуществленной группировки.

Применяют два способа образования новых групп. Первым, наиболее простым и распространенным способом является изменение (чаще укрупнение) первоначальных интервалов. Второй способ получил название долевой перегруппировки и состоит в образовании новых групп на основе закрепления за каждой группой определенной доли единиц совокупности. Проиллюстрируем методику вторичной группировки на следующем примере.

Пример:
Распределение сотрудников предприятия по уровню дохода

Группы работающих по уровню доходов, тыс.руб.	Число работающих, чел.
До 4	16
4-10	20
10-18	44
18-30	74
30-40	37
40 и более	9
Итого	200

Произведем перегруппировку данных, образовав новые группы с интервалами до 5, 5 - 10, 10 - 20, 20 - 30, свыше 30 тыс. руб.

В первую новую группу войдет полностью первая группа сотрудников и часть второй группы. Чтобы образовать группу до 5 тыс. руб., необходимо от интервала второй группы взять 1,0 тыс. руб. Величина интервала этой группы составляет 6,0 тыс. руб. Следовательно, необходимо взять от нее 1/6 (1,0:6,0) часть. Аналогичную же часть во вновь образуемую первую группу надо взять и от численности работающих, то есть $20 \times \frac{1}{6} = 3$ чел. Тогда в первой группе будет работающих: 16 + 3 = 19 чел.

Вторую новую группу образуют работающие второй группы за вычетом отнесенных к первой, то есть 20 - 3 = 17 чел. Во вновь образованную третью группу войдут все сотрудники третьей группы и часть сотрудников четвертой. Для определения этой части от интервала 18 - 30 (ширина интервала равна 12) нужно добавить к предыдущему 2,0 (чтобы верхняя граница интервала была равна 2,0 тыс. руб.). Следовательно, необходимо взять часть интервала, равную [2:12 = 1:6]. В этой группе 74 человека, значит надо взять 74×(1:6) = 12 чел. В новую третью группу войдут 44 +12 = 56 чел. Во вновь образованную четвертую группу войдут 74 - 12 = 62 чел., оставшихся от прежней четвертой группы. Пятую вновь образованную группу составят работающие пятой и шестой прежних групп: 37 +9 = 46 чел.

В результате получим следующие новые группы:

Группы работающих по уровню доходов, руб.	Число работающих
до 5	19
5-10	17
10-20	56
20-30	62
свыше 30	46
Итого	200

3.5. Статистическая таблица и ее элементы

Результаты сводки и группировки материалов статистического наблюдения, как правило, представляются в виде таблиц. Таблица является наиболее рациональной, наглядной и компактной формой представления статистического материала. Однако, не всякая таблица является статистической. Таблица умножения, опросный лист социологического обследования и так далее, могут носить табличную форму, но еще не являются статистическими таблицами.

Статистической называется таблица, которая содержит сводную числовую характеристику исследуемой совокупности по одному или нескольким существенным признакам, взаимосвязанным логикой экономического анализа.

Основные элементы статистической таблицы, составляющие как бы ее остов (основу), показаны на схеме 2.1.

Табличной называется такая форма расположения числовой информации, при которой число располагается на пересечении четко сформулированного заголовка по вертикальному столбцу, называемому графой, и названия по соответствующей горизонтальной полосе - строке.

Таким образом, внешне таблица представляет собой пересечение граф и строк, которые формируют остов таблицы.

Статистическая таблица содержит три вида заголовков: общий, верхние и боковые. Общий заголовок отражает содержание всей таблицы (к какому месту и времени она относится), располагается над макетом таблицы по центру и является внешним заголовком. Верхние заголовки характеризуют содержание граф (заголовки сказуемого), а боковые (заголовки подлежащего) - строк. Они являются внутренними заголовками.

Остов таблицы, заполненный заголовками, образует макет таблицы; если на пересечении граф и строк записать цифры, то получается полная статистическая таблица.

Название таблицы
(общий заголовок)

<table>
<tr><td>Содержание строк</td><td colspan="6">Наименование граф (верхние заголовки)</td></tr>
<tr><td>А</td><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td><td>...</td></tr>
<tr><td rowspan="3">Наименование строк (боковые заголовки)</td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>Итоговая строка</td><td></td><td></td><td></td><td></td><td></td><td>Итоговая графа</td></tr>
</table>

*) Примечания к таблице.

Схема 3.1. Остов (основа) статистической таблицы

Цифровой материал может быть представлен абсолютными (численность населения РФ), относительными (индексы цен на продовольственные товары) и средними (среднемесячный доход сотрудника коммерческого банка) величинами.

Таблицы могут сопровождаться примечанием, используемым с целью пояснения, в случае необходимости, заголовков, методики расчета некоторых показателей, источников информации и так далее.

По логическому содержанию таблица представляет собой «статистическое предложение», основными элементами которого являются подлежащее и сказуемое.

Подлежащим статистической таблицы называется объект, который характеризуется цифрами. Это может быть одна или несколько совокупностей, отдельные единицы совокупности в порядке их перечня или сгруппированные по каким-либо признакам, территориальные единицы и так далее. Обычно подлежащее таблицы дается в левой части, в наименовании строк.

Сказуемое статистической таблицы образует система показателей, которыми характеризуется объект изучения, то есть подлежащее таблицы. Сказуемое формирует верхние заголовки и составляет содержание граф с логически последовательным расположением показателей слева направо.

Расположение подлежащего и сказуемого в отдельных случаях может меняться местами для более полного и лучшего способа прочтения и анализа исходной информации об исследуемой совокупности.

3.6. Виды статистических таблиц

В практике экономико-статистического анализа используются различные виды статистических таблиц.

В зависимости от структуры подлежащего, от группировки единиц в нем, различают статистические таблицы простые и сложные, а последние, в свою очередь, подразделяются на групповые и комбинационные.

Простой называется такая таблица, в подлежащем которой дается перечень каких-либо объектов или территориальных единиц.

Простые таблицы различают монографические и перечневые. Монографические таблицы характеризуют не всю совокупность единиц изучаемого объекта, а только одну какую-либо группу из нее, выделенную по определенному признаку (табл. 3.5).

Таблица 3.5.

Ввод в действие зданий жилого назначения в Российской Федерации в 2003 г.

	Число зданий, ед.	Общий строительный объем зданий млн. куб.м.	Общая площадь зданий, млн. кв.м.
А	**1**	**2**	**3**
Введено в действие зданий жилого назначения	119727	150,1	40,5

Перестроив подлежащее таблицы 3.5, таким образом, чтобы были показаны все введенные в действие здания, то есть, показав каждую единицу совокупности, получаем перечневую таблицу (см. табл. 3.6.).

Таким образом, простыми перечневыми таблицами называются таблицы, подлежащее которых содержит перечень единиц изучаемого объекта.

Таблица 3.6.

Ввод в действие зданий в Российской Федерации в 2003 г.

Номер здания	Общий строительный объем, куб.м.	Общая площадь, кв.м.
А	**2**	**3**
1	…	…
2	…	…
…	…	…
…	…	…
Всего	…	…

Простые таблицы не дают возможности выявить социально-экономические типы изучаемых явлений, их структуру, а также взаимосвязи и взаимозависимости между характеризующими их признаками.

Эти задачи более полно могут быть решены с помощью сложных - групповых и, особенно, комбинационных таблиц.

Групповыми называются статистические таблицы, подлежащее которых содержит группировку единиц совокупности по одному количественному или атрибутивному признаку.

Простейшим видом групповых таблиц являются ряды распределения. Групповая таблица может быть более сложной, если в сказуемом дополнительно приводятся ряд показателей, характеризующих группы подлежащего. Такие таблицы часто используются в целях сопоставления обобщающих показателей по группам.

Таблица 3.7.

Распределение населения Российской Федерации по возрастным группам в 2002 г. (на начало года)

Группы населения по возрасту, лет	Численность населения, тыс. чел.	Численность населения, в % к итогу
1	2	3
0-4	6306	4,4
5-9	7123	4,9
10-14	10825	7,5
15-19	12208	8,5
20-24	10901	7,6
25-29	10422	7,2
30-34	9534	6,6
35-39	10588	7,4
40-44	12595	8,7
45-49	11625	8,1
50-54	9832	6,8
55-59	4841	3,4
60-64	8625	6,0
65-69	5974	4,2
70 и более	12555	8,7
Итого	143954	100,0

Таблица 3.7. отражает количественное распределение населения Российской Федерации по возрасту.

Таким образом, групповые таблицы позволяют выявить и охарактеризовать социально-экономические типы явлений, их структуру в зависимости только от одного признака.

Комбинационными называются статистические таблицы, подлежащее которых содержит группировку единиц совокупности одновременно по двум и более признакам: каждая из групп, построенная по одному признаку, разбивается, в свою очередь, на подгруппы по какому-либо другому признаку и так далее.

Таблица 3.8.

Группировка предприятий пищевой промышленности одного из регионов Российской Федерации по величине прибыли и численности промышленно-производственного персонала в 2003 г.

Группы предприятий по величине прибыли, млн. руб.	Группы предприятий по численности промышленно-производственного персонала (чел.)	Число предприятий
1	2	3
50-100	200-250 250-300 300-350	3 4 8
Итого по группе	-	15
100-150	200-250 250-300 300-350	1 2 2
Итого по группе	-	5
Итого по подгруппам	200-250 250-300 300-350	4 6 10
Всего		20

Подлежащим в таблице являются группы предприятий по величине прибыли и численности промышленно-производственного персонала.

Комбинационные таблицы позволяют характеризовать типические группы, выделенные по нескольким признакам и связь между ними. Последовательность разбиения единиц совокупности на однородные группы по признакам определяется либо важностью одного из них в их комбинации, либо порядком их изучения.

В сказуемом статистической таблицы, как уже говорилось, приводятся показатели, которые являются характеристикой изучаемого объекта.

По структурному строению сказуемого различают статистические таблицы с простой и сложной его разработкой.

При **простой разработке сказуемого**, показатель, определяющий его, не подразделяется на подгруппы и итоговые значения получаются путем простого суммирования значений по каждому признаку отдельно, независимо друг от друга. Примером простой разработки сказуемого может служить следующий фрагмент статистической таблицы:

Таблица 3.9.

Распределение строительных организаций различных форм собственности по объему работ, выполненных по договорам строительного подряда в 2003 г.

Строитель-ные органи-зации	Объем работ, выполненных по договорам строительного подряда - всего	в том числе по формам собственности				
		госу-дарст-венная	муни-ци-пальная	частная	смешан-ная россий-ская	прочие

После заполнения данного фрагмента таблицы получается подробная характеристика строительных организаций по структуре объема работ по формам собственности. По каждой строительной организации можно получить информацию об объеме работ, выполненных по договорам строительного подряда, как в целом, так и в разрезе форм собственности.

Сложная разработка сказуемого предполагает деление признака, формирующего его, на подгруппы:

Предп-риятия	Приобре-тено акций (всего)	в том числе			
		На льготных условиях		По цене определенной Госкомимуществом	
		привиле-гированные типа А	обыкно-венные	привиле-гированные типа А	обыкновенные

При этом получается более полная и подробная характеристика объекта.

Здесь оба признака сказуемого (ценовой и видовой) тесно связаны друг с другом. Можно проанализировать не только количество приобретенных акций по видам и условиям приобретения их сотрудниками приватизированных предприятий, но и определить число привилегированных и обыкновенных акций, приобретенных на разных ценовых условиях. То есть, при сложной разработке сказуемого явление или объект могут быть охарактеризованы различной комбинацией признаков, формирующих их.

Исследователь при построении статистических таблиц должен руководствоваться оптимальным соотношением показателей сказуемого.

3.7. Основные правила построения и анализа статистических таблиц

Статистические таблицы, как средство наглядного и компактного представления цифровой информации, должны быть статистически правильно оформлены.

Основными приемами, определяющими технику формирования статистических таблиц, являются следующие:

1. Таблица должна быть компактной и содержать только те данные, которые непосредственно отражают исследуемое явление в статике и динамике и необходимы для познания его сущности. Цифровой материал необходимо излагать таким образом, чтобы при анализе таблицы сущность явления раскрывалась чтением строк слева направо и сверху вниз;

2. Заголовок таблицы и названия граф и строк должны быть четкими, краткими, лаконичными, представлять собой законченное целое, органично вписывающееся в содержание текста. В названии таблицы должны найти отражение объект, признак, время и место совершения события. Например: «Курс доллара США на торгах ММВБ на 01.01.2004 г.» Названия таблицы, граф и строк пишутся полностью, без сокращений.

3. Информация, располагаемая в столбцах (графах) таблицы, завершается итоговой строкой. Существуют различные способы соединения слагаемых граф с их итогом:

- строка «Итого» или «Всего» завершает статистическую таблицу;
- итоговая строка располагается первой строкой таблицы и соединяется с совокупностью ее слагаемых словами «В том числе».

4. Если названия отдельных граф повторяются между собой, содержат повторяющиеся термины или несут единую смысловую нагрузку, то необходимо им присвоить объединяющий заголовок.

5. Графы и строки полезно нумеровать. Графы слева, заполненные названием строк, принято обозначать заглавными буквами алфавита (А), (В) и так далее, а все последующие графы - номерами в порядке возрастания.

6. Взаимосвязанные данные, характеризующие одну из сторон анализируемого явления (например, число коммерческих банков и удельный вес коммерческих банков (в % к итогу) и т.д.), целесообразно располагать в соседних друг с другом графах.

7. Графы и строки должны содержать единицы измерения, соответствующие поставленным в подлежащем и сказуемом показателям. При этом используются общепринятые сокращения единиц измерения (чел., руб., кВт/ч и так далее).

8. Числа целесообразно, по возможности, округлять. Округление чисел в пределах одной и той же графы или строки следует проводить с одинаковой степенью точности (до целого знака или до десятого и так далее).

Если все числа одной и той же графы или строки даны с одним десятичным знаком, а одно из чисел имеет точно два знака после запятой, то числа с одним знаком после запятой следует дополнять нулем, тем самым подчеркнув их одинаковую точность.

9. Отсутствие данных об анализируемом социально-экономическом явлении может быть обусловлено различными причинами и это по-разному отмечается:

а) если данная позиция (на пересечении соответствующих графы и строки) вообще не подлежит заполнению, то ставится знак «Х»;

б) если по какой-либо причине отсутствуют сведения, то ставится многоточие «...» или «нет свед.»;

в) если явление отсутствует полностью, то клетка заполняется тире (-);

г) для отображения очень малых чисел используют обозначения (0,0) или (0,00).

10. В случае необходимости дополнительной информации - разъяснений к таблице, могут даваться примечания.

Соблюдение приведенных правил построения и оформления статистических таблиц делает их основным средством представления, обработки и обобщения статистической информации о состоянии и развитии анализируемых социально-экономических явлений.

Анализу статистических таблиц предшествует этап ознакомления - чтения их.

«*Чтение*» предполагает, что исследователь, прочитав слова и числа таблицы, усвоил ее содержание в целом, сформулировал первые суждения об объекте, уяснил назначение таблицы, дал оценку явлению или процессу, описанному в таблице.

Анализ предполагает реализацию двух его направлений - структурного и содержательного.

Структурный анализ предполагает анализ строения таблицы и характеристику представленных в ней:

- совокупности и единиц наблюдения, формирующих ее;
- признаков и их комбинации, формирующих подлежащее и сказуемое таблицы;
- признаков - количественные или атрибутивные;
- соотношение признаков подлежащего с показателями сказуемого;
- вида таблицы - простая или сложная, а последняя - групповая или комбинационная;
- решаемых задач - анализ структуры, типов явлений или их взаимосвязей.

Содержательный анализ предполагает изучение внутреннего содержания таблицы: анализ отдельных групп подлежащего по соответствующим признакам сказуемого; выявление соотношений и пропорций между группами явлений по одному и разным признакам; сравнительный анализ и формулировка выводов по отдельным группам и по всей совокупности в целом, установление закономерностей и определение резервов развития изучаемого объекта.

Прежде чем приступить к анализу числовой информации, необходимо проверить ее достоверность и научную обоснованность, источники ее получения. Должна быть произведена проверка данных: логическая (например, абсурдно, если численность работающих на фирме составила 115,1 чел.) и счетная - выборочный расчет отдельных значений признаков по группе, либо итоговых значений.

Анализ отдельных признаков и групп необходимо начинать с изучения абсолютных величин, затем - связанных с ними относительных величин.

Анализ таблиц может быть дополнен расчетными средними величинами, графиками, диаграммами и т.д., если этого требуют задачи исследования.

Анализ данных таблиц производится по каждому признаку в отдельности, а затем в логико-экономическом сочетании признаков.

Соблюдение правил и последовательности работы со статистическими таблицами позволит исследователю осуществить научно-обоснованный экономико-статистический анализ объектов и процессов.

Глава 4. Графическое представление статистической информации

4.1. Роль и значение графического метода в статистике

В результате сводки и дальнейшей обработки данных отчетности, различного рода обследований, переписей, наблюдений и т.п. экономист получает большое количество различных статистических показателей, которые он располагает в виде таблиц. Применение табличного метода значительно облегчает ориентацию в материале. Однако из этого не следует, что можно ограничиться одними таблицами. Для того, чтобы сделать дальнейший шаг в понимании материала, надо от табличного метода перейти к графическому.

Графиком в статистике называется условные изображения статистических данных в виде различных геометрических образов: точек, линий, фигур и т.п. Главное достоинство графиков - наглядность.

В статистике графики используются, во–первых, в целях широкой популяризации данных и для облегчения их восприятия неспециалистами. Поэтому в различного рода докладах, речах и сообщениях представление статистических данных часто осуществляется при помощи графиков. Графики облегчают ознакомление масс со статистическими данными, оживляют таблицу, делают ее более доступной. Во–вторых, графики широко используются для обобщения и анализа статистических данных. Они находят большое применение в исследовательской работе. Именно при помощи графиков легче уяснить закономерности развития, распределения и размещения явлений. При помощи графиков в ряде случаев можно сделать выводы, которые на базе табличного метода были бы затруднительными. В–третьих, надо еще указать и на контрольное значение графиков. Под этим следует понимать тот факт, что во многих случаях различного рода ошибки и неточности выявляются при применении графиков, т. е. они иногда являются контролером точности расчётов и вычислений.

В настоящее время графики прочно вошли в практику экономического анализа в связи с внедрением в статистическую работу новых математических методов и современной вычислительной техники на базе ПЭВМ, с использованием пакетов прикладных программ компьютерной графики. Наиболее распространёнными пакетами прикладных программ являются: «Excel», «Stat Graff», «Super call», «Hazard graphics» и др. Эти программы облегчают задачу исследователя в практическом применении графиков, так как с помощью дисплеев можно демонстрировать графики на световом экране, при необходимости оперативно изменяя в них одни данные, вводя другие и т. д. Такого рода графики в принципе могут заменить громоздкие таблицы компактными изображениями.

Графики различаются по своему виду, и задача состоит в том, чтобы найти наиболее подходящий график. Нужно научиться правильно пользоваться орудием графического метода при изображении статистических данных. Кроме этого, график надо уметь строить, понимать принцип его построения. В противном случае можно выбрать правильный график, но сделать его таким, что он исказит действительную картину.

4.2. Общие правила построения графического изображения

Несмотря на большое разнообразие статистических графиков, существуют общие правила их построения.

При построении графика важно найти такие способы изображения, которые наилучшим образом отвечают содержанию и логической природе изображаемых показателей.

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

Графический образ (основа графика) - это геометрические знаки, то есть совокупность точек, линий, фигур, с помощью которых изображаются статистические показатели. Важно правильно выбрать графический образ, который должен соответствовать цели графика и способствовать наибольшей выразительности изображаемых статистических данных. Так например на рисунке 4.4 графический образ представляет собой ряд столбиков, на рисунке 4.7 - ряд квадратов и т.п.

Вспомогательные элементы делают возможным чтение графика, его понимание и использование. К ним относятся: 1) экспликация графика; 2) пространственные ориентиры; 3) масштабные ориентиры; 4) поле графика.

Рассмотрим каждый из них.

Экспликация графика – словесное описание его содержания. Оно включает в себя общий заголовок графика, подписи вдоль масштабных шкал и пояснения к отдельным частям графика.

Заголовок графика должен в краткой и ясной форме отражать основное содержание (тему) данных, изображенных на графике; в нем указываются ограниченный в пространстве и времени объект , к которому относятся данные. Если заголовок является частью текста (в книге, статье, дипломной работе и т.д.), то он обычно помещается под нижним краем графика. Если график представляется отдельно от текста, заголовок пишется вверху графика буквами и цифрами более крупного размера, чем все остальные надписи на графике.

В графике, кроме заголовка, обязательно даются словесные пояснения условных знаков и смысла отдельных элементов графического образа. Сюда относятся названия и цифры масштабов, названия ломаных линий, цифры, характеризующие величины отдельных частей графика, ссылки на источники и т.д.

Пояснительные надписи, раскрывающие смысл отдельных элементов графического образа, могут быть помещены либо на самом графике (на графическом образе или рядом с ним) в виде так называемых **ярлыков** (см. рис. 4.8), либо в виде **ключа,** вынесенного за пределы графического образа (рис. 4.5). Последний способ обычно применяется в тех случаях, когда на графике недостаточно места, а пояснения длинные.

Пространственные ориентиры графика задаются в виде системы координатных сеток. Системы координат бывают прямолинейные (декартовые) и криволинейные. Для построения графиков используется обычно только первый и, изредка, первый и четвертый квадранты. Криволинейные координаты – это окружность, разделенная на 360º. В практике графического изображения применяются также полярные координаты. Они необходимы для циклического движения во времени.

Масштабные ориентиры статистического графика определяются масштабом и системой масштабных шкал. **Масштаб** статистического графика- это мера перевода числовой величины в графическую. Например, 1 см высоты столбика равен 50 тыс. рублей уставного капитала коммерческого банка. Если график построен в виде площадей или объемов, масштабами служат единицы площадей или объемов (Например, 1см2=100км2 территории области).

Масштабы выбирают так, чтобы на графике ясно выступало различие изображаемых величин, но в то же время не терялась возможность их сравнения.

В случае, если на графике наносится не один, а два масштаба (в прямоугольной системе координат), соотношение их поля выбирается таким образом, чтобы стороны занятого графиком пространства по вертикали и горизонтали относились как $1:\sqrt{2}$. и $\sqrt{2}:1$.**Масштабной шкалой** называется линия, отдельные точки которой могут быть прочитаны как определённые числа. Шкала имеет большое значение в графике. В ней различают три элемента: линию (или носитель шкалы), определённое число помеченных чёрточками точек, которые расположены на носителе шкалы в определённом порядке, цифровое обозначение чисел, соответствующих отдельным помеченным точкам. Как правило, цифровым обозначением снабжаются не все помеченные точки, а лишь некоторые из них, расположенные в определённом порядке. По правилам числовое значение необходимо помещать строго против соответствующих точек, а не между ними (рис. 4.1).

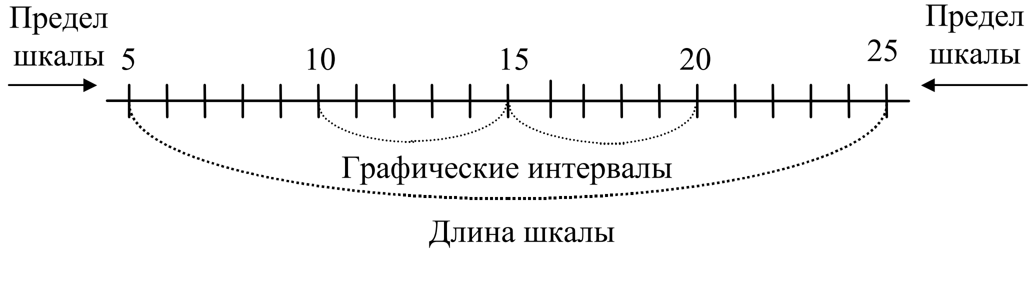

Рис. 4.1. Масштабная сетка

Графические и числовые интервалы могут быть равными и неравными. Если на всём протяжение шкалы равным графическим интервалам соответствуют равные числовые, такая шкала называется **равномерной**. Если же равным числовым интервалам соответствуют неравные графические, и наоборот ,- шкала называется **неравномерной**.

Масштабом равномерной шкалы называется длина отрезка (графический интервал), принятого за единицу и измеренного в каких-либо мерах. Чем меньше масштаб, тем гуще располагаются на шкале точки, имеющие одно и тоже значение. Построить шкалу – это значит на заданном носителе шкалы разместить точки и обозначить их соответствующими числами согласно условиям задачи. Из неравномерных наибольшее значение имеет логарифмическая шкала. Методика её построения несколько иная, так как на этой шкале отрезки пропорциональны не изображаемым величинам, а их логарифмам. Так при основании 10 lg1=0; lg10=1; lg100=2 и т. д.

Носитель шкалы может представлять собой как прямую, так и кривую линию. В соответствие с этим различают шкалы прямолинейные (например, миллиметровая линейка) и криволинейные - дуговые и круговые (циферблат часов).

Поле графика – то пространство, в котором размещаются образующие график геометрические знаки. Поле графика характеризуется его форматом, т.е. размером и пропорциями (соотношением сторон).

Например, лист бумаги, на котором располагается график, должен быть пропорциональным. Считается, что наиболее удобной для восприятия глазом человека пропорцией, является прямоугольник $1:\sqrt{2}$, т.е. 1:1,474(примерно 5:7). Это сочетание принято в стандарте писчей бумаги, предназначенной для копировально-множительной техники с форматом А4, т.е.210мм:297мм.

Примерно такие же пропорции должны быть выдержаны и в размерах большей части собственно графических изображений. При этом длинная сторона графика(сетки) может быть расположена по горизонтали (широкий график) и по вертикали (высокий график).

Приступая к графическому изображению статистических данных, необходимо прежде всего выбрать форму графика и определить методологию и технику его построения.

4.3. Классификация основных видов статистических графиков

Для графического представления статистических данных используются самые разнообразные виды графиков (рис. 4.2 и 4.3). Их можно классифицировать по разным признакам: характеру графического образа, способу построения и назначению (содержанию).

По характеру графического образа различают графики объемные, линейные и плоскостные (рис. 4.2).

По способу построения графики можно разделить на диаграммы и статистические карты (рис. 4.3).

Диаграмма представляет собой чертеж, показывающий соотношение статистических данных при помощи разнообразных геометрических и изобразительных средств.

Статистические карты предназначены для графического изображения одноименных показателей, относящихся к разным территориям. Для этого в основу изображения берется географическая карта. Изображение на карте статистических данных называется картограммой или картодиаграммой.

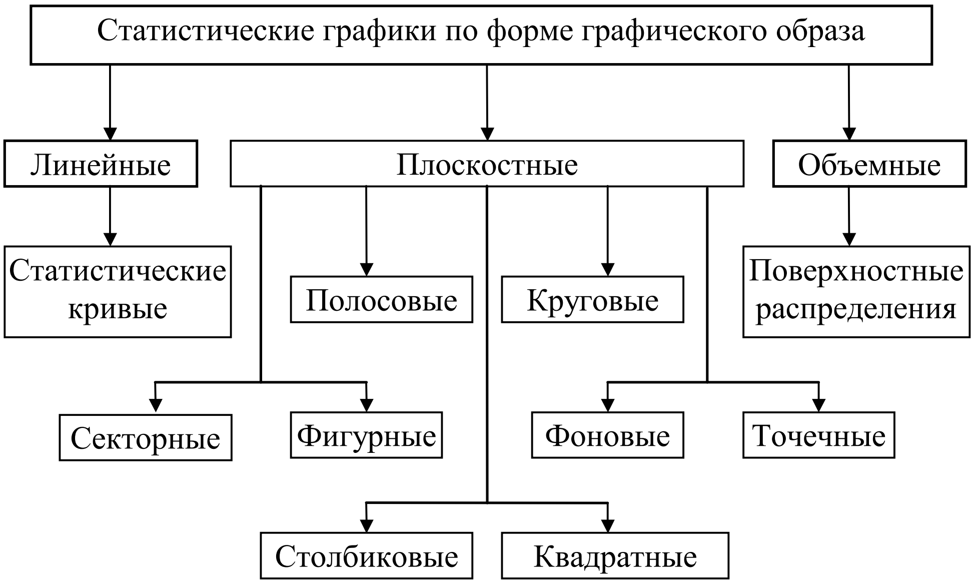

Рис. 4.2. Классификация статистических графиков по форме графического образа

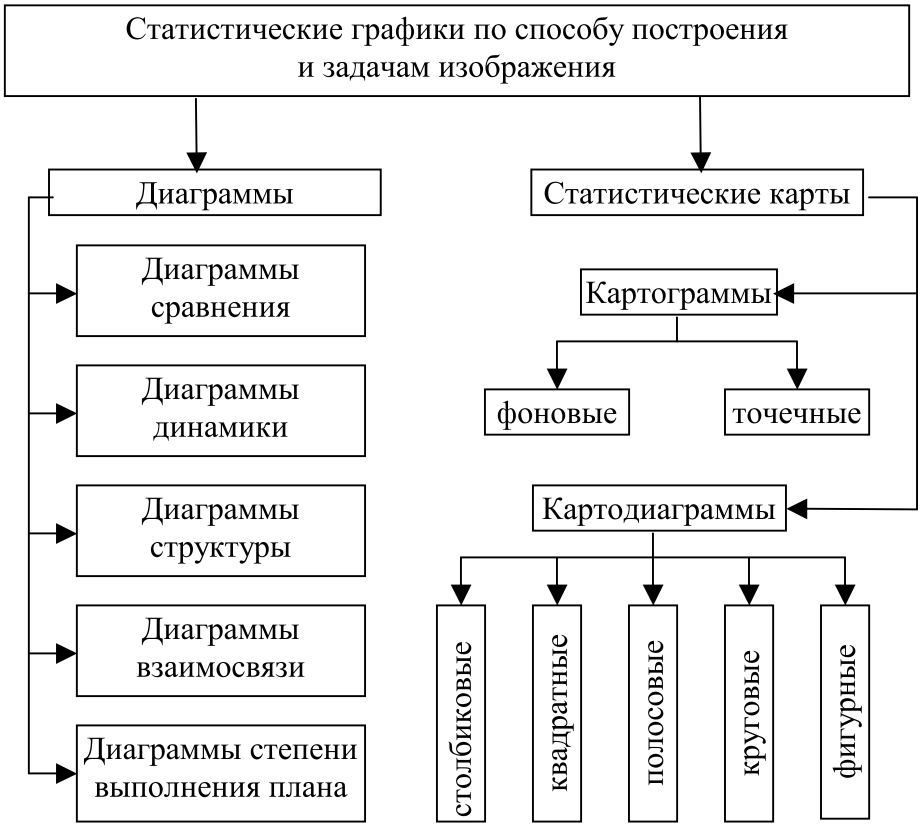

Рис.4.3. Классификация статистических графиков по способу построения и содержанию изображаемых данных

По содержанию или назначению можно выделить графики сравнения в пространстве, графики относительных величин (структуры, динамики и т.п.), графики вариационных рядов, графики взаимосвязанных показателей и графики размещения по территории (рис. 4.3).

4.4. Диаграммы сравнения

Различные виды диаграмм применяются для сравнения одноименных статистических данных, характеризующих разные территории или объекты. Наиболее распространённым видом таких диаграмм являются **столбиковые диаграммы**. Они представляют собой график, в котором различные величины представлены расположенными в высоту прямоугольниками («столбиками») одинаковой или разной высоты. Столбиковые диаграммы применяются для сравнения некоторых объектов во времени.

Построение такого рода диаграмм требует только одной **вертикальной масштабной шкалы,** которая определяет высоту каждого столбика.

Масштабная шкала должна начинаться с нуля, быть непрерывной и на ней записываются лишь круглые или округленные значения.

Столбики должны быть даны на некотором, друг к другу одинаковом для всех расстоянии или вплотную друг к другу. Ширина столбиков берется произвольной. На шкале должна быть указана единица измерения.

При выборе масштаба надо рассчитать так, чтобы максимальное число было представлено на графике.

Пример. Требуется изобразить с помощью столбиковой диаграммы данные о трудоустройстве граждан органами государственной службы занятости региона (цифры условные): в 2004 г. трудоустроено 2822 чел.; в 2003 г. – 2398 чел.; в 2002 г. – 2406 чел.; в 2001 г. – 2218 чел. Примем масштаб: 500 чел. соответствует 1 см. Тогда высота первого столбика (трудоустройство в 2001 г.) будет равна 4,4 см (1 см*2218/500), высота второго (в 2002 г.)-4,8 см; высота третьего (в 2003 г.)-4,79 см; высота четвертого (в 2004 г.)-5,6 см. Наглядность данной диаграммы достигается сравнением высоты столбиков (рис. 4.4).

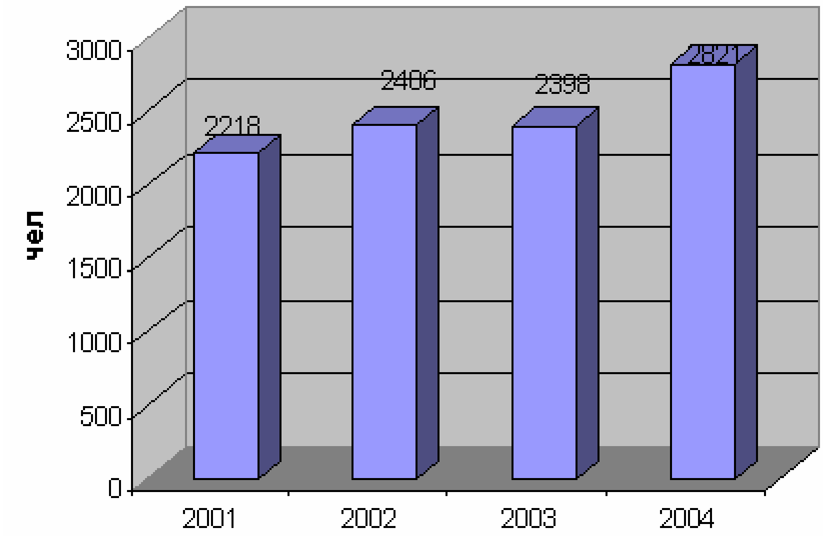

Рис. 4.4. Динамика трудоустройства граждан органами государственной службы занятости в регионе за 2001-2004 гг.

На рис.4.5 с помощью столбиковой диаграммы показана структура посевных площадей сельхозпредприятий N-ой области РФ за 2004 г. (цифры условные). На этой диаграмме столбики располагаются вплотную по группам объектов в пространстве.

Масштаб принят такой, что каждым 5000 тыс. га соответствует отрезок в 1 см.

Если прямоугольники, изображающие показатели, расположить не по вертикали, а по горизонтали, то диаграмма получит название **ленточной.** В качестве примера приведем полосовую диаграмму сравнения, характеризующую данные о реализации минеральных удобрений сельхозпредприятиями в N-ом регионе за 2001-2004 гг. (рис. 4.6).

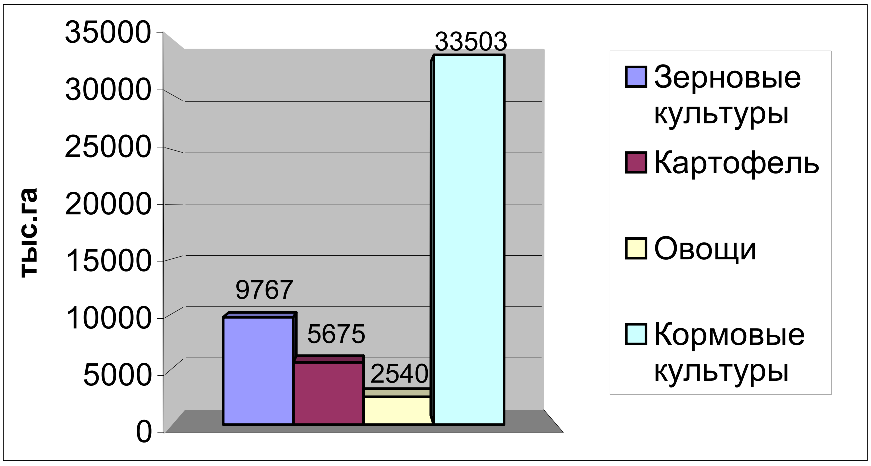

Рис. 4.5. Структура посевных площадей сельхозпредпртятий N-ой области РФ в 2004 г.

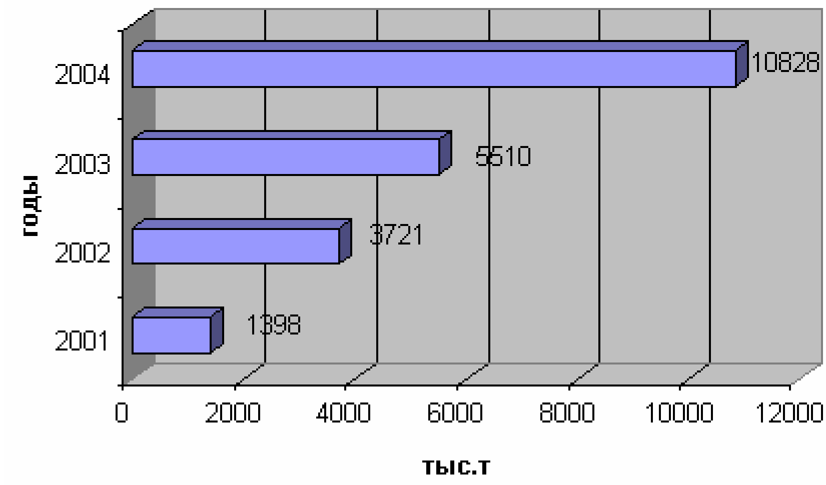

Рис. 4.6. Динамика реализации минеральных удобрений сельскохозяйственными предприятиями в N-ом регионе за 2001-2004 гг.

Иногда разница между наименьшими и наибольшими значениями сравниваемых данных настолько велика, что установление подходящего масштаба для столбиков или полос оказывается затруднительным. В этих случаях вместо столбиковой (полосовой) диаграммы целесообразно применить плоскостную (двухмерную) диаграмму – квадратную или круговую. Принцип построения этих диаграмм заключается в том, что величины сравниваемых данных изображаются площадями квадратов или кругов. Иными словами, площади квадратов (кругов) должны быть пропорциональны величинам изображаемых явлений, но сами площади квадратов (кругов) пропорциональны квадратам их сторон (радиусов). Следовательно, стороны квадратов или радиусов кругов должны быть пропорциональны корням квадратным из величин изображаемых статистических данных.

Пример. Необходимо с помощью квадратной диаграммы изобразить реализацию молочных продуктов предприятиями розничной торговли в одном из регионов за 2004г. по следующим данным:

Товар	Товарооборот,млн.руб
творог	11
сметана	16
молоко	19

Для построения квадратной диаграммы сначала извлечем квадратные корни из чисел: $\sqrt{11}=3{,}32$; $\sqrt{16}=4$; $\sqrt{19}=4{,}36$. Затем установим масштаб, например, примем 1см-1,5 млн.руб. Тогда сторона 1-го квадрата составит 2,2см (3,32:1,5); 2-го – 2,7см; 3-го – 2,9см (4,36:1,5). Далее строим квадраты.

Для правильного построения диаграммы квадраты необходимо расположить на одинаковом расстоянии друг от друга, а в каждой фигурке указать числовое значение, которое она изображает, не приводя масштаба измерения. (рис. 4.7)

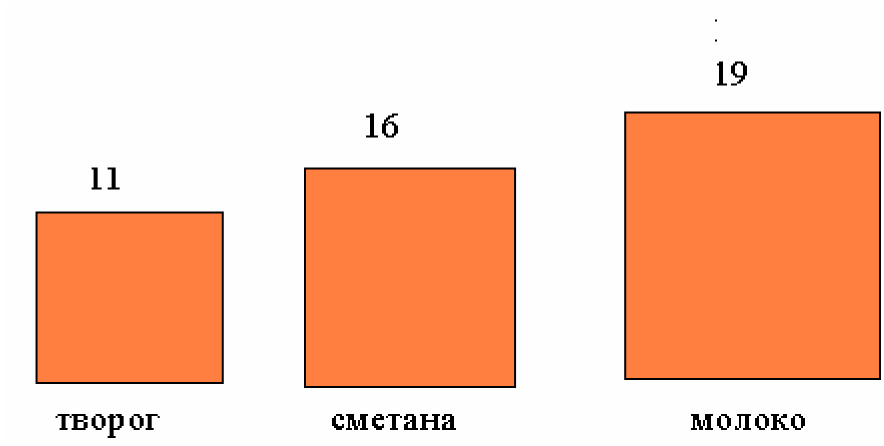

Рис. 4.7. Реализация молочных продуктов предприятиями розничной торговли в одном из регионов в 2004 г., млн.руб.

Круговая диаграмма строится аналогично квадратной с той разницей, что находим величину радиуса для каждого круга.

Пример. По данным об иностранных инвестициях в экономику РФ по основным странам-инвесторам за 2002 г. построить круговые диаграммы:

Страна	Германия	Кипр	Швейцария
Инвестиции, млн долларов США (x)	4001	2327	1349
$\sqrt{x}$	63,25	48,24	36,7
R	3,2	2,4	1,8

Примем 1см - 20млн дол, тогда радиус 1-го круга будет 3,2см (63,25:20), 2-го круга - 2,4см;3-го круга - 1,8см. (рис. 4.8).

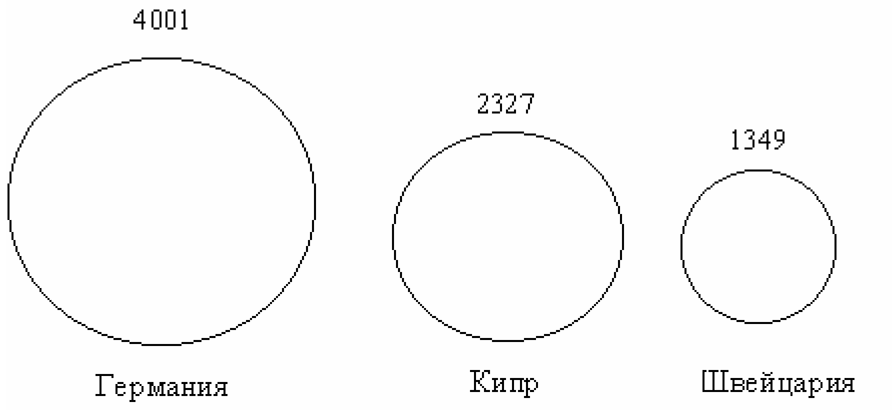

Рис. 4.8. Иностранные инвестиции в экономику РФ по основным странам-инвесторам за 2002 г., млн. долларов США

Диаграммы, предназначенные для популяризации, иногда строятся в виде стандартных фигур-рисунков, характерных для изображаемых статистических данных, что делает диаграмму более выразительной, привлекает к ней внимание. Такие диаграммы называются **фигурными** или **изобразительными**. Каждая фигурка имеет одинаковый размер и принимается за определённую величину изображаемых статистических данных.

Пример. Изобразим в виде фигурной диаграммы количество проданных магнитофонов в N-ом регионе за 2001-2004 гг. по следующим данным:

Годы	2001	2002	2003	2004
Продано,тыс.шт.	1977	862	875	995

Примем условно за один знак 300 тыс. штук магнитофонов. Тогда продажа магнитофонов в 2001 г. в размере 1977 тыс. штук будет изображена в количестве 6,6 магнитофона, в 2002 г. – 2,9 магнитофона, в 2003 г. – 2,9 магнитофона, в 2004 г. – 3,3 магнитофона. (рис. 4.9).

Недостаток фигурных диаграмм заключается в том, что во многих случаях приходится либо округлять изображаемые данные, либо изображать, кроме целых фигур, их части, размер которых на глаз оценивать трудно.

Рис. 4.9. Динамика продажи магнитофонов за 2001-2004 гг.

Для графического изображения трех взаимосвязанных показателей, один из которых равен произведению двух других, российский статистик проф. В.Е.Варзар предложил использовать **прямоугольную** диаграмму, названную им «статистическим знаком». В настоящее время такие диаграммы часто называют **знаком Варзара.**

Знак Варзара строится в виде прямоугольника, основание которого пропорционально одному показателю-сомножителю, а высота – второму показателю сомножителю. Тогда произведение этих показателей, т.е. третий показатель, будет изображаться площадью прямоугольника.

Пример. Имеются следующие данные в 2001 г. по всему миру:

ВНП – 46403 млрд. долл.
ВНП на душу населения – 7570 долл.
Средняя численность населения – 6,1298 млрд. чел.

Нужно изобразить эти данные с помощью знака Варзара (рис. 4.10). Взаимосвязь этих показателей можно представить в виде:

$$\left(\begin{array}{c}\text{Производство}\\ \text{ВНП на душу}\\ \text{населения}\end{array}\right) \text{X} \left(\begin{array}{c}\text{Средняя}\\ \text{численность}\\ \text{населения}\end{array}\right) = \left(\begin{array}{c}\text{Общее}\\ \text{производство}\\ \text{ВНП}\end{array}\right)$$

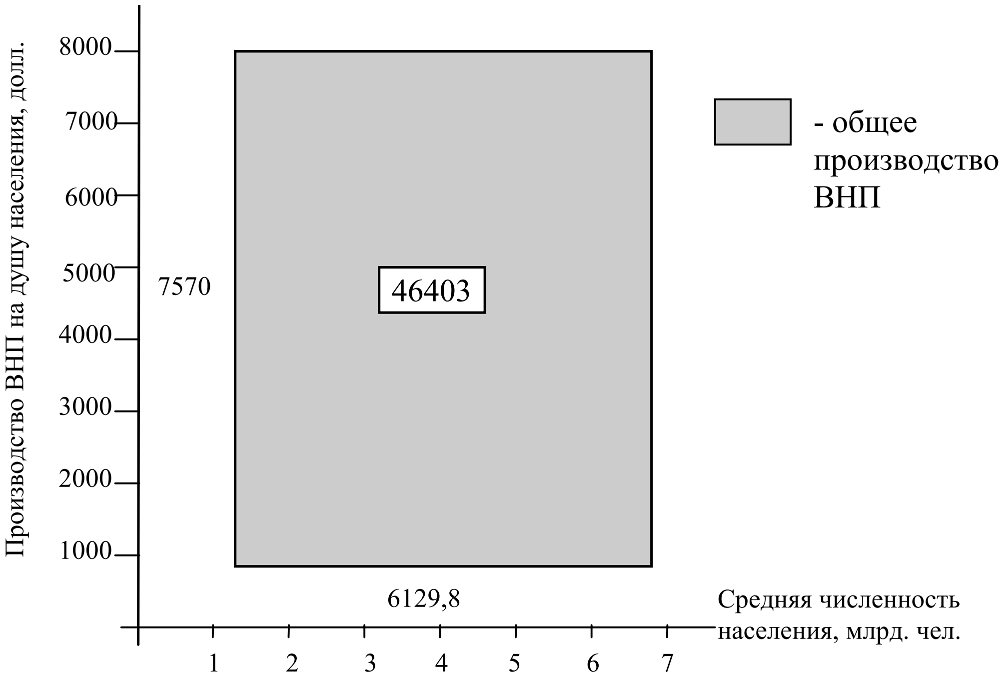

Рис. 4.10. Зависимость общего производства ВНП от производства ВНП на душу населения и средней численности населения мира в 2001 г.

4.5. Диаграммы структуры

Вторую большую группу показательных графиков составляют структурные диаграммы. Это такие диаграммы, в которых отдельные статистические совокупности сопоставляются по их структуре, характеризующейся соотношением разных параметров совокупности или ее отдельных частей.

Широко распространенный метод графического изображения структуры статистических данных заключается в составлении структурных круговых или секторных диаграмм. **Секторные диаграммы** удобно строить следующим образом: вся величина явления принимается за сто процентов, рассчитываются доли отдельных частей в процентах. Круг разбивается на секторы пропорционально частям изображаемого целого. Таким образом, на 1% приходится 3,6 градуса. Для получения центральных углов секторов, изображающих доли частей целого, необходимо их процентное выражение умножить на 3,6 градуса. Секторные диаграммы позволяют не только разделить целое на части, но и сгруппировать отдельные части, давая как бы комбинированную группировку долей по двум признакам.

Пример. Рассмотрим построение секторной диаграммы по следующим данным о структуре иностранных инвестиций в РФ в 2002 году:

Тип инвестиций	прямые	портфельные	прочие
Доля инвестиций, в%	20	2	78

Построение секторной диаграммы начинается с определения центральных углов секторов. Для этого процентное выражение отдельных частей совокупности умножим на 3,6 градуса, т.е. 20*3,6=72°; 2*3,6=7,2°; 78*3,6=280,8°. По найденным значениям углов круг делится на соответствующие сектора (рис. 4.11).

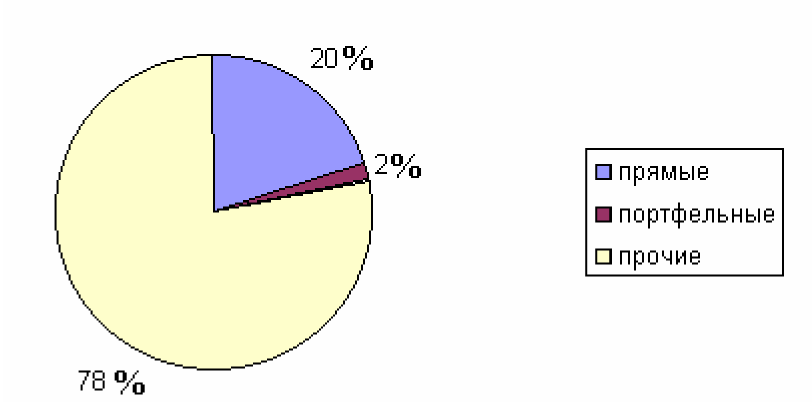

Рис. 4.11. Удельный вес иностранных инвестиций в РФ за 2002 г.

Другим видом структурных статистических диаграмм являются диаграммы удельных весов, отражающие структуры сравниваемых совокупностей по процентному соотношению в них отдельных частей, выделяемых по тому или иному количественному или атрибутивному признаку. Эти диаграммы получены путем преобразования простой полосовой диаграммы с подразделенными полосами. Полосовые диаграммы удельных весов могут вскрыть экономически существенные особенности многих изучаемых экономических явлений.

Пример. Необходимо изобразить графически следующие данные, характеризующие структуру потребительских расходов населения в N-ом регионе за 2003-2004 гг.(в процентах):

Показатели	2003	2004
все потребительские расходы; в том числе:	100	100
продукты питания	45,9	41,7
непродовольственные товары	34,4	36,2
алкогольные напитки	2,4	2,2
оплата услуг	17,3	19,9

Изобразим эти данные графически в виде полосовой диаграммы, цель которой - показать изменение удельных весов потребительских расходов населения за два года (рис. 4.12).

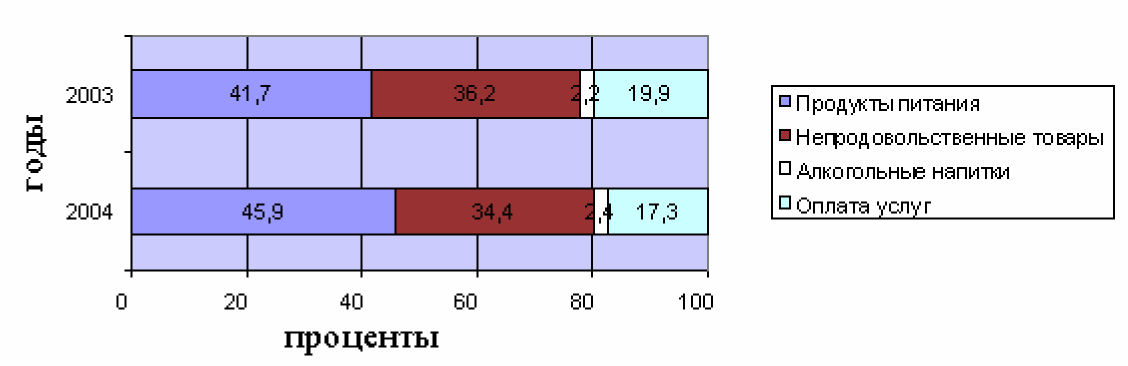

Рис. 4.12. Динамика удельного веса потребительских расходов населения в N-ом регионе за 2003-2004 гг.

Значительными преимуществами полосовых структурных диаграмм по сравнению с другими видами является их большая емкость, возможность отразить на небольшом пространстве большой объем полезной информации.

Секторные диаграммы выглядят убедительно при существенных различиях сравниваемых структур, а при небольших различиях они могут быть недостаточно выразительны.

4.6. Диаграммы динамики

Для изображения и внесения суждений о развитии явления во времени строятся диаграммы динамики. В рядах динамики используются для наглядного изображения явлений многие диаграммы: столбиковые, ленточные, квадратные, круговые, линейные, радиальные и другие. Выбор вида диаграмм зависит в основном от особенностей исходных данных, от цели исследования. Например, если имеется ряд динамики с неравноотстоящими уровнями во времени (1913, 1940, 1950, 1980, 2000, 2005 гг), то часто для наглядности используют столбиковые, квадратные или круговые диаграммы. Они зрительно впечатляют, хорошо запоминаются, но не годны для изображения большого числа уровней, так как громоздки, и если число уровней в ряду динамики велико, то целесообразно применять **линейные диаграммы**, которые воспроизводят непрерывность процесса развития в виде непрерывной ломаной линии.

Для построения линейных диаграмм используют систему прямоугольных координат. Обычно по оси абсцисс откладывается время (годы, месяцы и т.д.), а по оси ординат наносят масштабы для отображения явлений или процессов. Особое внимание следует обратить на масштаб осей координат, так как от этого зависит общий вид графика. Обеспечение равновесия, пропорциональности между осями координат необходимо в диаграмме, так как нарушение равновесия дает неправильное изображение развития явления. Если масштаб для шкалы на оси абсцисс очень растянут по сравнению с масштабом на оси ординат, то колебания в динамике явлений мало выделяются, и наоборот, преувеличение масштаба по оси ординат по сравнению с масштабом на оси абсцисс дает резкие колебания. Если в ряду динамики данные за некоторые годы отсутствуют, это должно быть учтено при построении графика. Равным периодам времени и размерам уровня должны соответствовать равные отрезки масштабной шкалы.

Пример. Рассмотрим построение линейной диаграммы на основании следующих данных:

Динамика валового сбора кормовых культур в регионе за 1995-2004 г.

Годы	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004
Валовый сбор, млн. тонн	237	179	189	158	186	192	172	191	210	211

Изображение динамики валового сбора кормовых культур на координатной сетке с неразрывной шкалой значений, начинающихся от нуля, вряд ли целесообразно, так как 2/3 поля диаграммы остается неиспользованным и ничего не дает для выразительности изображения. Поэтому в данных условиях рекомендуется строить шкалу без вертикального нуля, то есть шкала значений разрывается недалеко от нулевой линии и на диаграмму попадает лишь часть возможного поля графика. Это не приводит к искажениям в изображении динамики явления и процесс его изменения рисуется диаграммой более четко (рис. 4.13).

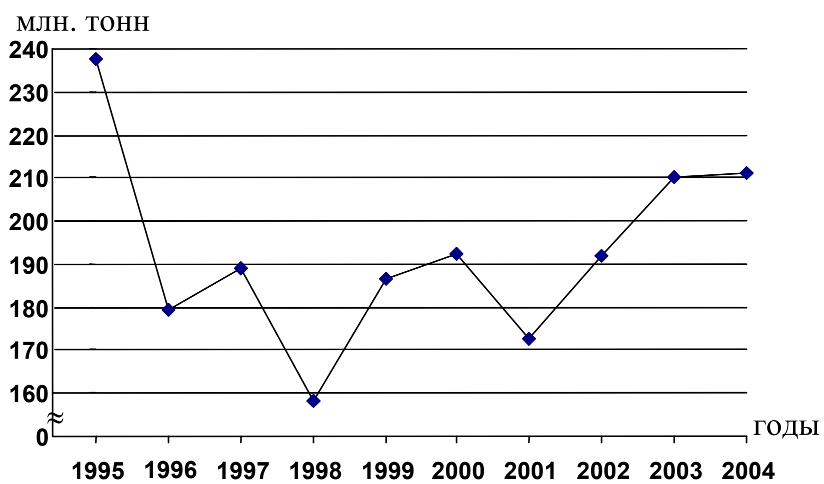

Рис. 4.13. Динамика валового сбора кормовых культур в регионе за 1995-2004 гг.

Нередко на одном линейном графике приводится несколько кривых, которые дают сравнительную характеристику динамики различных показателей или одного и того же показателя в разных странах. Примером графического изображения сразу нескольких показателей может служить рис. 4.14.

Линейные диаграммы с равномерной шкалой имеют недостаток, снижающий их познавательную ценность. Этот недостаток заключается в том, что равномерная шкала позволяет измерять и сравнивать только отраженные на диаграмме абсолютные приросты или уменьшения показателей на протяжении исследуемого периода. Однако при изучении динамики важно знать относительные изменения исследуемых показателей по сравнению с достигнутым уровнем или темпы их изменения.

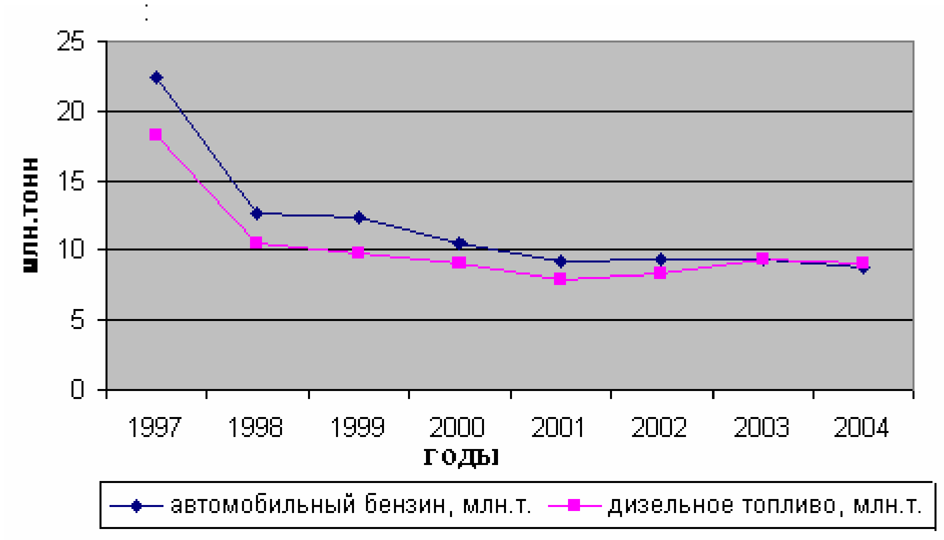

Рис. 4.14. Динамика потребления автобензина и дизельного топлива автотранспортом организаций отраслей экономики региона за 1997-2004 гг.

Именно относительные изменения экономических показателей в динамике искажаются при изображении их на координатной диаграмме с равномерной вертикальной шкалой. Кроме того, в обычных координатах теряет всякую наглядность и даже становится невозможным изображение рядов динамики с резко изменяющимися уровнями, которые обычно имеют место в динамических рядах за длительный период времени.

В этих случаях следует отказаться от равномерной шкалы и положить в основу графика полулогарифмическую систему. **Полулогарифмической системой** называется система, в которой на одной оси нанесен линейный масштаб, а на другой логарифмический. В данном случае логарифмический масштаб наносится на ось ординат, а на оси абсцисс располагают равномерную шкалу для отсчета времени по принятым интервалам (годам, кварталам, месяцам, дням и прочее).

Техника построения логарифмической шкалы следующая: необходимо найти логарифмы исходных чисел; начертить ординату и разделить на несколько равных частей. Затем нанести на ординату (или равную ей параллельную линию) отрезки, пропорциональные абсолютным приростам этих логарифмов. Далее записать соответствующие логарифмы чисел и их антилогарифмы, например (0,000; 0,3010; 0,4771; 0,6021; ... ; 1,000, что дает 1, 2, 3, 4 ..., 10). Полученные антилогарифмы окончательно дают вид искомой шкалы на ординате. Логарифмический масштаб лучше понять на примере.

Пример. Допустим, нужно изобразить на графике динамику производства газа в регионе за 1975-2004 гг., за эти годы его рост составил 9,1 раза. С этой целью находим логарифмы для каждого уровня ряда (см. таблицу 4.1).

Таблица 4.1.

Динамика производства газа в регионе за 1975-2004 гг. (млн. м³)

Годы	Yi	LgYi
1975	170	2,23
1980	292	2,46
1985	507	2,70
1990	741	2,84
1995	1039	3,02
2000	1294	3,11
2004	1544	3,19

Найдя минимальное и максимальное значения логарифмов производства газа, строим масштаб с таким расчетом, чтобы все данные разместились на графике. В соответствии с масштабом находим соответствующие точки, которые соединим прямыми линиями. В результате получим график (рис. 4.15) с использованием логарифмического масштаба на оси ординат.

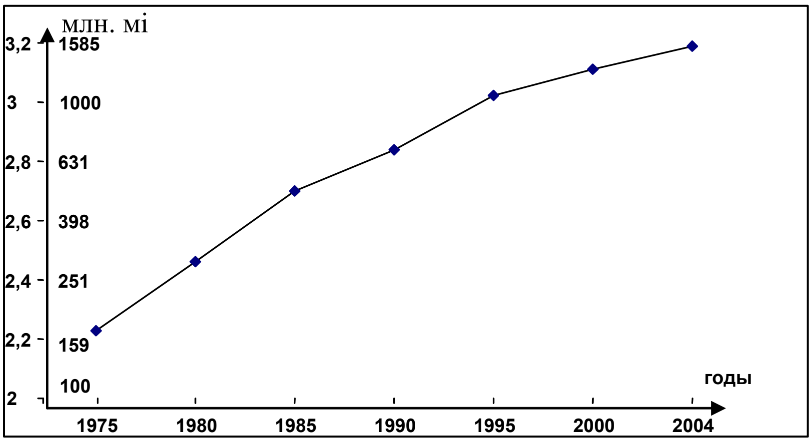

Рис. 4.15. Динамика производства газа в регионе за 1975-2004 гг.

К диаграммам динамики относятся и **радиальные диаграммы**, построенные в полярных координатах и предназначенные для отражения процессов, ритмически повторяющихся во времени. Чаще всего эти диаграммы применяются для иллюстрации сезонных колебаний, и в этом отношении они имеют преимущество перед статистическими кривыми. Радиальные диаграммы делятся на два вида: замкнутые и спиральные. Эти два вида диаграмм отличаются друг от друга по технике построения, все зависит от того, что взято в качестве базы отсчета - центр круга или окружность.

Замкнутые диаграммы отражают весь внутригодичный цикл динамики одного года. Их построение сводится к следующему: вычерчивается круг, среднемесячный показатель приравнивается к радиусу этого круга, затем весь круг делится на двенадцать равных секторов, посредством проведения радиусов, которые изображаются в виде тонких линий. Каждый радиус изображает месяц, причем расположение месяцев аналогично циферблату часов. На каждом радиусе делается отметка в определенном месте, согласно масштабу, исходя из данных на соответствующий месяц. Если данные превышают среднегодовой уровень, то отметка делается вне окружности на продолжении радиуса. Затем отметки различных месяцев соединяются отрезками.

Пример. Необходимо изобразить с помощью замкнутой диаграммы динамику уголовно-наказуемых преступлений в одном из городов за 2004 г. по следующим данным:

Месяцы	Количество преступлений	Месяцы	Количество преступлений
январь	8345	июль	7542
февраль	6419	август	6396
март	7720	сентябрь	6792
апрель	5976	октябрь	7296
май	5304	ноябрь	49999
июнь	6176	декабрь	6425

По данным приведенным в таблице определим среднемесячное количество преступлений (R=79420/12=6618). Масштаб 1см=1000 преступлений (рис. 4.16).

Если в качестве базы отсчета берется окружность, такого рода диаграммы называются спиральными. Спиральные диаграммы отличаются от замкнутых тем, что в них декабрь одного года соединяется не с январем данного же года, а с январем следующего года. Это дает возможность изобразить весь динамический ряд за несколько лет в виде одной кривой. Особенно наглядна такая диаграмма тогда, когда наряду с сезонным ритмом ряд обнаруживает неуклонный рост из года в год.

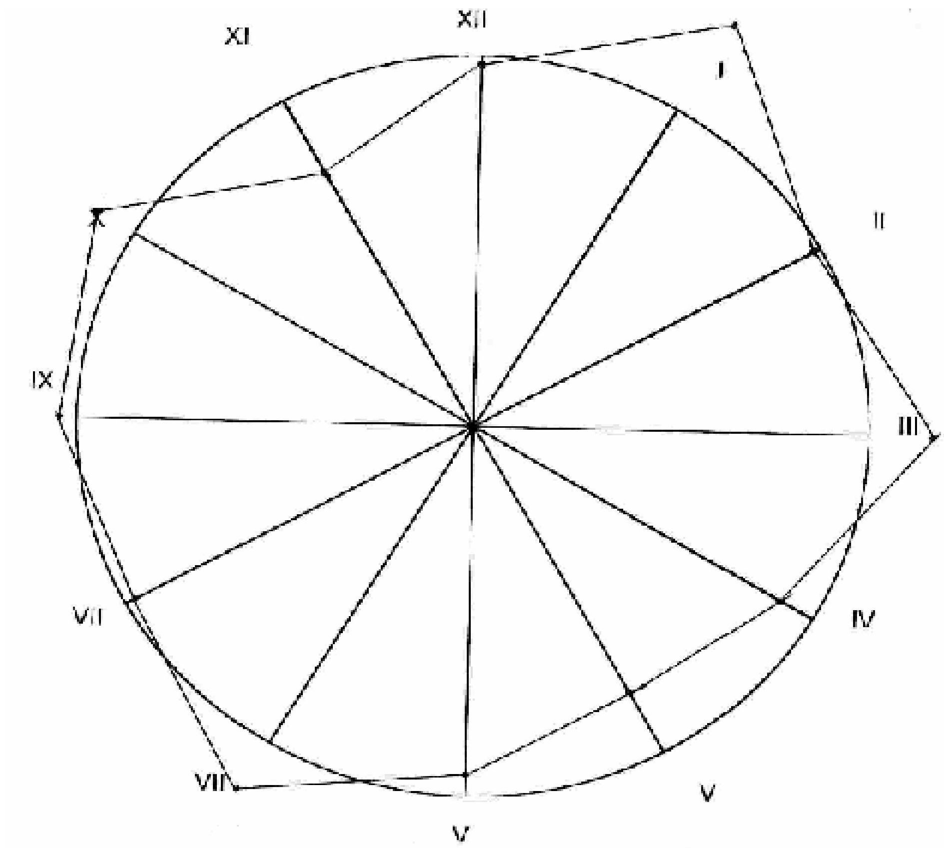

Рис. 4.16. Динамика уголовно-наказуемых преступлений в одном из городов за 2004 г.

Для отображения зависимости одного показателя от другого строится **диаграмма взаимосвязи**. Один показатель принимается за X, а другой за Y (т.е. функцию от X). Строится прямоугольная система координат с масштабами для показателей, в которой вычерчивается график. На рисунке 4.17 показана взаимосвязь между стоимостью основных производственных фондов и уровнем затрат на реализацию продукции.

Рис. 4.17. Зависимость уровня затрат на реализацию продукции от стоимости основных производственных фондов

Рис. 4.17 показывает, что с увеличением стоимости основных производственных фондов происходит увеличение затрат на реализацию продукции и данная зависимость этих показателей может быть выражена линейной связью.

Диаграммы взаимосвязи имеют большое значение на практике, так как множество различных показателей связаны между собой либо прямой, либо обратной формой связи. Они могут использоваться также для отображения различных циклических процессов (например инфляционной спирали), взаимонакладывающихся явлений и т.п.

4.6.Статистические карты

Карты статистические представляют собой вид графических изображений статистических данных на схематичной географической карте, характеризующих уровень или степень распространения того или иного явления на определенной территории.

Средствами изображения территориального размещения являются штриховка, фоновая раскраска или геометрические фигуры. Различают картограммы и картодиаграммы.

Картограмма - это схематическая географическая карта, на которой штриховкой различной густоты, точками или окраской различной степени насыщенности показывается сравнительная интенсивность какого-либо показателя в пределах каждой единицы нанесенного на карту территориального деления (например, плотность населения по областям или республикам, распределение районов по урожайности зерновых культур и т.п.). Картограммы делятся на фоновые и точечные.

Картограмма фоновая - вид картограммы, на которой штриховкой различной густоты или окраской различной степени насыщенности показывают интенсивность какого-либо показателя в пределах территориальной единицы. **Картограмма точечная** - вид картограммы, где уровень какого-либо явления изображается с помощью точек. Точка изображает одну единицу совокупности или некоторое их количество, чтобы показать на географической карте плотность или частоту появления определенного признака.

Пример. Необходимо с помощью точечной картограммы изучить размещение посевов картофеля по территории области (цифры условные). (рис. 4.18)

Таблица 4.2.

Номер района на контурной карте	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Площадь посева картофеля, га	5930	5810	4690	6200	5700	6050	4820	3920	4280	3880	4055	2910	2540	2430	2720
Число точек на карте, 1 точка= 100 га	59	58	47	62	57	60	48	39	43	39	41	29	25	24	27

Решение.

Составим точечную картограмму размещения посевных площадей картофеля.

1. Отразим на карте размеры площади посева картофеля в каждом районе в виде определенного числа точек. Для этого установим, какая площадь картофеля будет соответствовать одной точке, т.е. определим масштаб картограммы. Для наглядности картограммы нужно, чтобы число точек было оптимальным, так как при большом количестве точки сольются, а при малом не отразят существующие различия между районами. При выборе масштаба следует учитывать, что при данных размерах контурной карты на территории района может быть размещено максимум 60-80 точек и что величина масштаба должна быть округленным, удобным для пользования числом. Исходя из этого, целесообразно принять 1 точку, равную 100 га. При этом в районе 4, где находится максимум посевов картофеля – 6200 га, будет 62 точки (6200:100), что является оптимальным числом.

2. Определим в соответствии с принятым масштабом число точек, которые следует нанести в границах каждого района. Для этого посевную площадь каждого района разделим на величину масштаба 100 га и полученное число точек (с округлением до 1) запишем в таблице.

3. Нанесем на контурную карту данные по каждому району. При этом проследим, чтобы точки были одинакового размера и равномерно распределялись в границах района. Укажем на картограмме культуру, которой соответствует приведенные данные, а также обозначим масштаб.

1 точка=100 га

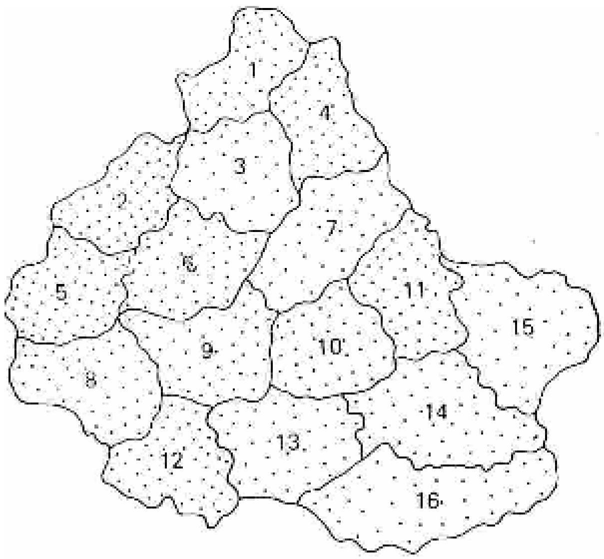

Рис. 4.18. Картограмма плотности размещения посевных площадей картофеля фермерских хозяйств по отдельным районам области

Картограмма показывает, что посевы картофеля сконцентрированы в основном в северо-западной части области, а к юго-востоку плотность размещения посевов заметно падает.

Пример 2. Построим фоновую картограмму урожайности картофеля фермерских хозяйств по отдельным районам области.

Решение.

1. Для построения фоновой картограммы предполагается предварительная группировка 16 районов по величине изучаемого признака – урожайности картофеля:

Группы районов	1	2	3	4
Урожайность, га	до 160	161-190	191-200	свыше 200

2. Установим для каждой группы районов вид штриховки. Интенсивность (густота) ее должна увеличиваться пропорционально нарастанию урожайности по группам районов и отражать различия в ней.

3. Заштрихуем районы, отнесенные к определенной группе, соответствующим видом штриховки. Укажем на картограмме культуру, интервалы урожайности и принятую для них штриховку (рис. 4.19).

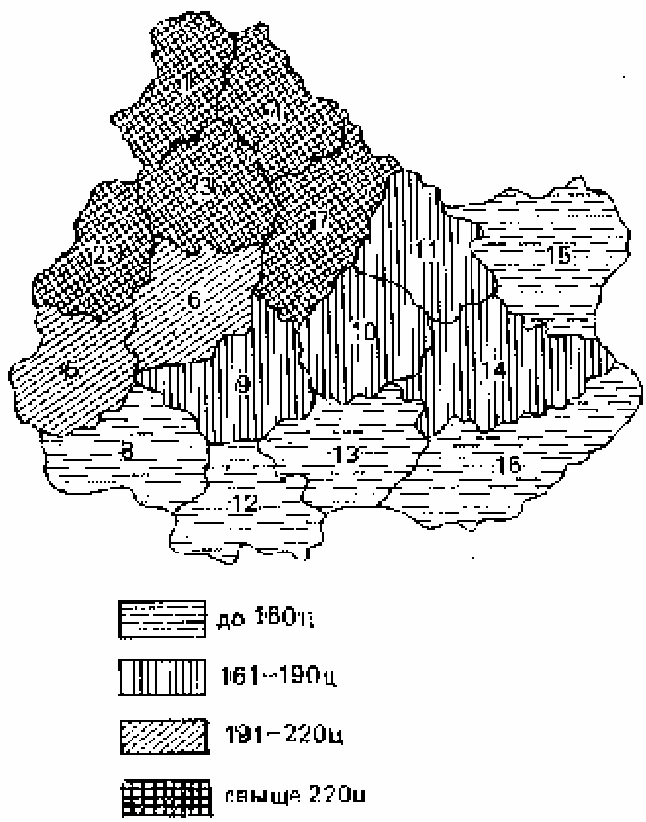

Рис. 4.19. Картограмме плотности размещения урожайности картофеля фермерских хозяйств по отдельным районам области

Картограмма показывает, что наиболее высокая урожайность картофеля в северной и западной частях области, самая низкая – в южных районах. Сравнивая картограммы посевных площадей и урожайности, необходимо отметить, что размещение посевов картофеля и урожайность взаимосвязаны: площадь посева картофеля относительно больше в северо-западной части области, где выше урожайность.

Таким образом, фоновые картограммы, как правило, используются для изображения средних или относительных величин, точечные – для объемных (количественных) показателей.

Вторую большую группу статистических карт составляют картограммы. Они представляют собой сочетание диаграмм с географической картой. В качестве изобразительных знаков в картодиаграммах используются диаграммные фигуры (столбики, квадраты, круги, фигуры, полосы), которые размещаются на контуре географической карты. Картодиаграммы дают возможность географически отразить более сложные статистико-географические построения, чем картограммы. Дальнейшим развитием данного подхода представления статистической информации являются географические информационные системы (ГИС).

Глава 5. Абсолютные, относительные и средние статистические показатели

5.1. Абсолютные показатели

Исходной, первичной формой выражения статистических показателей являются показатели в абсолютном выражении или абсолютные величины. Статистические показатели в форме абсолютных величин характеризуют абсолютные размеры изучаемых статистикой процессов и явлений, а именно, их массу, площадь, объем, протяженность, отражают их временные характеристики, а также могут представлять объем совокупности, т.е. число составляющих ее единиц.

Индивидуальные абсолютные показатели, как правило, получают непосредственно в процессе статистического наблюдения как результат замера, взвешивания, подсчета и оценки интересующего количественного признака. В ряде случаев индивидуальные абсолютные показатели имеют разностный характер: разность между численностью работников предприятия на конец и на начало года, разность между выручкой от реализации предприятия и общей суммой затрат и т.п.

Сводные абсолютные показатели, характеризующие объем признака или объем совокупности как в целом по изучаемому объекту, так и по какой-либо его части, получают в результате сводки и группировки индивидуальных значений. К таким показателям относятся общая численность занятых в отрасли, совокупные активы коммерческих банков региона и т.п.

Абсолютные статистические показатели всегда являются именованными числами. В зависимости от социально-экономической сущности исследуемых явлений, их физических свойств они выражаются в натуральных, стоимостных или трудовых единицах измерения.

В международной практике используются такие *натуральные единицы измерения* как тонны, килограммы, квадратные, кубические и простые метры, мили, километры, галлоны, литры, штуки и т.д. Например, производство электроэнергии в России в 2003 г. составило 915 млрд. кВт·ч, за этот же год добыто 408 млн. т нефти и 620 млрд. куб. м газа.

В группу натуральных также входят условно-натуральные измерители, используемые в тех случаях, когда какой-либо продукт имеет несколько разновидностей и общий объем можно определить только исходя из общего для всех разновидностей потребительского свойства. Так, различные виды органического топлива переводятся в условное топливо с теплотой сгорания 29,3 МДж/кг (7000 ккал/кг), мыло разных сортов - в условное мыло с 40%-ным содержанием жирных кислот, консервы различного объема - в условные консервные банки объемом 353,4 куб.см и т.д.

Перевод в условные единицы измерения осуществляется на основе специальных коэффициентов, рассчитываемых как отношение потребительских свойств отдельных разновидностей продукта к эталонному значению. Так, например, 100 т торфа, теплота сгорания которого - 24 МДж/кг, будут эквивалентны 81,9 т условного топлива (100 * 24,0/29,3), а 100 т нефти при теплоте сгорания 45 МДж/кг будут оцениваться в 153,6 т условного топлива (100 * 45,0/29,3).

В отдельных случаях для характеристики какого-либо явления или процесса одной единицы измерения недостаточно, и используется произведение двух единиц. Примером этому могут служить такие показатели как грузооборот и пассажирооборот, оцениваемые соответственно в тонно-километрах и пассажиро-километрах, производство электроэнергии, измеряемое в киловатт-часах и т.д.

В условиях рыночной экономики наибольшее значение и применение имеют *стоимостные единицы измерения*, позволяющие получить денежную оценку социально-экономических явлений и процессов. Так, одним из важнейших стоимостных показателей в системе национальных счетов, характеризующим общий уровень развития экономики страны, является валовой внутренний продукт, который в России за 1 квартал 2003 года составил 2893 млрд. рублей.

При анализе и сопоставлении стоимостных показателей необходимо иметь в виду, что в условиях высоких или относительно высоких темпов инфляции они становятся несопоставимыми. Так, сравнивать ВВП России за 2003 год с его величиной, например, за 1993 год вряд ли целесообразно, так как содержание рубля за этот период существенно изменилось. Для того, чтобы произвести подобные сравнения, там где это возможно, осуществляют пересчет в сопоставимые цены.

К *трудовым единицам измерения*, позволяющим учитывать как общие затраты труда на предприятии, так и трудоемкость отдельных операций технологического процесса, относятся человеко-дни и человеко-часы.

5.2. Относительные показатели

Относительный показатель представляет собой результат деления одного абсолютного показателя на другой и выражает соотношение между количественными характеристиками социально-экономических процессов и явлений. Поэтому, по отношению к абсолютным показателям, относительные показатели или показатели в форме относительных величин являются производными, вторичными. Без относительных показателей невозможно измерить интенсивность развития изучаемого явления во времени, оценить уровень развития одного явления на фоне других взаимосвязанных с ним явлений, осуществить пространственно-территориальные сравнения, в том числе и на международном уровне.

При расчете относительного показателя абсолютный показатель, находящийся в числителе получаемого отношения, называется текущим или сравниваемым. Показатель же, с которым производится сравнение и который находится в знаменателе, называется основанием или базой сравнения. Таким образом, рассчитываемая относительная величина показывает, во сколько раз сравниваемый абсолютный показатель больше базисного, или какую составляет от него долю, или сколько единиц первого приходится на 1, 100, 1000 и т. д. единиц второго.

Относительные показатели могут выражаться в коэффициентах, процентах, промилле, продецимилле или быть именованными числами. Если база сравнения принимается за 1, то относительный показатель выражается в коэффициентах, если база принимается за 100, 1000 или 10000, то относительный показатель соответственно выражается в процентах (%), промилле (‰) и продецимилле (‱).

Относительный показатель, полученный в результате соотнесения разноименных абсолютных показателей, в большинстве случаев должен быть именованным. Его наименование представляет собой сочетание наименований сравниваемого и базисного показателей (например, производство какой-либо продукции в соответствующих единицах измерения в расчете на душу населения).

Все используемые на практике относительные статистические показатели можно подразделить на следующие виды:

1) динамики;
2) плана;
3) реализации плана;
4) структуры;
5) координации;
6) интенсивности и уровня экономического развития;
7) сравнения.

Относительный показатель динамики (ОПД) представляет собой отношение уровня исследуемого процесса или явления за данный период времени (по состоянию на данный момент времени) к уровню этого же процесса или явления в прошлом:

$$\text{ОПД} = \frac{\text{Текущий уровень}}{\text{Предшествующий или базовый уровень}}$$

Рассчитанная. таким образом величина показывает, во сколько раз текущий уровень превышает предшествующий (базисный) или какую долю от последнего составляет. Данный показатель может быть выражен кратным отношением или переведен в проценты.

Различают относительные показатели динамики с постоянной и переменной базой сравнения. Если сравнение осуществляется с одним и тем же базисным уровнем, например, первым годом рассматриваемого периода, получают относительные показатели динамики с постоянной базой (базисные). При расчете относительных показателей динамики с переменной базой (цепных) сравнение осуществляется с предшествующим уровнем, т.е. основание относительной величины последовательно меняется.

Для примера воспользуемся данными таблицы 5.1.

Таблица 5.1.

Производство легковых автомобилей в РФ в 2000 – 2003гг. (тыс. шт.)

Год	2000	2001	2002	2003
Объем производства	969	1022	981	1011

Рассчитаем относительные показатели динамики с переменной и постоянной базой сравнения:

переменная база сравнения (цепные показатели)	постоянная база сравнения (базисные показатели)
$\frac{1022}{969} \cdot 100\% = 105{,}5\%$	$\frac{1022}{969} \cdot 100\% = 105{,}5\%$
$\frac{981}{1022} \cdot 100\% = 96{,}0\%$	$\frac{981}{969} \cdot 100\% = 101{,}2\%$
$\frac{1011}{981} \cdot 100\% = 103{,}1\%$	$\frac{1011}{969} \cdot 100\% = 104{,}4\%$

Относительные показатели динамики с переменной и постоянной базой сравнения взаимосвязаны между собой следующим образом: произведение всех относительных показателей с переменной базой равно относительному показателю с постоянной базой за исследуемый период. Так, для рассчитанных показателей (предварительно переведя их из процентов в коэффициенты) получим:

$$1{,}055 \cdot 0{,}960 \cdot 1{,}031 = 1{,}044$$

Относительные показатели плана и реализации плана. Все субъекты финансово-хозяйственной деятельности, от небольших индивидуальных частных предприятий и до крупных корпораций, в той или иной степени осуществляют как оперативное, так и стратегическое планирование, а также сравнивают реально достигнутые результаты с ранее намеченными. Для этой цели используются относительные показатели плана (ОПП) и реализации плана (ОПРП):

$$ОПП = \frac{Уровень,\ планируемый\ на\ (i+1)\ период}{Уровень,\ достигнутый\ в\ i\text{-}м\ периоде}$$

$$ОПРП = \frac{Уровень,\ достигнутый\ в\ (i+1)\ периоде}{Уровень,\ планируемый\ на\ (i+1)\ период}$$

Первый из этих показателей характеризует относительную высоту планового уровня, т.е. во сколько раз намечаемый объемный показатель превысит достигнутый уровень или сколько процентов от этого уровня составит. Второй показатель отражает фактический объем производства или реализации в процентах или коэффициентах по сравнению с плановым уровнем.

Предположим, оборот торговой фирмы в 2002 г. составил 3,0 млн.руб. Исходя из проведенного анализа складывающихся на рынке тенденций руководство фирмы считает реальным в следующем году довести оборот до 3,6 млн.руб. В этом случае относительный показатель плана, представляющий собой отношение планируемой величины к фактически достигнутой, составит 120% $(\frac{3,6}{3,0} \cdot 100\%)$. Предположим теперь, что фактический оборот фирмы за 2003 г. составил 3,8 млн. руб. Тогда относительный показатель реализации плана, определяемый как отношение фактически достигнутой величины к ранее запланированной, составит 105,6% $(\frac{3,8}{3,6} \cdot 100\%)$.

Между относительными показателями плана, реализации плана и динамики существует следующая взаимосвязь:

$$ОПП \cdot ОПРП = ОПД$$

В нашем примере:

$$1,20 \cdot 1,056 = 1,267 \quad \text{или} \quad \frac{3,8}{3,0} = 1,267$$

Основываясь на этой взаимосвязи по любым двум известным величинам при необходимости всегда можно определить третью неизвестную величину.

Относительный показатель структуры представляет собой соотношение структурных частей изучаемого объекта и их целого:

$$ОПС = \frac{Показатель,\ характеризующий\ часть\ совокупности}{Показатель\ по\ всей\ совокупности\ в\ целом}$$

Выражается относительный показатель структуры в долях единицы или в процентах. Рассчитанные величины, соответственно называемые долями или удельными весами, показывают, какой долей обладает или какой удельный вес имеет та или иная часть в общем итоге.

Рассмотрим структуру валового внутреннего продукта РФ в 1 квартале 2003г. (табл. 5.2.):

Таблица 5.2.

Структура валового внутреннего продукта РФ в 1 квартале 2003г.

	Объем	
	млрд.руб.	% к итогу
ВВП - всего	2893	100
в том числе:		
- производство товаров	917	31,7
- производство услуг	1635	56,5
- чистые налоги на продукты	341	11,8

Рассчитанные в последней графе данной таблицы проценты представляют собой относительные показатели структуры (в данном случае - удельные веса). Сумма всех удельных весов всегда должна быть строго равна 100% или 1.

Относительный показатель координации представляет собой отношение одной части совокупности к другой части этой же совокупности:

$$\text{ОПК} = \frac{\text{Показатель, характеризующий i - ую часть совокупности}}{\begin{array}{c}\text{Показатель, характеризующий часть совокупности,}\\ \text{выбранную в качестве базы сравнения}\end{array}}$$

При этом в качестве базы сравнения выбирается та часть, которая имеет наибольший удельный вес или является приоритетной с экономической, социальной или какой-либо другой точки зрения. В результате получают, во сколько раз данная часть больше базисной или сколько процентов от нее составляет, или сколько единиц данной структурной части приходится на 1 единицу (иногда - на 100, 1000 и т.д. единиц) базисной структурной части. Так, на основе данных приведенной выше таблицы 3.2 мы можем вычислить, что на каждый рубль произведенных товаров приходится 1,8 руб. произведенных услуг $(\frac{1635}{917})$ и 0,4 руб. чистых налогов на продукты $(\frac{341}{917})$.

Относительный показатель интенсивности характеризует степень распространения изучаемого процесса или явления и представляет собой отношение исследуемого показателя к размеру присущей ему среды:

$$ОПИ = \frac{\text{Показатель, характеризующий явление А}}{\text{Показатель, характеризующий среду распространиния явления А}}$$

Данный показатель получают сопоставлением уровней двух взаимосвязанных в своем развитии явлений. Поэтому, наиболее часто он представляет собой именованную величину, но может быть выражен и в процентах, промилле, продецимилле.

Обычно относительный показатель интенсивности рассчитывается в тех случаях, когда абсолютная величина оказывается недостаточной для формулировки обоснованных выводов о масштабах явления, его размерах, насыщенности, плотности распространения. Так, например, для определения уровня обеспеченности населения легковыми автомобилями рассчитывается число автомашин, приходящихся на 100 семей, для определения плотности населения рассчитывается число людей, приходящихся на 1 кв.км.

Так, по данным социальной статистики на конец 2003 г. общая численность безработных в РФ составляла 6,1 млн. чел., а экономически активное население – 70,9 млн. чел. Отсюда следует, что уровень безработицы составлял 8,6% $(\frac{6,1}{70,9} \cdot 100\%)$.

Разновидностью относительных показателей интенсивности являются **относительные показатели уровня экономического развития**, характеризующие производство продукции в расчете на душу населения и играющие важную роль в оценке развития экономики государства или региона. Так как объемные показатели производства продукции по своей природе являются интервальными, а показатель численности населения - моментным, в расчетах используют среднюю за период численность населения (предположим, среднегодовую).

Например, рассматривая лишь абсолютный размер ВВП России в 1 квартале 2003 года (2893 млрд. руб.), трудно оценить или "почувствовать" эту величину. Для того, чтобы на основе данной цифры сделать вывод об уровне развития экономики, необходимо сопоставить ее со среднеквартальной численностью населения страны (145,2 млн.чел), которая в простейшем случае рассчитывается как полусумма численности населения на начало и на конец квартала. В результате квартальный размер ВВП на душу населения составит 19,9 тыс. руб. $\left(\frac{2893000 \mathit{млн.руб}}{145{,}2 \mathit{млн.чел}}\right)$.

Относительный показатель сравнения представляет собой соотношение одноименных абсолютных показателей, характеризующих разные объекты (предприятия, фирмы, районы, области, страны и т.п.):

$$\text{ОПС} = \frac{\text{Показатель, характеризующий объект А}}{\text{Показатель, характеризующий объект В}}$$

Для выражения данного показателя могут использоваться как коэффициенты, так и проценты.

Например, согласно официальным статистическим данным, инвестиции в основной капитал в РФ в 2002 г. за счет средств федерального бюджета составили 81,6 млрд.руб., бюджетов субъектов Федерации и местных бюджетов – 184,5 млрд. руб., средств предприятий – 653,1 млрд.руб. Таким образом можно сделать вывод, что инвестиции за счет средств предприятий в 8 раз превышали инвестиции из средств федерального бюджета и в 3,5 раза превышали инвестиции из бюджетов субъектов Федерации и местных бюджетов.

5.3. Средние показатели

Наиболее распространенной формой статистических показателей, используемой в экономических исследованиях, является средняя величина, представляющая собой обобщенную количественную характеристику признака в статистической совокупности в конкретных условиях места и времени. Показатель в форме средней величины выражает типичные черты и дает обобщающую характеристику однотипных явлений по одному из варьирующих признаков. Он отражает уровень этого признака, отнесенный к единице совокупности. Широкое применение средних объясняется тем, что они имеют ряд положительных свойств, делающих их незаменимым инструментом анализа явлений и процессов в экономике.

Важнейшее свойство средней величины заключается в том, что она отражает то общее, что присуще всем единицам исследуемой совокупности. Значения признака отдельных единиц совокупности колеблются в ту или иную сторону под влиянием множества факторов, среди которых могут быть как основные, так и случайные. Например, курс акций корпорации в основном определяется финансовыми результатами ее деятельности. В то же время, в отдельные дни и на отдельных биржах эти акции в силу сложившихся обстоятельств могут продаваться по более высокому или заниженному курсу. Сущность средней в том и заключается, что в ней взаимопогашаются отклонения значений признака отдельных единиц совокупности, обусловленные действием случайных факторов, и учитываются изменения, вызванные действием факторов основных. Это позволяет средней отражать типичный уровень признака и абстрагироваться от индивидуальных особенностей, присущих отдельным единицам.

Типичность средней непосредственным образом связана с однородностью статистической совокупности. Средняя величина только тогда будет отражать типичный уровень признака, когда она рассчитана по качественно однородной совокупности. Так, если мы рассчитаем средний курс по акциям всех предприятий, реализуемых в данный день на данной бирже, то получим фиктивную среднюю. Это будет объясняться тем, что используемая для расчета совокупность является крайне неоднородной. В этом и подобных случаях метод средних используется в сочетании с методом группировок: если совокупность неоднородна - общие средние должны быть заменены или дополнены групповыми средними, т.е. средними, рассчитанными по качественно однородным группам.

Категорию средней можно раскрыть через понятие ее **определяющего свойства**. Согласно этому понятию средняя, являясь обобщающей характеристикой всей совокупности, должна ориентироваться на определенную величину, связанную со всеми единицами этой совокупности. Эту величину можно представить в виде функции:

$$f\,(x1,\ x2,\ ...,\ xn) \tag{5.1.}$$

Так как данная величина, в большинстве случаев, отражает реальную экономическую категорию, понятие определяющего свойства средней иногда заменяют понятием определяющего показателя.

Если в приведенной выше функции все величины x1, x2, ..., xn заменить их средней величиной $\bar{x}$, то значение этой функции должно остаться прежним:

$$f\,(x1,\ x2,\ ...,\ xn)=f(\bar{x},\ \bar{x},\ \ldots,\bar{x}) \tag{5.2.}$$

Исходя из данного равенства и определяется средняя. На практике определить среднюю во многих случаях можно через **исходное соотношение средней** (ИСС) или ее логическую формулу:

$$\text{ИСС} = \frac{\text{Суммарное значение или объем осредняемого признака}}{\text{Число единиц или объем совокупности}}$$

Так, например, для расчета средней заработной платы работников предприятия необходимо общий фонд заработной платы разделить на число работников:

$$\text{ИСС} = \frac{\text{Фонд заработной платы}\quad(\text{тыс. руб.})}{\text{Число работников}\quad(\text{чел})}$$

Числитель исходного соотношения средней представляет собой определяющий показатель. Для средней заработной платы таким определяющим показателем является фонд заработной платы. Независимо от того, какой первичной информацией мы располагаем - известен ли нам общий фонд заработной платы или заработная плата и численность работников, занятых на отдельных должностях, или какие-либо другие исходные данные - в любом случае среднюю заработную плату можно получить только через данное исходное соотношение средней.

Для каждого показателя, используемого в экономическом анализе, можно составить только одно истинное исходное соотношение для расчета средней. Если, например, требуется рассчитать средний размер вклада в банке, то исходное соотношение будет следующим:

$$\text{ИСС} = \frac{\text{Сумма всех вкладов (тыс. руб.)}}{\text{Число вкладов}}$$

Если же необходимо определить среднюю процентную ставку по кредитам, выданным на один и тот же срок, то потребуется следующее исходное соотношение:

$$\text{ИСС} = \frac{\text{Общая сумма выплат по процентам (из расчета за год, тыс. руб.)}}{\text{Общая сумма предоставленных кредитов (тыс. руб.)}}$$

Однако от того, в каком виде представлены исходные данные для расчета средней, зависит, каким именно образом будет реализовано ее исходное соотношение. В каждом конкретном случае для реализации исходного соотношения потребуется одна из следующих форм средней величины:

- средняя арифметическая,
- средняя гармоническая,
- средняя геометрическая,
- средняя квадратическая, кубическая и т.д.

Перечисленные средние объединяются в общей формуле **средней степенной** (при различной величине k):

$$\bar{x} = \sqrt[k]{\frac{\sum x_i^k f_i}{\sum f_i}} \qquad i=\overline{1,n}$$

где x_i - i-ый вариант осредняемого признака ($i=\overline{1,n}$)
f_i - вес i-го варианта.

Помимо степенных средних в экономической практике также используются средние структурные, среди которых наиболее распространены мода и медиана. При осреднении уровней динамических рядов применяются различные виды средней хронологической.

Наиболее распространенным видом средних величин является *средняя арифметическая*, которая, как и все средние, в зависимости от характера имеющихся данных может быть простой или взвешенной. Эта форма средней используется в тех случаях, когда расчет осуществляется по несгруппированным данным.

Предположим, шесть торговых предприятий фирмы имеют следующий объем товарооборота за месяц:

Торговое предприятие	1	2	3	4	5	6
Товарооборот (млн.руб.)	25	18	27	32	15	21

Для того, чтобы определить средний месячный товарооборот в расчете на одно предприятие, необходимо воспользоваться следующим исходным соотношением:

$$\text{ИСС} = \frac{\text{Общий объем товарооборота} \quad (\text{млн.руб.})}{\text{Число торговых центров}}$$

Используя приведенные в предыдущем параграфе условные обозначения, запишем формулу данной средней:

$$\bar{x} = \frac{x_1 + x_2 + \ldots + x_n}{n} = \frac{\sum x_i}{n} \qquad (5.3.)$$

С учетом имеющихся данных получим:

$$\bar{x} = \frac{25 + 18 + 27 + 32 + 15 + 21}{6} = 23 \text{ млн.руб.}$$

В данном случае мы использовали формулу средней арифметической простой (невзвешенной).

Средняя арифметическая взвешенная. При расчете средних величин отдельные значения осредняемого признака могут повторяться, встречаться по несколько раз. В подобных случаях расчет средней производится по сгруппированным данным или вариационным рядам, которые могут быть дискретными или интервальными.

Рассмотрим следующий условный пример:

Таблица 5.3.

Сделки по акциям эмитента «Х» за торговую сессию

Сделка	Количество проданных акций, шт.	Курс продажи, руб.
1	700	420
2	200	440
3	950	410

Определим по данному дискретному вариационному ряду средний курс продажи 1 акции, что можно сделать, только используя следующее исходное соотношение:

$$\text{ИСС} = \frac{\text{Общая сумма сделок (руб.)}}{\text{Количество проданных акций (шт.)}}$$

Чтобы получить общую сумму сделок необходимо по каждой сделке курс продажи умножить на количество проданных акций и полученные произведения сложить. В конечном итоге мы будем иметь следующий результат:

$$\bar{x} = \frac{420 \times 700 + 440 \times 200 + 410 \times 950}{700 + 200 + 950} = \frac{771500}{1850} = 417{,}03 \text{ руб.}$$

Расчет среднего курса продажи произведен по формуле средней арифметической взвешенной:

$$\bar{x} = \frac{\sum x_i f_i}{\sum f_i} \qquad (5.4.)$$

В отдельных случаях веса могут быть представлены не абсолютными величинами, а относительными (в процентах или долях единицы). Так, в приведенном выше примере количество проданных в ходе каждой сделки акций соответственно составляет 37,8% (0,378); 10,8% (0,108) и 51,4% (0,514) от их общего числа. Тогда, с учетом несложного преобразования формулы (5.4.) получим:

$$\bar{x} = \sum (x_i \frac{f_i}{\Sigma f_i}) \qquad (5.5.)$$

или

$$\bar{x} = 420 \times 0{,}378 + 440 \times 0{,}108 + 410 \times 0{,}514 = 417{,}03 \text{ руб.}$$

На практике наиболее часто встречаемая при расчете средних ошибка заключается в игнорировании весов в тех случаях, когда эти веса в действительности необходимы. Предположим, имеются следующие данные:

Таблица 5.4.

Себестоимость продукции «Z»

Предприятие	Себестоимость единицы продукции, руб.
1	37
2	39

Можно ли по имеющимся данным определить среднюю себестоимость данной продукции по двум предприятиям, вместе взятым? Можно, но только в том случае, когда объемы производства данной продукции на двух предприятиях совпадают. Тогда средняя себестоимость составит 38,0 руб. (доказательство этого правила будет приведено ниже.). Однако на первом предприятии за рассматриваемый период может быть произведено, к примеру, 50 единиц продукции, а на втором - 700 единиц. Тогда для расчета средней себестоимости потребуется уже средняя арифметическая взвешенная:

$$\bar{x} = \frac{37 \times 50 + 39 \times 700}{50 + 700} = 38{,}9\text{руб.}$$

Общий вывод заключается в следующем: использовать среднюю арифметическую невзвешенную можно только тогда, когда точно установлено отсутствие весов или их равенство.

При расчете средней по **интервальному вариационному ряду** для выполнения необходимых вычислений от интервалов переходят к их серединам. Рассмотрим следующий пример:

Таблица 5.5.

Распределение сотрудников предприятия по возрасту

Возраст (лет)	Число сотрудников (чел.)
до 25	8
25 - 30	32
30 - 40	68
40 - 50	49
50 - 60	21
60 и более	3
Итого:	181

Для определения среднего возраста персонала найдем середины возрастных интервалов. При этом величины открытых интервалов (первого и последнего) условно приравниваются к величинам интервалов, примыкающих к ним (второго и предпоследнего). С учетом этого середины интервалов будут следующими:

22, 5 27,5 35,0 45,0 55,0 65,0

Используя среднюю арифметическую взвешенную, определим средний возраст работников данного предприятия:

$$\bar{x} = \frac{22{,}5 \times 8 + 27{,}5 \times 32 + 35 \times 68 + 45 \times 49 + 55 \times 21 + 65 \times 3}{8 + 32 + 68 + 49 + 21 + 3} = 38{,}6$$ года.

Свойства средней арифметической. Средняя арифметическая обладает некоторыми математическими свойствами, более полно раскрывающими ее сущность и в ряде случаев используемыми при ее расчете. Рассмотрим эти свойства:

1. Произведение средней на сумму частот равно сумме произведений отдельных вариантов на соответствующие им частоты:

$$\bar{x}\sum f_i = \sum x_i f_i \qquad (5.6.)$$

Действительно, если мы обратимся к приведенному выше примеру расчета среднего курса продажи акций (табл. 5.1.), то получим следующее равенство (за счет округления среднего курса правая и левая части равенства в данном случае будут несколько отличаться):

417,03×1850=420×700+440×200+410×950

2. Сумма отклонений индивидуальных значений признака от средней арифметической равна нулю:

(5.7.)

Для нашего примера:

(420-417,03)×700+(440-417,03)×200+(410-417,03)×950 ≈ 0

Математическое доказательство данного свойства сводится к следующему:

3. Сумма квадратов отклонений индивидуальных значений признака от средней арифметической меньше, чем сумма квадратов их отклонений от любой другой произвольной величины С:

(5.8.)

Следовательно, сумма квадратов отклонений индивидуальных значений признака от произвольной величины С больше суммы квадратов их отклонений от своей средней на величину

или

На использовании этого свойства базируется расчет центральных моментов, представляющих собой характеристики вариационного ряда при $C = \bar{x}$: [1]

,

где к определяет порядок момента (центральный момент второго порядка представляет собой дисперсию).

4. Если все осредняемые варианты уменьшить или увеличить на постоянное число А, то средняя арифметическая соответственно уменьшится или увеличится на ту же величину:

(5.9.)

Так, если все курсы продажи акций увеличить на 15 руб., то средний курс также увеличится на 15 руб.:

$$\bar{x} = \frac{435 \times 700 + 455 \times 200 + 425 \times 950}{1850} = 417{,}03 + 15 = 432{,}03$$ руб.

5. Если все варианты значений признака уменьшить или увеличить в А раз, то средняя также соответственно увеличится или уменьшится в А раз:

$$\frac{\sum \frac{x_i}{A} f_i}{\sum f_i} = \frac{\frac{1}{A} \sum x_i f_i}{\sum f_i} = \frac{1}{A} \bar{x} \qquad (5.10.)$$

[1] При С=0 получают начальные моменты (начальный момент 1-го порядка - средняя арифметическая и т.д.).

Предположим, курс продажи в каждом случае возрастет в 2 раза. Тогда и средний курс также увеличится на 100%:

$$\bar{x} = \frac{420 \times 2 \times 700 + 440 \times 2 \times 200 + 410 \times 2 \times 950}{1850} = 417{,}03 \times 2 = 834{,}06$$ руб.

6. Если все веса уменьшить или увеличить в А раз, то средняя арифметическая от этого не изменится:

$$\frac{\sum x_i \frac{f_i}{A}}{\sum \frac{f_i}{A}} = \frac{\frac{1}{A}\sum x_i f_i}{\frac{1}{A}\sum f_i} = \bar{x} \qquad (5.11.)$$

Так, в нашем примере удобнее было бы рассчитывать среднюю, предварительно поделив все веса на 100:

$$\bar{x} = \frac{420 \times 7 + 440 \times 2 + 410 \times 9{,}5}{7 + 2 + 9{,}5} = \frac{7715}{18{,}5} = 417{,}03$$ руб.

Исходя из данного свойства, можно заключить, что если все веса равны между собой, то расчеты по средней арифметической взвешенной и средней арифметической невзвешенной приведут к одному и тому же результату.

Кроме средней арифметической при расчете статистических показателей могут использоваться и другие виды средних. Однако, в каждом конкретном случае, в зависимости от характера имеющихся данных, существует только одно истинное среднее значение показателя, являющееся следствием реализации его исходного соотношения.

Средняя гармоническая взвешенная используется, когда известен числитель исходного соотношения средней, но неизвестен его знаменатель. Рассмотрим расчет средней урожайности, являющейся одним из основных показателей эффективности производства в агробизнесе:

Таблица 5.6.

Валовой сбор и урожайность сельскохозяйственной культуры «Y» по районам области

Район	Валовый сбор, тыс. тонн	Урожайность, ц/га
А	36	13
Б	53	9
В	29	15
Г	78	8
Д	20	17

Средняя урожайность любой сельскохозяйственной культуры в среднем по нескольким территориям, агрофирмам, фермерским хозяйствам и т.п. может быть определена только на основе следующего исходного соотношения:

$$\text{ИСС} = \frac{\text{Общий валовой сбор} \quad (\text{тыс. ц.})}{\text{Общая посевная площадь} \quad (\text{тыс. га})}$$

Общий валовой сбор мы получим простым суммированием валового сбора по районам. Данные же о посевной площади отсутствуют, но их можно получить, разделив валовой сбор по каждого района на урожайность. С учетом этого определим искомую среднюю, предварительно переведя для сопоставимости тонны в центнеры:

$$\bar{x} = \frac{360+530+290+780+200}{\frac{360}{13}+\frac{530}{9}+\frac{290}{15}+\frac{780}{8}+\frac{200}{17}} = \frac{2160}{215{,}2} = 10{,}0 \text{ ц/га}$$

Таким образом, общая посевная площадь данной культуры в целом по области составляла 215,2 тыс.га, а средняя урожайность - 10,0 ц с одного гектара.

В данном случае расчет произведен по формуле средней гармонической взвешенной:

$$\bar{x} = \frac{\sum w_i}{\sum \frac{w_i}{x_i}}$$, где wi=xifi (5.12.)

Данная формула используется для расчета средних показателей не только в статике, но и в динамике, когда известны индивидуальные значения признака и веса W за ряд временных интервалов.

Средняя гармоническая невзвешенная. Эта форма средней, используемая значительно реже, имеет следующий вид:

$$\bar{x} = \frac{n}{\sum \frac{1}{x_i}} \qquad (5.13.)$$

Для иллюстрации области ее применения воспользуемся упрощенным условным примером. Предположим, в фирме, специализирующейся на торговле по почте на основе предварительных заказов, упаковкой и отправкой товаров занимаются два работника. Первый из них на обработку одного заказа затрачивает 5 мин., второй - 15 мин. Каковы средние затраты времени на 1 заказ, если общая продолжительность рабочего времени у работников равна?

На первый взгляд, ответ на этот вопрос заключается в осреднении индивидуальных значений затрат времени на 1 заказ, т.е. (5+15):2=10, мин. Проверим обоснованность такого подхода на примере одного часа работы. За этот час первый работник обрабатывает 12 заказов (60:5), второй - 4 заказа (60:15), что в сумме составляет 16 заказов. Если же заменить индивидуальные значения их предполагаемым средним значением, то общее число обработанных обоими работниками заказов в данном случае уменьшится:

$$\frac{60}{10}+\frac{60}{10}=12$$ заказов.

Подойдем к решению через исходное соотношение средней. Для определения средних затрат времени необходимо общие затраты времени за любой интервал (например, за час) разделить на общее число обработанных за этот интервал двумя работниками заказов:

$$\bar{x}=\frac{60+60}{\frac{60}{5}+\frac{60}{15}}=\frac{1+1}{\frac{1}{5}+\frac{1}{15}}=\frac{2}{0{,}200+0{,}067}=7{,}5$$ мин.

Если теперь мы заменим индивидуальные значения их средней величиной, то общее количество обработанных за час заказов не изменится:

$$\frac{60}{7{,}5}+\frac{60}{7{,}5}=16$$ заказов.

Подведем итог: средняя гармоническая невзвешенная может использоваться вместо взвешенной в тех случаях, когда значения wi для единиц совокупности равны (в рассмотренном примере рабочий день у сотрудников одинаковый).

Средняя геометрическая. Еще одной формулой, по которой может осуществляться расчет среднего показателя, является средняя геометрическая:

$$\bar{x}=\sqrt[к]{x_1\cdot x_2\cdot x_3\cdot\ldots\cdot x_к}=\sqrt[к]{\Pi x_i}$$ - невзвешенная

(5.14.)

$$\bar{x}=\sqrt[\Sigma m]{x_1^{m_1}\cdot x_2^{m_2}\cdot x_3^{m_3}\cdot\ldots\cdot x_к^{m_к}}=\sqrt[\Sigma m]{\Pi x_i^{m_i}}$$ - взвешенная

Наиболее широкое применение этот вид средней получил в анализе динамики для определения среднего темпа роста, что будет рассмотрено в соответствующей главе.

Средняя квадратическая. В основе вычислений ряда сводных расчетных показателей лежит средняя квадратическая:

$$\bar{x} = \sqrt{\frac{\sum x_i^2}{n}}$$ - невзвешенная

(5.15.)

$$\bar{x} = \sqrt{\frac{\sum x_i^2 f_i}{\sum f_i}}$$ - взвешенная

Наиболее широко этот вид средней используется при расчете показателей вариации.

В статистическом анализе также применяются степенные средние 3-го порядка и более высоких порядков.

5.4.Структурные средние

Наиболее часто используемыми в экономической практике структурными средними являются мода и медиана. **Мода** представляет собой значение изучаемого признака, повторяющееся с наибольшей частотой. **Медианой** называется значение признака, приходящееся на середину ранжированной (упорядоченной) совокупности.

Главное свойство медианы заключается в том, что сумма абсолютных отклонений значений признака от медианы меньше, чем от любой другой величины:

$$\sum |x_i - M_e| = \min$$

Рассмотрим определение моды и медианы по **несгруппированным данным**.

Предположим, что 9 торговых фирм города реализуют товар А по следующим оптовым ценам (тыс.руб.).

4,4 4,3 4,4 4,5 4,3 4,3 4,6 4,2 4,6

Так как чаще всего встречается цена 4,3 тыс.руб., то она и будет модальной.

Для определения медианы необходимо провести ранжирование:

4,2 4,3 4,3 4,3 4,4 4,4 4,5 4,6 4,6

Центральной в этом ряду является цена 4,4 тыс.руб., следовательно, данная цена и будет медианой. Если ранжированный ряд включает четное число единиц, то медиана определяется как средняя из двух центральных значений.

Если мода отражает типичный, наиболее распространенный вариант значения признака, то медиана практически выполняет функции средней для неоднородной, не подчиняющейся нормальном закону распределения совокупности. Она также используется в тех случаях, когда средняя не позволяет объективно оценить исследуемую совокупность вследствие сильного влияния максимальных и минимальных значений. Проиллюстрируем познавательное значение медианы следующим примером.

Допустим, нам необходимо дать характеристику среднего дохода группы людей, насчитывающей 100 человек, из которых 99 имеют доходы в интервале от 100 до 1000 долл. в месяц, а месячные доходы последнего составляют 50000 долл.:

№ п/п	1	2	3	4	...	50	51	...	99	100
Доход (долл.)	100	104	104	107	...	162	164	...	200	50000

Если мы воспользуемся средней арифметической, то получим средний доход, равный примерно 600-700 долл., который не только в несколько раз меньше дохода 100-го человека, но и имеет мало общего с доходами остальной части группы. Медиана же, равная в данном случае 163 долл., позволит дать объективную характеристику уровня доходов 99% данной совокупности людей.

Рассмотрим определение моды и медианы по **сгруппированным данным** (рядам распределения).

Предположим, распределение торговых предприятий города по уровню розничных цен на товар А имеет следующий вид:

Цена, руб.	Число торговых предприятий
52	12
53	48
54	56
55	60
56	14
Всего	190

Определение моды по дискретному вариационному ряду не составляет большого труда - наибольшую частоту (60 предп.) имеет цена 55 руб., следовательно она и является модальной.

Для определения медианного значения признака по следующей формуле находят номер медианной единицы ряда:

$$N_{me} = \frac{n+1}{2} \tag{5.16}$$

где n - объем совокупности.

В нашем случае $N_{me} = \frac{190+1}{2} = 95,5$.

Полученное дробное значение, всегда имеющее место при четном числе единиц в совокупности, указывает, что точная середина находится между 95 и 96 предприятиями. Необходимо определить, в какой группе находятся предприятия с этими порядковыми номерами. Это можно сделать, рассчитав накопленные частоты. Очевидно, что магазинов с этими номерами нет в первой группе, где всего лишь 12 торговых предприятий, нет их и во второй группе (12+48=60). 95-ое и 96-ое предприятия находятся в третьей группе (12+48+56=116) и, следовательно, медианой является цена 54 руб.

В отличие от дискретных вариационных рядов определение моды и медианы по **интервальным рядам** требует проведения определенных расчетов на основе следующих формул :

$$M_o = x_o + i \times \frac{(f_{M_o} - f_{M_o-1})}{(f_{M_o} - f_{M_o-1}) + (f_{M_o} - f_{M_o+1})} \qquad (5.17)$$

где Xo - нижняя граница модального интервала (модальным называется интервал, имеющий наибольшую частоту);

i - величина модального интервала;

fMo - частота модального интервала;

fMo-1 - частота интервала, предшествующего модальному;

fMo+1 - частота интервала, следующего за модальным.

и

$$M_e = x_0 + i \times \frac{\frac{1}{2}\sum f_i - s_{M_{e-1}}}{f_{M_e}} \qquad (5.18)$$

где Xo - нижняя граница медианного интервала (медианным называется первый интервал, накопленная частота которого превышает половину общей суммы частот);

i - величина медианного интервала:

Sme-1 - накопленная частота интервала, предшествующего медианному;

fMe - частота медианного интервала.

Проиллюстрируем применение этих формул, используя данные таблицы 5.7.

Информация, подобная представленной в этой таблице, необходима для получения четкого представления о покупательной способности населения страны или региона, для оценки эластичности спроса и, в конечном итоге, для выбора того или иного метода ценообразования и обоснования окончательной цены на товар.

Таблица 5.7.

Распределение населения региона по уровню среднедушевого денежного дохода

Среднедушевой денежный доход (в среднем за месяц), руб.	Удельный вес населения, %
400 и менее	2,4
400 - 500	15,4
500 - 600	20,1
600 -700	17,2
700 - 800	12,8
800 - 900	9,2
900 - 1000	6,5
1000 - 1100	4,5
1100 - 1200	3,2
1200 - 1300	2,3
свыше 1300	6,4
Всего	100,0

Интервал с границами 500 - 600 в данном распределении будет модальным, так как он имеет наибольшую частоту. Использую формулу (5.17), определим моду:

$$M_0 = 500 + 100 \times \frac{20{,}1 - 15{,}4}{(20{,}1 - 15{,}4) + (20{,}1 - 17{,}2)} = 562 \text{ руб.}$$

Для определения медианного интервала необходимо определять накопленную частоту каждого последующего интервала до тех пор, пока она не превысит 1/2 суммы накопленных частот (в нашем случае - 50%):

Интервал	Накопленная частота, %
400 и менее	2,4
400 - 500	17,8
500 - 600	37,9
600 - 700	55,1

Мы определили, что медианным является интервал с границами 600 - 700. Определим медиану:

$$M_e = 600 + 100 \times \frac{50,0 - 37,9}{17,2} = 670 \text{ руб.}$$

Соотношение моды, медианы и средней арифметической указывает на характер распределения признака в совокупности, позволяет оценить его асимметрию. Если $Mo<Me<\overline{X}$ - имеет место правосторонняя асимметрия, при $\overline{X}<Me<Mo$ следует сделать вывод о левосторонней асимметрии ряда.

На основе полученных в последнем примере значений структурных средних можно заключить, что наиболее распространенным, типичным является среднедушевой доход порядка 560 руб. в месяц. В то же время, более половины населения располагает доходом свыше 670 руб. при среднем уровне 735 руб. (средняя арифметическая взвешенная). Из соотношения этих показателей следует вывод о правосторонней асимметрии распределения населения по уровню среднедушевых денежных доходов, что позволяет предполагать о достаточной емкости рынка дорогих товаров повышенного качества и товаров престижной группы.

Глава 6. Анализ вариации

6.1. Основные показатели вариации

Информация о средних уровнях исследуемых показателей обычно бывает недостаточной для полного анализа изучаемого процесса или явления. Иногда совершенно непохожие по своему внутреннему строению совокупности могут иметь равные средние величины. Поэтому для более детального изучения того или иного явления необходимо учитывать разброс или вариацию значений отдельных единиц совокупности. Измерение вариации признаков имеет как теоретическое, так и практическое значение.

Так, например, для выявления наиболее стабильно работающего коллектива или предприятия наравне с другими показателями рассчитывают и основные показатели вариации. Эти показатели дают возможность количественно определить размеры устойчивости производительности труда, уровня квалификации, цен на основные виды выпускаемой продукции и т.п. Измерение размеров вариации такого показателя как « выполнение работ в срок» имеет важное значение для принятия решений заказчиками и инвесторами, т.к. ситуация, в которой присутствует изменчивость признака, часто содержит риск. Особое значение показатели вариации приобретают в анализе рынка ценных бумаг, где мера колеблемости отождествляется с мерой рискованности вложения денежных средств.

Основными показателями, характеризующими вариацию, являются:

- - размах,
- - дисперсия,
- - среднее квадратическое отклонение,
- - коэффициент вариации.

Для иллюстрации расчетов этих показателей воспользуемся следующими данными:

Имеются данные о продаже основных марок холодильников:

Таблица.6.1.

Модель	Цена ($)	Объем продаж (шт.)
Siemens	1000	30
Bosch	800	26
AEG Santo	900	24
Miele KF	1200	30
Gorenje	870	20
Haier	570	23
Samsung	760	30
Zanussi	700	20
Daewoo	460	20
Beko	650	25
Candy	480	20
Whirpool	470	21

Простейшим показателем, уже использованным выше при группировке данных, является **размах вариации.** Он представляет собой разность максимального и минимального значений признака:

$$R = X\ max - X\ min$$
$$R = 1200 – 460 = 740\ \$$$

Этот показатель служит незаменимой мерой разброса экстремальных значений признака. Кроме характеристики границ разброса признака, размах вариации может быть использован для выявления ошибок. При наличии очень больших (или очень малых) ошибочно записанных значений признака размах вариации сразу резко возрастает, что требует проверки и корректировки исходных данных.

Недостатком данного показателя является то, что он оценивает только границы варьирующего признака и не отражает его колеблемость внутри этих границ. Вследствие этого размах вариации может неправильно характеризовать общую колеблемость признака.

Этого недостатка лишен другой показатель - **дисперсия**, рассчитываемый как средний квадрат отклонений значений признака от их средней величины. Между индивидуальными отклонениями от средней и колеблемостью признака существует прямая зависимость: чем сильнее колеблемость признака, тем больше отклонения его значений от средней величины и менее устойчив изучаемый показатель.

Как и средняя величина этот показатель может быть рассчитан в двух формах: взвешенной и невзвешенной:

$$\sigma^2 = \frac{\sum (x_I - \bar{x})^2}{n}$$ невзвешенная форма

$$\sigma^2 = \frac{\sum (x_I - \bar{x})^2 f_i}{\sum f_i}$$ взвешенная форма

где x_i - отдельные значения признака

$\bar{x}$ - общая средняя

f_i - вес варианта признака в общей совокупности.

По приведенным выше данным определим средневзвешенную цену холодильника и рассчитаем дисперсию:

$$\bar{x} = \frac{1000x30 + 800x26 + 900x24 + ... + 470x21}{30 + 26 + 24 + ... + 21} = 763\$$$

$$\sigma^2 = \frac{(1000 - 763)30 + (800 - 763)26 + ... + (470 - 763)21}{30 + 26 + 24 + ... + 20 + 21} = 73500{,}12$$

Дисперсию в отдельных случаях удобнее рассчитывать по другой формуле:

$\sigma^2 = \overline{x^2} - (\overline{x})^2$, т.е. дисперсия равна разности средней из квадратов индивидуальных значений признака и квадратом средней величины.

Эту формулу можно представить иначе:

$$\sigma^2 = \frac{\sum x_i^2}{n} - \left(\frac{\sum x_i}{n}\right)^2$$ - невзвешенная форма

$$\sigma^2 = \frac{\sum x_i^2 f_i}{\sum f_i} - \left(\frac{\sum x\ f_i}{\sum f_i}\right)^2$$ - взвешенная форма

Следует отметить, что дисперсия еще не дает представления об однородности совокупности, и этому показателю трудно дать экономическую интерпретацию, т.к он рассчитан в квадратных единицах. Поэтому следующим шагом в исследовании однородности совокупности является расчет среднего квадратического отклонения, показывающего на сколько в среднем отклоняются конкретные варианты признака от его среднего значения. Оно определяется как квадратный корень из дисперсии и имеет ту же размерность что и изучаемый признак:

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

$$\sigma = \sqrt{\frac{\sum \left(x_i - \bar{x}\right)^2 f_i}{\sum f_i}}$$ - взвешенная форма

В нашем примере среднее квадратическое отклонение равно:

$\sigma = \sqrt{73500{,}12} = 271{,}1$ $

Таким образом, цена каждой марки холодильника отклоняется от средней цены в среднем на 271,1$

Рассмотренные показатели позволяют получить абсолютное значение вариации признака. Однако для сравнения разных совокупностей с точки зрения устойчивости какого-либо одного признака или для определения однородности совокупности рассчитывают относительные показатели.

Эти показатели вычисляются как отношение размаха вариации, среднего линейного отклонения или среднего квадратического отклонения к средней арифметической или медиане. Чаще всего эти показатели выражаются в процентах.

Коэффициент осцилляции (V_r):

$$V_r = \frac{R}{\bar{x}} 100\%$$

Линейный коэффициент вариации ($V_{\bar{d}}$):

$$V_{\bar{d}} = \frac{\bar{d}}{\bar{x}} 100\%$$

Наиболее распространенным показателем является **коэффициент вариации**:

$$V_\sigma = \frac{\sigma}{\bar{x}} x 100 \%$$

Определим значение этого показателя по данным таблицы 1:

$$V_\sigma = \frac{271{,}1}{763} x 100\% = 35{,}5\%$$

Рассчитанная величина свидетельствует о неоднородности цен на холодильники, т.к. однородной совокупность считается, если коэффициент вариации меньше 33% (для распределений близких к нормальному).

Следует отметить, что коэффициент вариации может быть более 100%, что, в частности, может быть при наличии значений сильно отличающихся от средней величины. Такой результат означает, что в исследуемой совокупности сильна вариация признаков по отношению к средней величине.

Если исследуется вариация альтернативных признаков, т.е.признаков, которыми одни единицы совокупности обладают, а другие – нет, то дисперсия альтернативного признака определяется по формуле :

$$\sigma^2 = pq$$, где

p – доля единиц, обладающих данным признаком,
q – доля единиц не обладающих данным признаком.

Максимальное значение дисперсии доли равно 0,25 (когда p=q=0,5).

Информативность показателей вариации повышается, если они рассчитываются для целей сравнительного анализа. При этом показатели рассчитанные по одной совокупности сопоставляются с показателями, рассчитанными по другой аналогичной совокупности или по той же самой, но относящейся к другому периоду времени. Например, исследуется динамика вариации на товары длительного пользования по месячным или ежегодным данным в одном и том же торговом предприятии или за один и тот же период времени , но по разным регионам.

6.2. Использование показателей вариации в анализе взаимосвязей

Изучая вариацию интересующего нас признака в пределах исследуемой совокупности и опираясь на общую среднюю в расчетах, трудно оценить степень воздействия на него какого-либо отдельного признака.

При проведении такого анализа исходная совокупность должна представлять собой множество единиц, каждая из которых характеризуется двумя признаками – факторным (оказывающим влияние на взаимосвязанный с ним признак) и результативным (подверженным влиянию).

Для выявления взаимосвязи исходная совокупность делится по факторному признаку на группы. Выводы о степени взаимосвязи базируются на анализе вариации результативного признака. Если статистическая совокупность разбита на группы по какому-либо признаку, то для оценки влияния различных факторов, определяющих вариацию индивидуальных значений признака, используют правило сложения дисперсий.

Общая дисперсия представляет собой сумму средней из внутригрупповой и межгрупповой и дисперсий:

$$\sigma_o^2 = \overline{\sigma^2} + \delta^2,$$ где

σ_o^2 - общая дисперсия

$\overline{\sigma^2}$ - средняя из внутригрупповых дисперсий

δ^2 - межгрупповая дисперсия

Общая дисперсия характеризует вариацию признака по всей совокупности как результат влияния всех факторов, определяющих индивидуальные различия единиц совокупности.

$$\sigma^2 = \frac{\sum (x_I - \overline{x})^2 f_i}{\sum f_i},$$

где x_i - отдельные значения признака

$\overline{x}_i$ - общая средняя варьирующего признака

f_i - вес варианта признака в общей совокупности.

Межгрупповая дисперсия характеризует вариацию, обусловленную влиянием фактора, положенного в основу группировки.

$$\delta^2 = \frac{\sum (\overline{x}_j - \overline{x})^2 n_j}{\sum f_j},$$

где $\overline{x}$ - общая средняя варьирующего признака

$\overline{x_j}$ - средняя j- ой группы

n_j - число единиц в j-ой группе ($\sum n_j = \sum f_i$)

Средняя из внутригрупповых дисперсий отражает ту часть вариации результативного признака, которая обусловлена действием всех прочих неучтенных факторов, кроме фактора, по которому осуществлялась группировка. Другими словами **внутригрупповая дисперсия** отражает случайную вариацию. Внутригрупповая дисперсия рассчитывается отдельно по каждой j-ой группе.

$$\sigma_i^2 = \frac{\sum (x_j - \overline{x_j})^2}{n_j},$$

где x_i - значение признака у отдельных элементов j-ой группы

$\overline{x_j}$ - средняя j-ой группы

n_j - число единиц j-ой группы

Для всех групп в целом вычисляется **средняя из внутригрупповых** дисперсий, взвешенных на частоты соответствующих групп по формуле:

$$\overline{\sigma^2} = \frac{\sum \sigma_j^2 n_j}{\sum f_i}$$

Взаимосвязь между тремя видами дисперсий получила название правила сложения дисперсий. Таким образом, зная два вида дисперсий всегда можно определить третий: $\sigma_o^2 = \overline{\sigma^2} + \delta^2$. Из этого равенства следует, что общая дисперсия, как правило, будет больше средней из групповых дисперсий. Это обусловлено тем, что при расчленении общей совокупности единиц на части по какому-либо признаку образуются более или менее однородные группы, в результате чего сокращается колеблемость признаков в пределах каждой группы. Это приводит к тому, что средняя из групповых дисперсий оказывается меньше дисперсии признака по всей совокупности единиц, причем разница между этими показателями будет тем больше, чем однороднее получаются группы в результате расчленения общей совокупности.

Теснота связи между факторным и результативным признаками оценивается на основе эмпирического корреляционного отношения:

$$\eta_э = \sqrt{\frac{\delta^2}{\sigma^2}}$$

Данный показатель может принимать значения от 0 до 1. Чем ближе к 1 будет его величина, тем сильнее взаимосвязь между рассматриваемыми признаками.

На следующем условном примере исследуем зависимость объема выполненных работ от формы собственности проектно-изыскательских организаций.

Таблица 6.2.

Выполнение работ проектно-изыскательскими организациями разной формы собственности

Форма собственности	Количество предприятий	Объем выполненных работ (млн.р.)	Итого
Государственная	4	10, 30, 20, 40	100
Негосударственная	6	20, 40, 60, 20, 50, 50	240
Итого	10		340

Решение:

1. Определяется средний объем выполненных работ для предприятий двух форм собственности.

$\overline{X} = \frac{340}{10} = 34$ млн.р.

2. Определяется средний объем выполненных работ для каждой формы собственности.

$\overline{X}\ госуд = \frac{100}{4} = 25$ млн.р.

$\overline{X}негосуд. = \frac{240}{6} = 40$ млн.р.

3. Рассчитывается общая и внутригрупповые (т.е. для каждой группы) дисперсии.

$$\sigma_{об}{}^{2} = \frac{(10-34)^2 + (30-34)^2 + (20-34)^2 + (40-34)^2 + (20-34)^2 + ... + (50-34)^2}{10} = 243{,}8$$

$$\sigma^2_{госуд.} = \frac{(10-25)^2 + (30-25)^2 + (20-25)^2 + (40-25)^2}{4} = 125$$

$$\sigma^2_{негосуд..} = \frac{(20-40)^2 + (40-40)^2 + (60-40)^2 + (20-40)^2 + (50-40)^2 + (50-40)^2}{6} \approx 233$$

4. Определяется средняя из внутригрупповых и межгрупповая дисперсия. Для этого расчета полученные ранее данные заносятся в таблицу.

Вспомогательная таблица

Форма собственности	Число предприятий	Средняя по группе	Внутригрупповые дисперсии
Государственная	4	25	125
Негосударственная	6	40	233
Итого	10		

- Средняя из внутригрупповых дисперсий

$$\overline{\sigma^2} = \frac{125x4 + 233x6}{10} = 189{,}8$$

- Межгрупповая дисперсия

$$\delta^2 = \frac{(25-34)^2 x4 + (40-34)^2 x6}{10} = 54$$

На последнем этапе решения задачи необходимо проверить тождество, отражающее закон сложения дисперсий:

$$54,0+189,8=243,8$$

Таким образом, можно сделать вывод о том, что объем работ, выполненных проектно-изыскательскими организациями на 22% [(54,0/243,8)х100%] зависит от фактора, положенного в основание группировки, т.е. от формы собственности., а на 78% [(189,8/243,8)х100%)] – от прочих факторов.

Вывод о том, что объем выполненных работ в гораздо большей степени зависит от каких-либо других факторов, чем от формы собственности предприятий подтверждается и величиной эмпирического корреляционного отношения:

$$\eta = \sqrt{\frac{54}{243,8}} = 0,47$$

Величина этого показателя свидетельствует о том, что зависимость объема работ от формы собственности предприятия невелика.

Глава 7. Выборочное наблюдение

7.1. Цели и этапы выборочного наблюдения

Выборочное наблюдение в настоящее время находит достаточно широкое применение в обследованиях промышленных и сельскохозяйственных предприятий, изучении цен на потребительском рынке, в обследованиях бюджетов и занятости населения. Выборочный метод является важнейшим источником информации в контроле качества продукции, в маркетинговых и социологических исследованиях.

Выборочным наблюдением называется такое несплошное обследование, при котором признаки регистрируются у отдельных единиц изучаемой статистической совокупности, отобранных с использованием специальных методов, а полученные в процессе обследования результаты с определенным уровнем вероятности распространяются на всю исходную совокупность.

Выборочное наблюдение нельзя отождествлять с несплошным обследованием вообще, так как оно является лишь одним из видов последнего, наиболее проработанным с методологической и организационной точек зрения. Помимо выборочного наблюдения несплошное обследование может осуществляться путем монографического описания, методом основного массива или на основе различных видов анкетирования, когда отсутствуют какие-либо специальные методы отбора респондентов и процент заполненных и возвращенных анкет заранее не известен.

Преимущества выборочного наблюдения заключаются в существенной экономии различного вида ресурсов, а именно:

а) финансовых средств, затрачиваемых на сбор и обработку данных, подготовку и оплату кадров;

б) материально-технических ресурсов (канцелярские товары, оргтехника, расходные материалы, транспортное обслуживание и т.п.);

в) трудовых ресурсов, привлекаемых к обследованию на всех его этапах;

г) сокращении времени, затрачиваемого как на получение первичной информации, так и на ее последующую обработку вплоть до публикации итоговых материалов.

В то же время, необходимо четко представлять, что выборочное наблюдение, как бы грамотно с методологической точки зрения оно не было организовано, всегда связано с определенными, пусть небольшими и измеряемыми ошибками. Поэтому, когда вариация регистрируемых признаков очень сильная и процент отбора для получения выборочных значений с заданной точностью достигает 20-25%, следует правильно оценить целесообразность несплошного обследования, сопоставив достаточно большие затраты всех ресурсов на такую объемную выборку и ожидаемые погрешности статистических характеристик. Вполне вероятно, что проведение сплошного обследования в подобных случаях будет более оправданным.

В то же время, при решении ряда задач выборочное наблюдение является единственно возможным способом получения необходимой информации. Так, контроль многих видов продукции связан с их порчей, потерей товарного вида, нарушением герметизации и т.п. Например, нельзя проверить каждую производимую предприятием электролампу на соблюдение требований по продолжительности горения. Нельзя проверить на соответствие стандартам каждого пакета с соком или молочной продукцией, так как это связано с вскрытием их упаковки. В подобных случаях контроль качества может осуществляться только с использованием выборочного метода.

Реализация выборочного метода базируется на понятиях генеральной и выборочной совокупностей.

Генеральной совокупностью называется вся исходная изучаемая статистическая совокупность, из которой на основе отбора единиц или групп единиц формируется **совокупность выборочная**. Поэтому генеральную совокупность также называют основой выборки.

Отбор единиц в выборочную совокупность может быть повторным или бесповторным.

При **повторном отборе** попавшая в выборку единица подвергается обследованию, т.е. регистрации значений ее признаков, возвращается в генеральную совокупность и наравне с другими единицами участвует в дальнейшей процедуре отбора. Таким образом, некоторые единицы могут попадать в выборку дважды, трижды или даже большее число раз. И при изучении выборочной совокупности они будут рассматриваться как отдельные независимые наблюдения.

Отметим, что число единиц генеральной совокупности, участвующих в отборе, при таком подходе остается постоянным. Поэтому вероятность попадания в выборку для всех единиц совокупности на протяжении всего процесса отбора также не меняется.

На практике методология повторного отбора обычно используется в тех случаях, когда объем генеральной совокупности не известен и теоретически возможно повторение единиц с уже встречавшимися значениями всех регистрируемых признаков.

Например, при проведении маркетинговых исследований мы не можем сколько-нибудь точно оценить, какое число потребителей предпочитают стиральный порошок конкретной торговой марки, сколько покупателей предпочитают делать покупки именно в данном супермаркете и т.д. Поэтому возможно повторение совершенно идентичных единиц как по причине практически неограниченных объемов совокупности, так и вследствие возможной повторной регистрации. Предположим, при проведении обследования один и тот же покупатель может дважды прийти в магазин и дважды подвергнуться обследованию.

При выборочном контроле качества продукции объем генеральной совокупности также часто не определен, так как процесс производства может осуществляться постоянно, каждый день дополняя генеральную совокупность новыми единицами - изделиями. Поэтому в выборочную совокупность могут попасть два и более изделий с абсолютно одинаковыми характеристиками. Следовательно, и в этом случае при обработке результатов выборки необходимо ориентироваться на методологию, используемую при повторном отборе.

При **бесповторном отборе** попавшая в выборку единица подвергается обследованию и в дальнейшей процедуре отбора не участвует. Такой отбор целесообразен и практически возможен в тех случаях, когда объем генеральной совокупности четко определен. Получаемые при этом результаты, как правило, являются более точными по сравнению с результатами, основанными на повторной выборке.

Как уже отмечалось выше, выборочное наблюдение всегда связано с определенными ошибками получаемых характеристик. Эти ошибки называются ошибками репрезентативности (представительности).

Ошибки репрезентативности обусловлены тем обстоятельством, что выборочная совокупность не может по всем параметрам в точности воспроизвести совокупность генеральную. Получаемые расхождения или ошибки репрезентативности позволяют заключить, в какой степени попавшие в выборку единицы могут представлять всю генеральную совокупность. При этом следует различать систематические и случайные ошибки репрезентативности.

Систематические ошибки репрезентативности связаны с нарушением принципов формирования выборочной совокупности. Например, вследствие каких-либо причин, связанных с организацией отбора, в выборку попали единицы, характеризующиеся несколько большими или, наоборот, несколько меньшими по сравнению с другими единицами значениями наблюдаемых признаков. В этом случае и рассчитанные выборочные характеристики будут завышенными или заниженными.

Случайные ошибки репрезентативности обусловлены действием случайных факторов, не содержащих каких-либо элементов системности в направлении воздействия на рассчитываемые выборочные характеристики. Но даже при строгом соблюдении всех принципов формирования выборочной совокупности выборочные и генеральные характеристики будут несколько различаться. Получаемые случайные ошибки могут быть статистически оценены и учтены при распространении результатов выборочного наблюдения на всю генеральную совокупность. Оценка ошибок выборочного наблюдения основана на теоремах теории вероятностей.

При дальнейшем рассмотрении теории и методов выборочного наблюдения в данной главе используются следующие общепринятые **условные обозначения**:

N- объем (число единиц) генеральной совокупности;

n - объем (число единиц) выборочной совокупности;

□ - генеральная средняя, т.е. среднее значение изучаемого признака по генеральной совокупности (средняя прибыль, средняя величина активов, средняя численность работников предприятия и т.п);

□- выборочная средняя, т.е. среднее значение изучаемого признака по выборочной совокупности;

M - численность единиц генеральной совокупности, обладающих определенным вариантом или вариантами изучаемого признака (численность городского населения, численность сельского населения, количество бракованных изделий, число нерентабельных предприятий и т.п.);

p - генеральная доля, т.е. доля единиц, обладающих определенным вариантом или вариантами изучаемого признака, во всей генеральной совокупности (доля городского населения в общей численности населения, доля бракованной продукции в общем выпуске, доля нерентабельных предприятий в общей численности предприятий и т.п.); определяется как $\frac{M}{N}$;

m - численность единиц выборочной совокупности, обладающих определенным вариантом или вариантами изучаемого признака;

w - выборочная доля, т.е. доля единиц, обладающих определенным вариантом или вариантами изучаемого признака, в выборочной совокупности; определяется как $\frac{m}{n}$;

□ - средняя ошибка выборки;

□ - предельная ошибка выборки.

Ошибка выборки или отклонение выборочной средней от средней генеральной находится в прямой зависимости от дисперсии изучаемого признака в генеральной совокупности, и в обратной зависимости - от объема выборки. Таким образом среднюю ошибку выборки можно представить как

$$\mu = \sqrt{\frac{\sigma^2_{ген}}{n}}.$$

При проведении выборочного наблюдения дисперсия изучаемого признака в генеральной совокупности, как правило, не известна. В то же время, между генеральной дисперсией и средней из всех возможных выборочных дисперсий существует следующее соотношение:

$$\sigma^2_{ген} = \overline{\sigma}^2 \frac{n}{n-1}.$$

В связи с тем, что на практике в большинстве случаев из генеральной совокупности в определенный момент времени производится только одна выборка, дисперсия изучаемого признака по этой выборке и используется при расчете ошибки. Учитывая, что при достаточно большом объеме выборки отношение $\frac{n}{n-1}$ близко к 1, формула средней ошибки повторной выборки принимает следующий вид:

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

где $\square$ - дисперсия изучаемого признака по выборочной совокупности.

При определении возможных границ значений характеристик генеральной совокупности рассчитывается предельная ошибка выборки, которая зависит от величины ее средней ошибки и уровня вероятности, с которым гарантируется, что генеральная средняя не выйдет за указанные границы. Согласно теореме А.М.Ляпунова, вероятность той или иной величины предельной ошибки, при достаточно большом объеме выборочной совокупности, подчиняется нормальному закону распределения и может быть определена на основе интеграла Лапласа.

Значения интеграла Лапласа при различных величинах t табулированы и представлены в статистических справочниках. При обобщении результатов выборочного наблюдения наиболее часто используются следующие уровни вероятности и соответствующие им значения t:

P	0,683	0,950	0,954	0,997
t	1	1,96	2	3

Например, если при расчете предельной ошибки выборки мы используем значение t=2, то с вероятностью 0,954 можно утверждать, что расхождение между выборочной средней и генеральной средней не превысит двухкратной величины средней ошибки выборки.

Теоретической основой для определения границ генеральной доли, т.е. доли единиц, обладающих тем или иным вариантом признака, является теорема Бернули. Согласно данной теореме вероятность получения сколь угодно малого расхождения между выборочной долей и генеральной долей при достаточно большом объеме выборки будет стремиться к единице. С учетом того, что вероятность расхождения между выборочной и генеральной долями подчиняется нормальному закону распределения, эта вероятность также определяется по функции F(t) при заданном значении t.

Процесс подготовки и проведения выборочного наблюдения включает ряд последовательных этапов:

1. Определение цели обследования.
2. Установление границ генеральной совокупности.
3. Составление программы наблюдения и программы разработки данных.
4. Определение вида выборки, процента отбора и метода отбора.
5. Отбор и регистрация наблюдаемых признаков у отобранных единиц.
6. Расчет выборочных характеристик и их ошибок.
7. Распространение полученных результатов на генеральную совокупность.

В зависимости от состава и структуры генеральной совокупности выбирается вид выборки или способ отбора. К наиболее распространенным на практике видам относятся:

- собственно-случайная (простая случайная) выборка;
- механическая (систематическая) выборка;
- типическая (стратифицированная, расслоенная) выборка;
- серийная (гнездовая) выборка.

Отбор единиц из генеральной совокупности может быть комбинированным, многоступенчатым и многофазным.

Комбинированный отбор предполагает объединение нескольких видов выборки. Так, например, можно комбинировать типическую и серийную, серийную и собственно-случайную выборки. Ошибка такой выборки определяется ступенчатостью отбора.

Многоступенчатым называется отбор, при котором из генеральной совокупности сначала извлекаются укрупненные группы, потом – более мелкие и так до тех пор, пока не будут отобраны те единицы, которые подвергаются обследованию.

Многофазная выборка, в отличие от многоступенчатой, предполагает сохранение одной и той же единицы отбора на всех этапах его проведения; при этом отобранные на каждой стадии единицы подвергаются обследованию, каждый раз – по более расширенной программе.

7.2. Собственно-случайная (простая случайная) выборка

Собственно-случайная выборка заключается в отборе единиц из генеральной совокупности в целом, без разделения ее на группы, подгруппы или серии отдельных единиц. При этом единицы отбираются в случайном порядке, не зависящем ни от последовательности расположения единиц в совокупности, ни от значений их признаков.

Прежде чем производить собственно-случайный отбор, необходимо убедиться, что все без исключения единицы генеральной совокупности имеют абсолютно равные шансы попадания в выборку, в списках или перечне отсутствуют пропуски, игнорирования отдельных единиц и т.п. Следует также установить четкие границы генеральной совокупности таким образом, чтобы включение или невключение в нее отдельных единиц не вызывало сомнений. Так, например, при обследовании торговых предприятий необходимо указать, включит ли генеральная совокупность торговые павильоны, коммерческие палатки, передвижные торговые точки и прочие подобные объекты; при обследовании студентов важно определиться, будут ли приниматься во внимание студенты-заочники, экстерны, учащиеся в магистратуре, лица, находящиеся в академическом отпуске и т.п.

Для проведения отбора единиц в выборочную совокупность используется один из математических алгоритмов, например, **метод прямой реализации**, включающий следующие этапы:

1. Все единицы генеральной совокупности, расположенные в случайном порядке или ранжированные по какому-либо признаку, нумеруются от 1 до N.

2. С помощью процессора случайных чисел получают n значений в интервале от 1 до N. Если первоначально случайные числа получены в интервале от 0 до 1, их необходимо умножить на N и округлить по правилам до целого значения.

3. Из сформированного списка единиц генеральной совокупности отбираются единицы, соответствующие по номеру полученным случайным числам.

Упрощенным вариантом метода прямой реализации является отбор единиц в выборочную совокупность на основе **таблицы случайных чисел** (см. Приложение ...). Для проведения отбора могут быть использованы цифры любого столбца данной таблицы, при этом необходимо учитывать объем генеральной совокупности.

Рассмотрим процедуру отбора на основе фрагмента таблицы случайных чисел. Предположим, объем генеральной совокупности составляет 70000 единиц и требуется сформировать выборку объемом 500 единиц, то цифры таблицы следует перегруппировать для получения пятизначных чисел следующим образом:

5489	5583	3156	0835	1988
3522	0935	7877	5665	7020
7555	7579	2550	2487	9477
5759	3554	5080	9074	7001
6303	6895	3371	3196	7231

Для формирования выборки мы должны взять 500 чисел в интервале от 00001 до 70000. Таким образом, нам следует из списка единиц генеральной совокупности отобрать единицы под номером 54895, 35220, 57593 и т.д. При этом номера свыше 70000 (75557, 93578 и подобные) будут проигнорированы.

При проведении бесповторного отбора повторяющиеся номера следует учитывать только один раз. При повторном отборе, если тот или иной номер случайно встретится еще один или более раз, соответствующая этому номеру единица в каждом случае повторно включается в выборочную совокупность.

После проведения отбора с использованием какого-либо алгоритма, реализующего принцип случайности, или на основе таблицы случайных чисел, необходимо определить границы генеральных характеристик. Для этого рассчитываются средняя и предельная ошибки выборки.

Средняя ошибка повторной собственно-случайной выборки определяется по формуле:

$$\mu = \frac{\sigma}{\sqrt{n}} \tag{7.1}$$

С учетом выбранного уровня вероятности и соответствующего ему значения *t* **предельная ошибка повторной собственно-случайной выборки** выборки составит:

(7.2)

Тогда можно утверждать, что при заданной вероятности генеральная средняя будет находиться в следующих границах:

$$\tilde{x} - \Delta_{\tilde{x}} \le \bar{x} \le \tilde{x} + \Delta_{\tilde{x}}. \tag{7.3}$$

Предположим, в результате выборочного обследования доходов домохозяйств региона, осуществленного на основе собственно-случайной повторной выборки, получен следующий ряд распределения (табл. 7.1).

Таблица 7.1.

Результаты выборочного обследования доходов домохозяйств региона

Доход, тыс.руб.	До 5	5-10	10-15	15-20	20 и более
Число домохозяйств	52	354	475	170	49

Рассмотрим определение **границ генеральной средней**, в данном примере – среднего дохода домохозяйства в целом по данному региону, опираясь только на результаты выборочного обследования. Для определения средней ошибки выборки нам необходимо прежде всего рассчитать выборочную среднюю величину и дисперсию изучаемого признака (табл. 8.2).

Таблица 7.2.

Расчет среднего дохода домохозяйства и дисперсии

Доход, тыс.руб.	Число домохозяйств f	Середина интервала x	xf	$x^2 f$
До 5	52	2,5	130,0	325,0
5 - 10	354	7,5	2655,0	19912,5
10 - 15	475	12,5	5937,5	74218,75
15 - 20	170	17,5	2975,0	52062,5
20 и более	49	22,5	1102,5	24806,25
Итого	1100	-	12800	171325

$$\tilde{x} = \frac{12800}{1100} = 11{,}6;$$

$$\sigma^2 = \frac{171325}{1100} - 11{,}6^2 = 21{,}19;$$

$$\sigma = \sqrt{21{,}19} = 4{,}6.$$

Средняя ошибка выборки составит:

$$\mu_{\tilde{x}} = \frac{4{,}6}{\sqrt{1100}} = 0{,}14.$$

Определим предельную ошибку выборки с вероятностью 0,954 (t=2):

$$\Delta_{\tilde{x}} = 2 \cdot 0{,}14 = 0{,}28.$$

Установим границы генеральной средней (тыс.руб.):

$$11{,}6 - 0{,}28 \le \bar{x} \le 11{,}6 + 0{,}28$$

или

$$11{,}32 \le \bar{x} \le 11{,}88.$$

Таким образом, на основании проведенного выборочного обследования с вероятностью 0,954 можно заключить, что средний доход домохозяйства в целом по региону лежит в пределах от 11,3 до 11,9 тыс.руб.

При расчете **средней ошибки собственно-случайной бесповторной выборки** необходимо учитывать поправку на бесповторность отбора:

$$\mu_{\tilde{x}} = \sqrt{\frac{\sigma^2}{n}\left(1 - \frac{n}{N}\right)}. \qquad (7.4)$$

Если предположить, что представленные в табл. 7.1 данные являются результатом 5%-ного бесповторного отбора (следовательно, генеральная совокупность включает 22000 домохозяйств), то средняя ошибка выборки будет несколько меньше:

$$\mu_{\tilde{x}} = \sqrt{\frac{21{,}19}{1100}\left(1 - \frac{1100}{22000}\right)} = 0{,}135.$$

Соответственно уменьшится и предельная ошибка выборки, что вызовет сужение границ генеральной средней. Особенно ощутимо влияние поправки на бесповторность отбора при относительно большом проценте выборки.

Мы рассмотрели определение границ генеральной средней. Рассмотрим теперь, как определяются **границы генеральной доли**, т.е. границы доли единиц, обладающих тем или иным значением признака.

Воспользуемся еще раз данными табл. 7.1 для того, чтобы определить границы доли домохозяйств, доходы которых составляют менее 10 тыс.руб. Согласно результатам обследования, численность таких домохозяйств составила 52+354=406. Определим выборочную долю и дисперсию:

$$w = \frac{406}{1100} = 0{,}369;$$

$$\sigma_w^2 = w(1 - w) = 0{,}369 \cdot 0{,}631 = 0{,}2328.$$

Рассчитаем среднюю ошибку выборки:

$$\mu_w = \sqrt{\frac{0{,}2328}{1100}\left(1 - \frac{1100}{22000}\right)} = 0{,}014.$$

Предельная ошибка выборки с заданной вероятностью составит:

$$\Delta_w = 2 \cdot 0{,}014 = 0{,}028.$$

Определим границы генеральной доли:

$$0{,}369 - 0{,}028 \le p \le 0{,}369 + 0{,}028$$

или

$$0{,}341 \le p \le 0{,}397.$$

Следовательно, с вероятностью 0.954 можно утверждать, что доля домохозяйств, имеющих доходы менее 10 тыс.руб., в целом по данному региону находится в пределах от 34,1 до 39,7%.

Мы рассмотрели определение границ генеральной средней и генеральной доли по результатам уже проведенного выборочного наблюдения, при известном объеме выборки или проценте отбора. На этапе же проектирования выборочного наблюдения именно объем выборочной совокупности и требует определения.

Чем больше объем выборки, тем меньше значения средней и предельной ошибок выборочного наблюдения и, следовательно, тем уже границы генеральной средней и генеральной доли. В то же время, необходимо учитывать, что большой объем выборки приводит к удорожанию обследования, увеличению сроков сбора и обработки материалов, требует привлечения дополнительного персонала и соответствующего материально-технического обеспечения. Затраты всех ресурсов на 20-30%-ное выборочное наблюдение уже сопоставимы с расходами на сплошное обследование. При этом не следует забывать, что статистические характеристики, полученные по выборочной совокупности, всегда имеют вероятностную основу и всегда будут уступать результатам сплошного наблюдения по точности и надежности. Поэтому при подготовке выборочного наблюдения необходимо определить тот минимально необходимый объем выборки, который обеспечит требуемую точность полученных статистических характеристик при заданном уровне вероятности.

Представим формулу (7.2) следующим образом:

$$\Delta_{\tilde{x}} = t\frac{\sigma}{\sqrt{n}}. \tag{7.5}$$

Отсюда можно вывести формулу для определения **необходимого объема собственно-случайной повторной выборки**:

(7.6)

Полученный на основе использования данной формулы результат всегда округляется в большую сторону. Например, если мы получили, что необходимый объем выборки составляет 493,1 единицы, то обследовав 493 единицы мы не достигнем требуемой точности. Поэтому, для достижения желаемого результата обследованием должны быть охвачены 494 единицы. С другой стороны, рассчитанное значение необходимого объема выборки свободно может быть увеличено в большую сторону на несколько единиц. Если мы располагаем необходимыми ресурсами, если по причинам организационного порядка (компактность расположения единиц, фиксированная нагрузка на каждого регистратора и т.п.) мы вполне можем охватить больший объем, то включение в выборочную совокупность 500 или, например, 550 единиц только уменьшит значения полученных случайной и предельной ошибок.

Как видно из формулы (7.6) необходимый объем выборки будет тем больше, чем выше заданный уровень вероятности и чем сильнее варьирует наблюдаемый признак. В то же время повышение допустимой предельной ошибки выборки приводит к снижению необходимого ее объема.

Расчет необходимого объема выборки предполагает, что организаторы выборочного наблюдения уже на этапе его проектирования располагают по крайней мере косвенными данными о вариации изучаемых признаков. Источниками таких данных могут служить:

а) результаты исследования данного объекта в предшествующие периоды;

б) результаты исследования аналогичных объектов (жителей других населенных пунктов, предприятий других регионов и т.п.);

в) специально проведенное небольшое по объему выборочное обследование данного объекта, ставящее целью лишь изучение вариации наблюдаемых признаков.

При определении необходимого объема выборки для определения границ генеральной доли задача оценки вариации решается значительно проще. Если дисперсия изучаемого альтернативного признака неизвестна, то можно использовать ее максимальное возможное значение:

$$\sigma^2_{w\max} = w(1-w) = 0{,}5(1-0{,}5) = 0{,}25.$$

Например, предприятию связи с вероятностью 0,954 необходимо определить удельный вес телефонный разговоров продолжительностью менее 1 минуты с предельной ошибкой 2%. Сколько разговоров нужно обследовать в порядке собственно-случайного повторного отбора для решения этой задачи?

Для получения ответа на поставленный вопрос воспользуемся формулой (7.6) и будем ориентироваться на максимальную возможную дисперсию доли телефонных разговоров такой продолжительности. Расчет приводит к следующему результату:

$$n = \frac{2^2 \cdot 0{,}25^2}{0{,}02^2} = 2500.$$

Таким образом, обследованием должны быть охвачены не менее 2500 разговоров на предмет их продолжительности.

Необходимый объем собственно-случайной бесповторной выборки может быть определен по следующей формуле:

$$n = \frac{t^2 \sigma^2 N}{t^2 \sigma^2 + \Delta^2_{\tilde{x}} N}. \tag{7.7}$$

Укажем на одну особенность формулы (7.7). При проведении вычислений объем генеральной совокупности должен быть выражен только в единицах, а не в тысячах или в миллионах единиц. Например, подставив в данную формулу общую численность населения региона, выраженную в тысячах человек, мы не получим правильное значение необходимой численности выборки, также выраженное в тысячах человек, как это иногда бывает в других расчетах. Результат вычислений будет неверен.

7.3. Механическая (систематическая) выборка

Механическая выборка может быть применена в тех случаях, когда генеральная совокупность каким-либо образом упорядочена, т.е. имеется определенная последовательность в расположении единиц (табельные номера работников, списки избирателей, телефонные номера респондентов, номера домов и квартир и т.п.). Для проведения отбора желательно, чтобы все единицы также имели порядковые номера от 1 до N.

Для проведения механической выборки устанавливается пропорция отбора, которая определяется соотнесением объемов выборочной и генеральной совокупностей. Так, если их совокупности в 500000 единиц предполагается отобрать 10000 единиц, то пропорция отбора составит . Отбор единиц осуществляется в соответствии с установленной пропорцией через равные интервалы. Например, при пропорции 1:50 (2%-ная выборка) отбирается каждая 50-я единица, при пропорции 1:20 (5%-ная выборка) - каждая 20-я единица и т.д.

Интервал отбора также можно определить как частное от деления 100% на установленный процент отбора. Так, при 2%-ном отборе интервал составит 50 (100%:2%), при 4%-ном отборе - 25 (100%:4%). В тех случаях, когда результат деления получается дробным, сформировать выборку механическим способом при строгом соблюдении процента отбора не представляется возможным. Например, по этой причине нельзя сформировать 3%-ную или 6%-ную выборки.

Генеральную совокупность при механическом отборе можно ранжировать или упорядочить по величине изучаемого или коррелирующего с ним признака, что позволит повысить репрезентативность выборки. Однако в этом случае возрастает опасность систематической ошибки, связанной с занижением значений изучаемого признака (если из каждого интервала регистрируется первое значение) или его завышением (если из каждого интервала регистрируется последнее значение). Поэтому целесообразно из каждого интервала отбирать центральную или одну из двух центральных единиц.

Например, при 5%-ной выборке интервал отбора составит 20 единиц, тогда отбор целесообразно начинать с 10-й или с 11-й единицы. В первом случае в выборку попадут 10, 30, 50, 70 и с таким же интервалом последующие единицы; во втором случае - единицы с номерами 11, 31, 51, 71 и т.д.

При механической выборке также может появиться опасность систематической ошибки, обусловленной случайным совпадением выбранного интервала и циклических закономерностей в расположении единиц генеральной совокупности. Так, при переписи населения 1989 г. в ходе 25%-го выборочного обследования семей имела место опасность попадания в выборку квартир только одного типа (например, только однокомнатных или только трехкомнатных), так как на лестничных площадках многих типовых домов располагаются именно по 4 квартиры. Чтобы избежать систематической ошибки, в каждом новом подъезде счетчик менял начало отбора.

Для определения средней ошибки механической выборки, а также необходимой ее численности, используются соответствующие формулы, применяемые при собственно-случайном бесповторном отборе (7.4 и 7.7). При этом, определив необходимую численность выборки и сопоставив ее с объемом генеральной совокупности, как правило, приходится производить соответствующее округление для получения целочисленного интервала отбора.

Например, в области зарегистрировано 12000 фермерских хозяйств. Определим, сколько из них нужно отобрать в порядке механического отбора для определения средней площади сельхозугодий с ошибкой □2 га. (Р=0,997). По результатам ранее проведенного обследования известно, что среднее квадратическое отклонение площади сельхозугодий составляет 8 га. Произведем расчет, воспользовавшись формулой 7.7.:

$$n = \frac{3^2 \cdot 8^2 \cdot 12000}{3^2 \cdot 8^2 + 2^2 \cdot 12000} = 142{,}3 \approx 143.$$

С учетом полученного необходимого объема выборки (143 фермерских хозяйства) определим интервал отбора: 12000:143=83,9. Определенный таким способом интервал всегда округляется в меньшую сторону, так как при округлении в большую сторону произведенная выборка не достигнет рассчитанного по формуле необходимого объема. Следовательно, в нашем примере, из общего списка фермерских хозяйств необходимо отобрать для обследования каждое 83-е хозяйство. При этом процент отбора составит 1,2% (100%:83).

7.4. Типическая (стратифицированная) выборка

Типический отбор целесообразно использовать в тех случаях, когда все единицы генеральной совокупности объединены в несколько крупных типических групп. Такие группы также называют стратами или слоями, в связи с чем типический отбор также называют стратифицированным или расслоенным. При обследованиях населения в качестве типических групп могут быть выбраны области, районы, социальные, возрастные или образовательные группы, при обследовании предприятий - отрасли или подотрасли, формы собственности и т.п.

Рассматривать генеральную совокупность в разрезе нескольких крупных групп единиц имеет смысл только в том случае, если средние значения изучаемых признаков по группам существенно различаются. Например, с большой уверенностью можно предположить, что доходы населения крупного города будут в среднем выше доходов населения, проживающего в сельской местности; численность работников промышленного предприятия в среднем будет выше численности работников торгового или сельскохозяйственного предприятия; средний возраст студентов будет значительно меньше среднего возраста занятого населения и, тем более, пенсионеров. В то же время, нет никакого смысла при выделении типических групп ориентироваться на признак, не связанный или очень слабо связанный с изучаемым.

Отбор единиц в выборочную совокупность из каждой типической группы осуществляется собственно-случайным или механическим способом. Поскольку в выборочную совокупность в той или иной пропорции обязательно попадают представители всех групп, типизация генеральной совокупности позволяет исключить влияние межгрупповой дисперсии на среднюю ошибку выборки. В то же время, в выделенных типических группах обследуются далеко не все единицы, а только включенные в выборку. Следовательно, на величине полученной ошибки будет сказываться различие между единицами внутри этих групп, т.е. внутригрупповая вариация. Поэтому, ошибка типической выборки будет определяться величиной не общей дисперсии, а только ее части - средней из внутригрупповых дисперсий.

При типической выборке, *пропорциональной объему типических групп*, число единиц, подлежащих отбору из каждой группы, определяется следующим образом:

$$n_i = n\frac{N_i}{N}, \qquad (7.8)$$

где N_i - объем i-й группы;

n_i - объем выборки из i-й группы.

Предположим, общая численность населения области составляет 1,5 млн. чел., в том числе городское - 900 тыс. чел. и сельское - 600 тыс.

чел. Если в ходе выборочного наблюдения планируется обследовать 100 тыс. жителей, то эта численность должна быть поделена пропорционально объему типических групп следующим образом:

городское население - $n_{\langle} = 100000 \dfrac{900000}{1500000} = 60000$ чел.;

сельское население - $n_{\_} = 100000 \dfrac{6000000}{1500000} = 40000$ чел..

Средняя ошибка типической выборки определяется по формулам:

$$\mu = \sqrt{\frac{\overline{\sigma}^2}{n}} \quad \text{(повторный отбор),} \qquad (7.9)$$

$$\mu = \sqrt{\frac{\overline{\sigma}^2}{n}\left(1 - \frac{n}{N}\right)} \quad \text{(бесповторный отбор),} \qquad (7.10)$$

где $\square$ - средняя из внутригрупповых дисперсий.

Рассмотрим данный вариант типической выборки на условном примере.

Предположим, 10%-ный бесповторный типический отбор безработного населения, пропорциональный размерам районов, проведенный с целью оценки продолжительности периода поиска работы, привел к следующим результатам (табл. 7.3).

Таблица 7.3.

Результаты обследования безработного населения области

Район	Всего зарегистрировано безработных, чел.	Обследовано, чел.	Число недель поиска работы	
			средняя	дисперсия
А	5000	500	7	36
Б	8200	820	15	64
В	2100	210	5	9

Рассчитаем среднюю из внутригрупповых дисперсий:

$$\overline{\sigma}^2 = \frac{\sum \sigma_i^2 n_i}{\sum n_i} = \frac{36 \cdot 500 + 64 \cdot 820 + 9 \cdot 210}{500 + 820 + 210} = 47{,}0.$$

Определим среднюю и предельную ошибки выборки (с вероятностью 0,954):

$$\mu_{\tilde{x}} = \sqrt{\frac{47{,}0}{1530}\left(1 - \frac{1530}{15300}\right)} = 0{,}17;$$

$\Delta_{\tilde{x}} = 2 \cdot 0{,}17 = 0{,}34.$

Рассчитаем выборочную среднюю:

$$\tilde{x} = \frac{\sum x_i n_i}{\sum n_i} = \frac{7 \cdot 500 + 15 \cdot 820 + 5 \cdot 210}{500 + 820 + 210} = 11{,}0$$ недель.

В результате проведенных расчетов с вероятностью 0,954 можно сделать вывод, что среднее число недель, затрачиваемых на поиск работы, в целом по области находится в пределах:

$11{,}0 - 0{,}34 \le \bar{x} \le 11{,}0 + 0{,}34.$

При определении **необходимого объема типической выборки** в рассмотренных выше формулах (7.6) и (7.7) общую дисперсию наблюдаемого признака необходимо заменить на среднюю из внутригрупповых дисперсий. Тогда данные формулы примут следующий вид:

$$n = \frac{t^2 \bar{\sigma}^2}{\Delta_{\tilde{x}}^2} \quad \text{(повторный отбор)} \tag{7.11}$$

$$n = \frac{t^2 \bar{\sigma}^2 N}{t^2 \bar{\sigma}^2 + \Delta_{\tilde{x}}^2 N} \quad \text{(бесповторный отбор)} \tag{7.12}$$

Предположим, в рассмотренном выше примере нам необходимо определить среднее число недель, затрачиваемых на поиск работы, с предельной ошибкой □1 неделя. Учитывая величину полученной ранее средней из внутригрупповых дисперсий определим необходимый объем типической выборки при условии бесповторного отбора:

$$n = \frac{2^2 \cdot 47{,}0 \cdot 15300}{2^2 \cdot 47{,}0 + 1^2 \cdot 15300} = 185{,}7.$$

Таким образом мы получили, что при заданных условиях для достижения требуемой точности достаточно обследовать выборочным методом всего 186 чел. Распределим эту численность на три района рассматриваемой области пропорционально их размерам по числу зарегистрированных безработных:

$$n_А = 186 \frac{5000}{15300} = 60{,}8;$$

$$n_Б = 186 \frac{8200}{15300} = 99{,}7;$$

$$n_В = 186 \frac{2100}{15300} = 25{,}5.$$

Расчеты показывают, что в районе А необходимо обследовать 61 чел., в районе Б - 100 чел., и в районе В - 25 чел.

Мы рассмотрели типический отбор, пропорциональный объему типических групп. Второй вариант формирования типической выборки заключается в отборе единиц, *пропорциональном вариации признака* в типических группах. Логика такого отбора заключается в следующем: если внутри какой-либо типической группы наблюдаемый признак варьирует слабо, то для определения границ генеральных характеристик из данной группы достаточно обследовать относительно небольшое число единиц; при сильной же вариации признака объем выборки должен быть соответственно увеличен.

7.5. Серийная выборка

Сущность серийной выборки заключается в собственно-случайном либо механическом отборе групп единиц (серий), внутри которых производится сплошное обследование. Единицей отбора при этой выборке является группа или серия, а не отдельная единица генеральной совокупности, как это имело место в рассматриваемых ранее выборках.

Данный способ отбора удобен в тех случаях, когда единицы генеральной совокупности изначально объединены в небольшие более или менее равновеликие группы или серии. В качестве таких серий могут выступать упаковки с определенным количеством готовой продукции, партии товара, студенческие группы, бригады и другие подобные объединения.

В большинстве случаев серийная выборка имеет не столько методологические, сколько организационные преимуществами перед другими способами формирования выборочной совокупности. Например, в Великобритании серийный отбор используется в обследованиях населения, когда серией являются домохозяйства, объединенные общим почтовым индексом. В случайном порядке производится выборка индексов и под обследование попадают все домохозяйства, имеющие индекс попавших в выборочную совокупность почтовых отделений.

В связи с тем, что при серийном отборе внутри отобранных групп обследуются все без исключения единицы, внутригрупповая вариация признака не отразится на ошибках выборочного наблюдения. В то же время, обследуются не все группы, а только попавшие в выборку. Следовательно на ошибках получаемых характеристик будут отражаться различия между группами, которые определяются межгрупповой дисперсией. Поэтому средняя ошибка серийной выборки определяется по формулам:

$$\mu = \sqrt{\frac{\delta^2}{r}} \quad \text{(повторный отбор),} \tag{7.13}$$

$$\mu = \sqrt{\frac{\delta^2}{r}\left(1 - \frac{r}{}\right)} \quad \text{(бесповторный отбор) ,} \tag{7.14}$$

где r - число отобранных серий;

R- общее число серий.

Межгрупповую дисперсию при равновеликих группах вычисляют следующим образом:

(7.15)

где - средняя i-й серии;

- общая средняя по всей выборочной совокупности.

Рассмотрим следующий пример. Предположим, партия готовой продукции предприятия упакована в 160 ящиков по 25 изделий в каждом. В целях контроля соблюдения параметров технологического процесса проведена 5%-ная серийная выборка, в ходе которой отбирался каждый 20-й ящик. Все изделия, находящиеся в отобранных ящиках были подвергнуты сплошному обследованию, заключающемуся в определении их точного веса. Полученные результаты представлены в следующей таблице:

Таблица 7.4.

Результаты выборочного обследования готовой продукции

Номер коробки	1	2	3	4	5	6	7	8
Средний вес изделия в ящике, г	563	545	548	560	555	561	547	552

С вероятностью 0,954 требуется определить границы среднего веса изделия во всей партии.

На основе приведенных в таблице внутригрупповых средних определим средний вес изделия по выборочной совокупности:

$$\tilde{x} = \frac{563 + 545 + ... + 552}{8} = 553{,}9$$ г.

С учетом полученной средней рассчитаем межгрупповую дисперсию:

$$\delta^2 = \frac{(563 - 553{,}9)^2 + (545 - 553{,}9)^2 + ... + (552 - 553{,}9)^2}{8} = 42{,}11.$$

Рассчитаем среднюю и предельную ошибки выборки:

$$\mu = \sqrt{\frac{42{,}11}{8}\left(1 - \frac{8}{160}\right)} = 2{,}2;$$ г;

$$\Delta_{\tilde{x}} = 2 \cdot 2{,}2 = 4{,}4.$$ г.

Определим границы генеральной средней:

$$553{,}9 - 4{,}4 \le \tilde{x} \le 553{,}9 + 4{,}4.$$

На основе результатов проведенных расчетов с вероятностью 0,954 можно утверждать, что средний вес изделия в целом по всей партии продукции находится в пределах от 549,5 г до 558,3 г.

Для определения необходимого объема серийной выборки при заданной предельной ошибке используются следующие формулы:

(повторный отбор);

$$r = \frac{t^2 \delta^2 R}{t^2 \delta^2 + \Delta_{\tilde{x}}^2 R}$$ (бесповторный отбор). (7.16)

Предположим, в рассмотренном выше примере необходимо определить границы среднего веса изделия с предельной ошибкой □ 3,0 г. Используя полученные выше данные о вариации веса определим, сколько ящиков с изделиями нужно обследовать в порядке бесповторной серийной выборки, чтобы получить результат с заданной точностью и при выбранном уровне вероятности:

$$r = \frac{2^2 \cdot 42{,}11 \cdot 160}{2^2 \cdot 42{,}11 + 3{,}0^2 \cdot 160} = 16{,}8.$$

Выполненный расчет позволяет заключить, что для получения границ генеральной средней с заданной точностью необходимо обследовать не менее 17 ящиков с изделиями, отобранных собственно-случайным или механическим способом.

Глава 8. Статистическое изучение взаимосвязи социально-экономических явлений

8.1. Причинность, регрессия, корреляция

Исследование объективно существующих связей между социально-экономическими явлениями и процессами является важнейшей задачей теории статистики. В процессе статистического исследования зависимостей вскрываются причинно-следственные отношения между явлениями, что позволяет выявлять факторы (признаки), оказывающие основное влияние на вариацию изучаемых явлений и процессов. Причинно-следственные отношения - это такая связь явлений и процессов, когда изменение одного из них - причины ведет к изменению другого - следствия.

Финансово-экономические процессы представляют собой результат одновременного воздействия большого числа причин. Следовательно, при изучении этих процессов необходимо выявлять главные, основные причины, абстрагируясь от второстепенных.

В основе первого этапа статистического изучения связи лежит качественный анализ, связанный с анализом природы социального или экономического явления методами экономической теории, социологии, конкретной экономики. Второй этап - построение модели связи, базируется на методах статистики: группировках, средних величинах, и так далее. Третий, последний этап - интерпретация результатов, вновь связан с качественными особенностями изучаемого явления. Статистика разработала множество методов изучения связей. Выбор метода изучения связи зависит от познавательной цели и задач исследования.

Признаки по их сущности и значению для изучения взаимосвязи делятся на два класса. Признаки, обуславливающие изменения других, связанных с ними признаков, называются **факторными**, или просто факторами. Признаки, изменяющиеся под действием факторных признаков, называются **результативными**.

В статистике различают функциональную и стохастическую зависимости. **Функциональной** называют такую связь, при которой определенному значению факторного признака соответствует одно и только одно значение результативного признака.

Если причинная зависимость проявляется не в каждом отдельном случае, а в общем, среднем, при большом числе наблюдений, то такая зависимость называется **стохастической.** Частным случаем стохастической связи является **корреляционная** связь, при которой изменение среднего значения результативного признака обусловлено изменением факторных признаков.

Связи между явлениями и их признаками классифицируются по степени тесноты, направлению и аналитическому выражению.

По степени тесноты связи различают (табл.8.1):

Таблица 8.1.

Количественные критерии оценки тесноты связи

Величина показателя связи	Характер связи
До ±0,3	практически отсутствует
±0,3 - ±0,5	слабая
±0,5 - ±0,7	умеренная
±0,7 - ±1,0	сильная

По направлению выделяют связь **прямую** и **обратную. Прямая -** это связь, при которой с увеличением или с уменьшением значений факторного признака происходит увеличение или уменьшение значений результативного признака. Так, рост объемов производства способствует увеличению прибыли предприятия. В случае **обратной** связи значения результативного признака изменяются под воздействием факторного, но в противоположном направлении по сравнению с изменением факторного признака, то есть **обратная** – это связь, при которой с увеличением или с уменьшением значений одного признака происходит уменьшение или увеличение значений другого признака. Так снижение себестоимости единицы производимой продукции влечет за собой рост рентабельности.

По аналитическому выражению выделяют связи **прямолинейные** (или просто **линейные**) и **нелинейные**. Если статистическая связь между явлениями может быть приблизительно выражена уравнением прямой линии, то ее называют **линейной** связью вида:

$$\bar{y}_x = a_0 + a_1 x \tag{8.1}$$

Если же связь может быть выражена уравнением какой-либо кривой, то такую связь называют нелинейной или криволинейной, например:

параболы - $\bar{y}_x = a_0 + a_1 x + a_2 x^2$ (8.2)

гиперболы - $\bar{y}_x = a_0 + a_1 \frac{1}{x}$; и т.д..

Для выявления наличия связи, ее характера и направления в статистике используются методы: приведения параллельных данных; графический; аналитических группировок; корреляции, регрессии.

Метод приведения параллельных данных основан на сопоставлении двух или нескольких рядов статистических величин. Такое сопоставление позволяет установить наличие связи и получить представление о ее характере.

Графически взаимосвязь двух признаков изображается с помощью поля корреляции. В системе координат на оси абсцисс откладываются значения факторного признака, а на оси ординат - результативного. Каждое пересечение линий, проводимых через эти оси, обозначаются точкой. При отсутствии тесных связей имеет место беспорядочное расположение точек на графике. Чем сильнее связь между признаками, тем теснее будут группироваться точки вокруг определенной линии, выражающей форму связи.

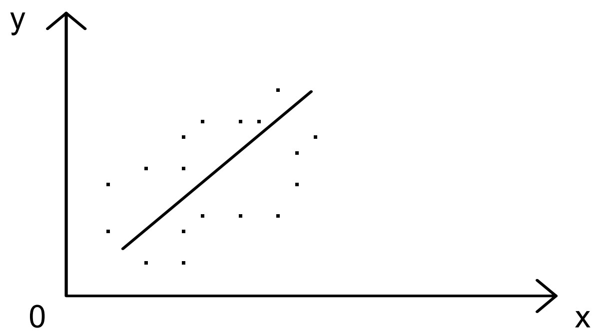

Рис. 8.1. График корреляционного поля

В статистике принято различать следующие виды зависимостей:

1. Парная корреляция - связь между двумя признаками (результативным и факторным, или двумя факторными).
2. Частная корреляция - зависимость между результативным и одним факторным признаками при фиксированном значении других факторных признаков.
3. Множественная корреляция - зависимость результативного и двух или более факторных признаков, включенных в исследование.

Корреляционный анализ имеет своей задачей количественное определение тесноты и направления связи между двумя признаками (при парной связи) и между результативным и множеством факторных признаков (при многофакторной связи).

Теснота связи количественно выражается величиной коэффициентов корреляции, которые, давая количественную характеристику тесноты связи между признаками, позволяют определять «полезность» факторных признаков при построении уравнения множественной регрессии. Знаки при коэффициентах корреляции характеризуют направление связи между признаками.

Регрессия тесно связана с корреляцией и позволяет исследовать аналитическое выражение взаимосвязи между признаками.

Регрессионный анализ заключается в определении аналитического выражения связи, в котором изменение одной величины (называемой зависимой или результативным признаком), обусловлено влиянием одной или нескольких независимых величин (факторных признаков).

Одной из проблем построения уравнений регрессии является их размерность, то есть определение числа факторных признаков, включаемых в модель. Их число должно быть оптимальным. Сокращение размерности за счет исключения второстепенных, несущественных факторов позволяет получить модель, быстрее и качественнее реализуемую. В то же время, построение модели малой размерности может привести к тому, что она будет недостаточно полно описывать исследуемое явление или процесс.

При построении моделей регрессии должны соблюдаться следующие требования:

1. Совокупность исследуемых исходных данных должна быть однородной и математически описываться непрерывными функциями.
2. Возможность описания моделируемого явления одним или несколькими уравнениями причинно-следственных связей.
3. Все факторные признаки должны иметь количественное (числовое) выражение.
4. Наличие достаточно большого объема исследуемой совокупности (в последующих примерах в целях упрощения изложения материала это условие нарушено, т.е. объем очень мал).
5. Причинно-следственные связи между явлениями и процессами должны описываться линейной или приводимой к линейной форме зависимостью.
6. Отсутствие количественных ограничений на параметры модели связи.
7. Постоянство территориальной и временной структуры изучаемой совокупности.

Соблюдение данных требований позволяет построить модель, наилучшим образом описывающую реальные социально-экономические явления и процессы.

8.2. Парная регрессия на основе метода наименьших квадратов

Парная регрессия позволяет получить аналитическое выражение связи между двумя признаками: результативным и факторным.

Определить тип уравнения можно, исследуя зависимость графически, однако существуют более общие указания, позволяющие выявить уравнение связи, не прибегая к графическому изображению. Если результативный и факторный признаки возрастают одинаково, то это свидетельствует о том, что связь между ними линейная, а при обратной связи - гиперболическая. Если результативный признак увеличивается в арифметической прогрессии, а факторный значительно быстрее, то используется параболическая или степенная регрессия.

Оценка параметров уравнений регрессии ($a0$, $a1$, и $a2$ - в уравнении параболы второго порядка) осуществляется методом наименьших квадратов, в основе которого лежит предположение о независимости наблюдений исследуемой совокупности и нахождении параметров модели (*a0* , *a1*), при которых минимизируется сумма квадратов отклонений эмпирических (фактических) значений результативного признака от теоретических, полученных по выбранному уравнению регрессии:

$$\mathbf{S} = \sum_{i=1}^{n} (y_i - \bar{y}_x)^2 \rightarrow \mathbf{min}$$

Система нормальных уравнений для нахождения параметров линейной парной регрессии методом наименьших квадратов имеет следующий вид:

$$\begin{cases} na_0 + a_1\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i \\ a_0\sum_{i=1}^{n} x_i + a_1\sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} x_i y_i \end{cases} \tag{8.3}$$

где *n* - объем исследуемой совокупности (число единиц наблюдения).

В уравнениях регрессии параметр *a0* показывает усредненное влияние на результативный признак неучтенных в уравнении факторных признаков. Коэффициент регрессии *a1* показывает, на сколько в среднем изменяется значение результативного признака при увеличении факторного признака на единицу собственного измерения.

Пример.

Имеются следующие данные о размере страховой суммы и страховых возмещений на автотранспортные средства одной из страховых компаний г. Москвы на 01.01.2004 г.

Таблица 8.2.

Зависимость между размером страховых возмещений и страховой суммой на автотранспорт одной из страховых компаний г. Москвы на 01.01.2004 г.

№ автомобиля в регистре	Объем страхового возмещения (тыс.долл.США), Yi	Стоимость застрахованного автомобиля (тыс.долл.США), Xi
1	0,1	8,8
2	1,3	9,4
3	0,1	10,0
4	2,6	10,6
5	0,1	11,0
6	0,3	11,9
7	4,6	12,7
8	0,3	13,5
9	0,4	15,5
10	7,3	16,7
Итого	17,1	120,1

Предположим наличие линейной зависимости между рассматриваемыми признаками.

Построим расчетную таблицу для определения параметров линейного уравнения регрессии объема страхового возмещения (табл. 8.3).

Таблица 8.3.

Расчетная таблица для определения параметров уравнения регрессии

№ автомобиля в регистре	Объем страхового возмещения (тыс.долл. США), Yi	Стоимость застрахованного автомобиля (тыс.долл.США), Xi	*x2*	*xy*	$\bar{y}_x$
1	0,1	8,8	77,44	0,88	0,052
2	1,3	9,4	88,36	12,22	0,362
3	0,1	10,0	100,00	1,00	0,672
4	2,6	10,6	112,36	27,56	0,982
5	0,1	11,0	121,00	1,10	1,188
6	0,3	11,9	141,61	3,57	1,653
7	4,6	12,7	161,29	58,42	2,066
8	0,3	13,5	182,25	4,05	2,479
9	0,4	15,5	240,25	6,20	3,513
10	7,3	16,7	278,89	121,91	4,133
Итого	**17,1**	**120,1**	**1503,45**	**236,91**	**17,100**

Система нормальных уравнений для данного примера имеет вид:

$$\begin{cases} 10a_0 + 120{,}1a_1 = 17{,}1 \\ 120{,}1a_0 + 1503{,}45a_1 = 236{,}91 \end{cases}$$

Отсюда: a***0*** = -4,4944; a1 = 0,5166.

Следовательно, $\bar{y}_x$ =-4,4944+0,5166 х.

Значения $\bar{y}_x$ в таблице 8.3 получены путем подстановки значений факторного признака xi (стоимость застрахованного автомобиля) в уравнение регрессии $\bar{y}_x$ =-4,4944+0,5166 х.

Коэффициент регрессии a1 = 0,5166 означает, что при увеличении стоимости застрахованного автомобиля на 1 тыс.долл.США, объем страхового возмещения (тыс.долл.США) возрастет в среднем на 0,5166 тыс.долл. США.

8.3. Множественная (многофакторная) регрессия

Изучение связи между тремя и более связанными между собой признаками носит название **множественной (многофакторной)** регрессии:

$$\bar{y}_{1,2,\dots,k} = f(x_1, x_2, \dots, x_k)$$

Построение моделей множественной регрессии включает несколько этапов:

1. Выбор формы связи (уравнения регрессии);

2. Отбор факторных признаков;

3. Обеспечение достаточного объема совокупности.

Выбор типа уравнения затрудняется тем, что для любой формы зависимости можно выбрать целый ряд уравнений, которые в определенной степени будут описывать эти связи. Основное значение имеют линейные модели в силу простоты и логичности их экономической интерпретации.

Важным этапом построения уже выбранного уравнения множественной регрессии является отбор и последующее включение факторных признаков.

С одной стороны, чем больше факторных признаков включено в уравнение, тем оно лучше описывает явление. Однако модель размерностью 100 и более факторных признаков сложно реализуема и требует больших затрат машинного времени. Сокращение размерности модели за счет исключения второстепенных, экономически и статистически несущественных факторов способствует простоте и качеству ее реализации. В то же время построение модели регрессии малой размерности может привести к тому, что такая модель будет недостаточно адекватна исследуемым явлениям и процессам.

Проблема отбора факторных признаков для построения моделей взаимосвязи может быть решена на основе интуитивно-логических или многомерных математико-статистических методов анализа.

Наиболее приемлемым способом отбора факторных признаков является **шаговая регрессия** (шаговый регрессионный анализ). Сущность метода шаговой регрессии заключается в реализации алгоритмов последовательного “включения”, “исключения” или “включения-исключения” факторов в уравнение регрессии и последующей проверке их статистической значимости. Алгоритм “включения” заключается в том, что факторы поочередно вводятся в уравнение так называемым «прямым методом». При проверке значимости введенного фактора определяется, на сколько уменьшается сумма квадратов остатков и увеличивается величина множественного коэффициента корреляции ($R2$). Одновременно используется и алгоритм последовательного «исключения», сущность которого заключается в том, что исключаются факторы, ставшие незначимыми по статистическим критериям.

Фактор является незначимым, если его включение в уравнение регрессии только изменяет значения коэффициентов регрессии, не уменьшая суммы квадратов остатков и не увеличивая их значения. Если при включении в модель соответствующего факторного признака величина множественного коэффициента корреляции увеличивается, а коэффициента регрессии не изменяется (или меняется несущественно), то данный признак существенен и его включение в уравнение регрессии необходимо. В противном случае, фактор нецелесообразно включать в модель регрессии.

При построении модели регрессии возможна проблема мультиколлинеарности, под которой понимается тесная зависимость между факторными признаками, включенными в модель $(r_{x_{ij}} > 0{,}8)$.

Наличие мультиколлинеарности между признаками вызывает:

- искажение величины параметров модели, которые имеют тенденцию к завышению, чем осложняется процесс определения наиболее существенных факторных признаков;
- изменение смысла экономической интерпретации коэффициентов регрессии.

В качестве причин возникновения мультиколлинеарности между признаками можно выделить следующие:

- изучаемые факторные признаки являются характеристикой одной и той же стороны изучаемого явления или процесса. Например: показатели объема производимой продукции и среднегодовой стоимости основных фондов одновременно включать в модель не рекомендуется, так как они оба характеризуют размер предприятия;
- факторные признаки являются составляющими элементами друг друга. Например: показатели выработки продукции на одного работающего и численность работающих одновременно в модель включать нельзя, так как в основе расчета показателей лежит один и тот же показатель – численность работающих на предприятии.
- факторные признаки по экономическому смыслу дублируют друг друга.

Устранение мультиколлинеарности может реализовываться через исключение из корреляционной модели одного или нескольких линейно-связанных факторных признаков или преобразование исходных факторных признаков в новые, укрупненные факторы.

Вопрос о том, какой из факторов следует отбросить, решается на основании качественного, логического анализа изучаемого явления, а также на основе анализа тесноты связи между результативным (у) с каждым из сильно коллинеарно связанных факторных признаков. Из дальнейшего анализа целесообразно исключить тот факторный признак, связь которого с результативным наименьшая.

Качество уравнения регрессии зависит от степени достоверности и надежности исходных данных и объема совокупности. Исследователь должен стремиться к увеличению числа наблюдений, так как большой объем наблюдений является одной из предпосылок построения адекватных статистических моделей.

Аналитическая форма связи результативного признака от нескольких факторных выражается и называется многофакторным (множественным) уравнением регрессии или моделью связи.

Линейное уравнение множественной регрессии имеет вид:

$$\overline{y}_{1,2,...,k} = a_0 + a_1 x_1 + a_2 x_2 + \ldots + a_k x_k$$

где $\overline{y}_{1,2,3,...,k}$ - теоретические значения результативного признака, полученные в результате подстановки соответствующих значений факторных признаков в уравнение регрессии;

$x_1, x_2, \ldots, x_k$ - факторные признаки;

$a_1, a_2, \ldots, a_k$ - параметры модели (коэффициенты регрессии).

Параметры уравнения могут быть определены графическим методом или методом наименьших квадратов.

Пример.

По следующим данным о выручке (y), спросу по номиналу ($x1$) и объему продаж ($x2$) корпоративных ценных бумаг определим зависимость между признаками.

Таблица 8.4.

Основные характеристики корпоративных ценных бумаг

Серия ценной бумаги	Выручка, млрд. руб., у	Спрос по номиналу, млрд. руб., х1	Объем продаж по номиналу, млрд. руб., х2
0001	3,0	6,8	3,5
0002	5,4	11,2	6,7
0003	5,9	9,1	6,8
0004	4,8	6,9	5,9
0005	3,3	6,4	3,8
0006	3,4	6,9	4,3
0007	5,3	12,2	6,9
Итого	**31,1**	**59,5**	**37,9**

Система нормальных линейных уравнений имеет вид:

$$\begin{cases} na_0 + a_1 \sum x_1 + a_2 \sum x_2 = \sum y \\ a_0 \sum x_1 + a_1 \sum x_1^2 + a_2 \sum x_1 x_2 = \sum x_1 y \\ a_0 \sum x_2 + a_1 \sum x_1 x_2 + a_2 \sum x_2^2 = \sum x_2 y \end{cases}$$

Для определения параметров линейного уравнения регрессии составим расчетную таблицу:

Таблица 8.5.

Расчетная таблица для определения параметров уравнения регрессии выручки от реализации корпоративных ценных бумаг

Серия ценной бумаги	Выручка, млрд. руб., y	Спрос по номиналу, млрд. руб., x1	Объем продаж по номиналу, млрд. руб., x2	x_1^2	x1 x2	x1 y	x_2^2	x2 y
0001	3,0	6,8	3,5	46,24	23,80	20,40	12,25	10,50
0002	5,4	11,2	6,7	125,44	75,04	60,48	44,89	36,18
0003	5,9	9,1	6,8	82,81	61,88	53,69	46,24	40,12
0004	4,8	6,9	5,9	47,61	40,71	33,12	34,81	28,32
0005	3,3	6,4	3,8	40,96	24,32	21,12	14,44	12,54
0006	3,4	6,9	4,3	47,61	29,67	23,46	18,49	14,62
0007	5,3	12,2	6,9	148,84	84,18	64,66	47,61	36,57
Итого	**31,1**	**59,5**	**37,9**	**539,51**	**339,6**	**276,93**	**218,73**	**178,85**

Система уравнений примет следующий вид:

$$\begin{cases} 7a_0 + 59{,}5a_1 + 37{,}9a_2 = 31{,}1 \\ 59{,}5a_0 + 539{,}51a_1 + 339{,}6a_2 = 276{,}93 \\ 37{,}9a_0 + 339{,}60a_1 + 218{,}73a_2 = 178{,}85 \end{cases}$$

Таким образом:

$$\bar{y}_{x_1,x_2} = 0{,}378 - 0{,}082x_1 + 0{,}879x_2.$$

8.4. Собственно-корреляционные параметрические методы изучения связи

Измерение тесноты (силы) и направления связи является важной задачей изучения и количественного измерения взаимосвязи социально-экономических явлений. Оценка тесноты связи между признаками предполагает определение меры соответствия вариации результативного признака и одного (при изучении парных зависимостей) или нескольких (множественных зависимостей) факторных признаков.

Линейный коэффициент корреляции (К. Пирсона) характеризует тесноту и направление связи между двумя коррелируемыми признаками в случае наличия между ними линейной зависимости.

В теории разработаны и на практике применяются различные модификации формулы расчета данного коэффициента:

$$r = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\sigma_x \cdot \sigma_y} \tag{8.4}$$

Производя расчет по итоговым значениям исходных переменных, линейный коэффициент корреляции можно вычислить по формуле:

$$r = \frac{n\sum xy - \sum x \sum y}{\sqrt{\left[n\sum x^2 - \left(\sum x\right)^2\right] \cdot \left[n\sum y^2 - \left(\sum y\right)^2\right]}} \quad (8.5)$$

Между линейным коэффициентом корреляции и коэффициентом регрессии существует определенная зависимость, выражаемая формулой:

$$r = a_i \frac{\sigma_{x_i}}{\sigma_y} \quad (8.6)$$

где a_i - коэффициент регрессии в уравнении связи;

σ_{x_i} - среднее квадратическое отклонение соответствующего, статистически существенного, факторного признака.

Линейный коэффициент корреляции изменяется в пределах от -1 до 1: $[-1 \le r \le 1]$. Знаки коэффициентов регрессии и корреляции совпадают. При этом интерпретацию выходных значений коэффициента корреляции можно осуществлять следующим образом (табл.8.6):

Таблица 8.6.

Оценка линейного коэффициента корреляции

Значение линейного коэффициента связи	Характеристика связи	Интерпретация связи
r = 0	отсутствует	-
0<r<1	прямая	с увеличением *x* увеличивается *y*
-1<*r*<0	обратная	с увеличением *x* уменьшается *y* и наоборот
r=1	функциональная	каждому значению факторного признака строго соответствует одно значение результативного признака

Пример

На основе выборочных данных о деятельности 6 предприятий одной из отраслей промышленности Российской Федерации оценить тесноту связи между трудоемкостью продукции предприятия (X, чел.-час.) и объемом ее производства (Y, млн. руб.)

Таблица 8.7.

Расчетная таблица для определения коэффициента корреляции

№ п/п	Объем произведенной продукции, млн. руб., Y	Затраты на 100 изделий, чел.-час, X	yx	y2	x2
1	221	96	21216	48841	9216
2	1070	77	82390	1144900	5929
3	1001	77	77077	1002000	5929
4	606	89	53934	367236	7921
5	779	82	63878	606841	6724
6	789	81	63909	622520	6561
Сумма	4466	502	362404	3792338	42280
Средняя	744,33	83,67	60400,67	632056,33	7046,67

1. Используя формулу (8.4) получаем:

$$\sigma_y^2 = \overline{y}^2 - (\overline{y})^2 = 632056{,}3 - (744{,}3)^2 = 78029{,}3$$

$$\sigma_x^2 = \overline{x}^2 - (\overline{x})^2 = 7046{,}67 - (83{,}67)^2 = 46$$

$$r = \frac{60400{,}67 - 744{,}33 \cdot 83{,}67}{\sqrt{78029{,}3 \cdot 46}} = -0{,}98$$

2. По формуле (8.5) значение коэффициента корреляции составило:

$$r = \frac{6 \cdot 362404 - 4466 \cdot 502}{\sqrt{\left[6 \cdot 42280 - (502)^2\right] \cdot \left[6 \cdot 3792338 - (4466)^2\right]}} =$$

$$= \frac{2174424 - 2241932}{\sqrt{(253680 - 252004) \cdot (22754028 - 19945156)}} =$$

$$= \frac{-67508}{\sqrt{1676 \cdot 2808872}} = \frac{-67508}{68612{,}46} = -0{,}98$$

Таким образом, результат по всем формулам одинаков и свидетельствует о сильной обратной зависимости между изучаемыми признаками.

В случае наличия линейной или нелинейной зависимости между двумя признаками для измерения тесноты связи применяют так называемое **корреляционное отношение.** Различают эмпирическое и теоретическое корреляционное отношение.

Эмпирическое корреляционное отношение рассчитывается по данным группировки, когда δ^2 характеризует отклонения групповых средних результативного показателя от общей средней:

$$\eta = \sqrt{\frac{\sigma^2 - \bar{\sigma}^2}{\sigma^2}} = \sqrt{1 - \frac{\bar{\sigma}^2}{\sigma^2}} = \sqrt{\frac{\delta^2}{\sigma^2}} \tag{8.7}$$

где η - корреляционное отношение;

σ^2 - общая дисперсия;

$\bar{\sigma}^2$ - средняя из частных (групповых) дисперсий;

δ^2 - межгрупповая дисперсия (дисперсия групповых средних).

Все эти дисперсии есть дисперсии результативного признака.

Теоретическое корреляционное отношение определяется по формуле:

$$\eta = \sqrt{\frac{\delta^2}{\sigma^2}} = \sqrt{1 - \frac{\sigma^2_{ост}}{\sigma^2}} \tag{8.8}$$

где δ^2 - дисперсия выровненных значений результативного признака, то есть рассчитанных по уравнению регрессии;

σ^2 - дисперсия эмпирических (фактических) значений результативного признака;

$\sigma^2_{ост}$ - остаточная дисперсия.

Корреляционное отношение изменяется в пределах от 0 до 1 $(0 \le \eta \le 1)$.

Для измерения тесноты связи при множественной корреляционной зависимости, то есть при исследовании трех и более признаков одновременно, вычисляется множественный и частные коэффициенты корреляции.

Множественный коэффициент корреляции вычисляется при наличии линейной связи между результативным и несколькими факторными признаками, а также между каждой парой факторных признаков.

Множественный коэффициент корреляции для двух факторных признаков вычисляется по формуле:

$$R_{y/x_1x_2} = \sqrt{\frac{r^2_{yx_1} + r^2_{yx_2} - 2r_{yx_1} \cdot r_{yx_2} \cdot r_{x_1x_2}}{1 - r^2_{x_1x_2}}} \tag{8.9}$$

где r_{yx_i} - парные коэффициенты корреляции между признаками.

Множественный коэффициент корреляции изменяется в пределах от 0 до 1 и по определению положителен: $0 \le R \le 1$. Приближение R к единице свидетельствует о сильной зависимости между признаками.

На основе данных таблицы 8.4 рассчитаем коэффициент множественной корреляции:

$$r_{yx_1} = \frac{\overline{yx_1} - \bar{y} \cdot \bar{x}_1}{\sigma_y \cdot \sigma_{x_1}} = 0{,}748, \quad r_{yx_2} = \frac{\overline{yx_2} - \bar{y} \cdot \bar{x}_2}{\sigma_y \cdot \sigma_{x_2}} = 0{,}983;$$

$$r_{x_1 x_2} = \frac{x_1 x_2 - x_1 * x_2}{\sigma_{x_1} \cdot \sigma_{x_2}} = 0{,}817.$$

Множественный коэффициент корреляции составит:

$$R_{y/x_1x_2} = \sqrt{\frac{-0{,}748^2 + 0{,}983^2 - 2 \cdot 0{,}748 \cdot 0{,}983 \cdot 0{,}817}{1 - 0{,}817^2}} = 0{,}975.$$

Частные коэффициенты корреляции характеризуют степень тесноты связи между двумя признаками x_1 и x_2 при фиксированном значении других $(k-2)$ факторных признаков, то есть когда влияние x_3 исключается, то есть оценивается связь между x_1 и x_2 в «чистом виде».

В случае зависимости y от двух факторных признаков x_1 и x_2 коэффициенты частной корреляции имеют вид:

$$r_{yx_1/x_2} = \frac{r_{yx_1} - r_{x_1x_2} \cdot r_{yx_2}}{\sqrt{\left(1 - r^2_{x_2y}\right) \cdot \left(1 - r^2_{x_1x_2}\right)}}$$

$$r_{yx_2/x_1} = \frac{r_{yx_2} - r_{x_1y} \cdot r_{x_1x_2}}{\sqrt{\left(1 - r^2_{x_1y}\right) \cdot \left(1 - r^2_{x_1x_2}\right)}} \tag{8.10}$$

где r - парные коэффициенты корреляции между указанными в индексе переменными.

В первом случае исключено влияние факторного признака x_2, во втором - x_1.

На основании приведенных выше данных о зависимости трех факторов деятельности предприятий вычислим частные коэффициенты корреляции (табл. 8.4):

$$r_{yx_1/x_2} = \frac{0{,}748 - 0{,}817 \cdot 0{,}983}{\sqrt{\left(1 - 0{,}983^2\right)} \cdot \sqrt{\left(1 - 0{,}817^2\right)}} = -0{,}517$$

$$r_{yx_2/x_1} = 0{,}972; \quad r_{x_1x_2/y} = 0{,}668.$$

8.5. Принятие решений на основе уравнений регрессии

Интерпретация моделей регрессии осуществляется методами той отрасли знаний, к которой относится исследуемое явление. Но всякая интерпретация начинается со статистической оценки уравнения регрессии в целом и оценки значимости входящих в модель факторных признаков.

Прежде всего необходимо рассмотреть коэффициенты регрессии. Чем больше величина коэффициента регрессии, тем значительнее влияние данного признака на моделируемый.

Знаки коэффициентов регрессии говорят о характере влияния на результативный признак. Если факторный признак имеет знак плюс, то с увеличением данного фактора результативный признак возрастает; если факторный признак имеет знак минус, то с его увеличением результативный признак уменьшается.

Если экономическая теория подсказывает, что факторный признак должен иметь положительное значение, а он имеет знак минус, то необходимо проверить расчеты параметров уравнения регрессии. Такое явление чаще всего бывает в силу допущенных ошибок при решении. Однако следует иметь ввиду, что когда рассматривается совокупное влияние факторов, то в силу наличия взаимосвязей между ними характер их влияния может меняться.

С целью расширения возможностей экономического анализа, используются **частные коэффициенты эластичности**, определяемые по формуле:

$$Э_{x_i} = a_1 \cdot \frac{\bar{x}_i}{\bar{y}} \qquad (8.11)$$

где $\bar{x}_i$ - среднее значение соответствующего факторного признака;

$\bar{y}$ - среднее значение результативного признака;

a_1 - коэффициент регрессии при соответствующем факторном признаке.

Коэффициент эластичности показывает на сколько процентов в среднем изменится значение результативного признака при изменении факторного признака на 1%.

Рассчитаем коэффициент эластичности ($Э_{x_i}$) по исходным данным о зависимости между выручкой (y), спросом по номиналу (x_1) и объемом продаж по номиналу (x_2) корпоративных ценных бумаг одной из корпораций, приведенным в таблице 8.4.

$a_1 = -0{,}082$; $a_2 = 0{,}879.$

$\bar{y} = \frac{31{,}1}{7} = 4{,}44$; $\bar{x}_1 = \frac{59{,}5}{7} = 8{,}5;$ $\bar{x}_2 = \frac{37{,}9}{7} = 5{,}41.$

$$Э_{x_1} = a_1 \cdot \frac{\bar{x}_1}{\bar{y}} = -0{,}082 \cdot \frac{8{,}5}{4{,}44} = -0{,}16; \quad Э_{x_2} = a_2 \frac{\bar{x}_2}{\bar{y}} = 0{,}879 \cdot \frac{5{,}41}{4{,}44} = 1{,}07.$$

Это значит, что при увеличении спроса по номиналу на ценные бумаги на 1%, выручка от их реализации снизится на 0,16%, а при увеличении объема продаж по номиналу на 1%, выручка увеличится на 1,07%.

Частный коэффициент детерминации:

$$d_{x_i} = r_{yx_i} \cdot \beta_{x_i} \qquad (8.12)$$

где r_{yx_i} - парный коэффициент корреляции между результативным и i-ым факторным признаком;

β_{x_i} - соответствующий стандартизованный коэффициент уравнения множественной регрессии:

$$\beta_{x_1} = a_1 \cdot \frac{\sigma_{x_i}}{\sigma_y} \qquad (8.13)$$

Частный коэффициент детерминации показывает на сколько процентов вариация результативного признака объясняется вариацией i-го признака, входящего в множественное уравнение регрессии.

По данным, приведенным в таблице 8.4 рассчитаем частный коэффициент детерминации для фактора x_1 - спрос по номиналу на ценные бумаги:

$$d_{x_1} = r_{yx_1}\beta_{x_1}; \quad r_{yx_1} = \frac{\overline{yx_1} - \bar{y} \cdot \bar{x}}{\sigma_y \cdot \sigma_{x_1}}; \quad \beta_{x_1} = a_1 \frac{\sigma_{x_1}}{\sigma_y}$$

$$\overline{yx_1} = \frac{\sum yx_1}{n} = \frac{276{,}93}{7} = 39{,}56; \quad \bar{y} = \frac{\sum y}{n} = \frac{31{,}1}{7} = 4{,}44;$$

$$\bar{x}_1 = \frac{\sum x_1}{n} = \frac{59{,}5}{7} = 8{,}5;$$

$$\sigma_y^2 = \overline{y^2} - (\bar{y})^2 = \frac{146{,}55}{7} - (4{,}44)^2 = 1{,}23;$$

$$\sigma_x^2 = \bar{x}_1^2 - (\bar{x}_1)^2 = \frac{539{,}51}{7} - (8{,}5)^2 = 4{,}82;$$

$$\sigma_y = \sqrt{\sigma_y^2} = \sqrt{1{,}23} = 1{,}109; \quad \sigma_x = \sqrt{4{,}82} = 2{,}195$$

$$r_{yx_1} = \frac{39{,}56 - 4{,}44 \cdot 8{,}5}{1{,}109 \cdot 2{,}195} = 0{,}748$$

$$\beta_{x_1} = -0{,}082 \cdot \frac{2{,}195}{1{,}109} = -0{,}16; \quad d_{x_1} = 0{,}748 \cdot (-0{,}16) = -0{,}12.$$

Определим частный коэффициент детерминации для фактора x_2 - объем продаж ценных бумаг по номиналу:

$$d_{x_2} = r_{yx_2} \cdot \beta_{x_2} = 0{,}006.$$

$r_{y\,x_2} = 0{,}983;$ $\beta_{x2} = a2 \cdot \frac{\sigma_{x2}}{\sigma_y} = 0{,}879 \cdot \frac{1{,}390}{1{,}109} = 1{,}10;$ $d_{x_2} = 0{,}983 \cdot 1{,}10 \approx 1{,}10.$

Полная экономическая интерпретация моделей регрессии позволяет выявить резервы развития и повышения деловой активности субъектов рыночной экономики.

8.6. Методы изучения связи качественных признаков

При наличии соотношения между вариацией качественных признаков говорят об их ассоциации, взаимосвязанности. Для оценки связи в этом случае используют ряд показателей.

Коэффициент ассоциации и контингенции. Для определения тесноты связи двух качественных признаков, каждый из которых состоит только из двух групп, применяются коэффициенты ассоциации и контингенции.

Для их вычисления строится таблица, которая показывает связь между двумя явлениями, каждое из которых должно быть альтернативным, то есть состоящим из двух качественно отличных друг от друга значений признака (например, изделие годное или бракованное).

Таблица 8.8.

Таблица для вычисления коэффициентов ассоциации и контингенции

a	b	a+b
c	d	c+d
a+c	b+d	a+b+c+d

Коэффициенты вычисляются по формулам:

ассоциации: $$K_a = \frac{ad - bc}{ad + bc} \qquad (8.14)$$

контингенции: $$K_k = \frac{ad - bc}{\sqrt{(a+b)\cdot(b+d)\cdot(a+c)\cdot(c+d)}} \qquad (8.15)$$

Коэффициент контингенции всегда меньше коэффициента ассоциации. Связь считается подтвержденной, если $K_a \geq 0{,}5$ или $K_k \geq 0{,}3$.

Пример. Исследуем связь между участием населения одного из городов в экологических акциях и уровнем его образования. Результаты обследования характеризуются следующими данными:

Таблица 8.9.

Зависимость участия населения города в экологических акциях от образовательного уровня

Группы рабочих	Числен-ность населения города	Из них	
		участвую-щих в акциях	**не участвующих в акциях**
Имеют среднее образование	100	78	22
Не имеют среднего образования	100	32	68
Итого	200	110	90

$$K_a = \frac{78\cdot 68 - 32\cdot 22}{78\cdot 68 + 32\cdot 22} = \frac{4600}{6608} = 0{,}766$$

$$K_k = \frac{78\cdot 68 - 32\cdot 22}{\sqrt{(78+22)\cdot(22+68)\cdot(78+32)\cdot(32+68)}} = \frac{5304-704}{\sqrt{99000000}} = 0{,}46$$

Таким образом, связь между участием населения города в экологических акциях и его образовательным уровнем имеет место, но не столь существенна.

Когда каждый из качественных признаков состоит более чем из двух групп, то для определения тесноты связи возможно применение **коэффициентов взаимной сопряженности Пирсона** и **Чупрова.** Эти коэффициенты вычисляются по следующим формулам:

$$K_n = \sqrt{\frac{\varphi^2}{1+\varphi^2}}\,; \qquad K = \sqrt{\frac{\varphi^2}{\sqrt{(K_1-1)\cdot(K_2-1)}}} \qquad (8.16)$$

где φ^2 - показатель взаимной сопряженности;

φ - определяется как сумма отношений квадратов частот каждой клетки таблицы к произведению итоговых частот соответствующего столбца и строки. Вычитая из этой суммы «1», получим величину φ^2:

$$\varphi^2 = \sum\frac{n_{xy}^2}{n_x n_y} - 1\,;$$

K_1 - число значений (групп) первого признака;

K_2 - число значений (групп) второго признака.

Чем ближе величина K_n и $K\text{ч}$ к 1, тем теснее связь.

Таблица 8.10.

Вспомогательная таблица для расчета коэффициента взаимной сопряженности

у / х	I	II	III	Всего
I			n_{xy}	n_x
II				n_x
III				n_x
Итого	n_y	n_y	n_y	n

$$1+\varphi^2 = \sum \frac{\sum \frac{n_{xy}^2}{n_x}}{n_y} = \sum \frac{\sum \frac{n_{xy}^2}{n_y}}{n_x}$$

Пример.

С помощью коэффициента взаимной сопряженности исследуем связь между себестоимостью продукции и накладными расходами на ее реализацию.

Таблица 8.11.

Зависимость между себестоимостью продукции и накладными расходами на ее реализацию

Накладные расходы	Себестоимость			Итого
	Низкая	Средняя	Высокая	
Низкие	19	12	9	40
Средние	7	18	15	40
Высокие	4	10	26	40
Итого	**30**	**40**	**50**	**120**

$$1+\varphi^2 = \frac{\frac{19^2}{30}+\frac{12^2}{40}+\frac{9^2}{50}}{40} + \frac{\frac{7^2}{30}+\frac{18^2}{40}+\frac{15^2}{50}}{40} + \frac{\frac{4^2}{30}+\frac{10^2}{40}+\frac{26^2}{50}}{40} =$$

$$= 0{,}431 + 0{,}356 + 0{,}414 = 1{,}183$$

$1+\varphi^2 = 1{,}183$; $\varphi^2 = 0{,}183$

$K_n = \sqrt{\frac{0{,}183}{1{,}183}} = \sqrt{0{,}155} = 0{,}39$; Кч= $\sqrt{\frac{0{,}183}{\sqrt{2\cdot 2}}} = 0{,}21$.

Связь слабая.

Особое значение для оценки связи имеет **биссериальный коэффициент корреляции,** который дает возможность оценить связь между качественным альтернативным и количественным варьирующим признаками. Данный коэффициент вычисляется по формуле:

$$r = \frac{\left|\overline{y_2} - \overline{y_1}\right|}{\sigma_y} \cdot \frac{pq}{Z} \tag{8.17}$$

где $\overline{y_2}$ и $\overline{y_1}$ - средние в группах;

σ_y - среднее квадратическое отклонение фактических значений признака от среднего уровня;

p - доля первой группы;

q - доля второй группы;

Z - табулированные (табличные) значения Z-распределения в зависимости от p.

Пример.

Распределение предприятий одной из отраслей промышленности по уровню дохода и источникам средств существования характеризуется следующими данными:

Таблица 8.12.

Зависимость уровня доходов сотрудников коммерческой структуры от уровня их образования

Источник средств	Уровень доходов, (млн.руб.)				**Всего**
	200-300	300-400	400-500	500-600	
	250	350	450	550	
Банковский кредит Собственные средства	5 9	7 4	6 2	4 1	22 16
Итого	**14**	**11**	**8**	**5**	**38**

$$\overline{y_1} = \frac{250 \cdot 5 + 350 \cdot 7 + 450 \cdot 6 + 550 \cdot 4}{22} = \frac{8600}{22} = 390{,}9$$

$$\overline{y_1} = \frac{250 \cdot 9 + 350 \cdot 4 + 450 \cdot 2 + 550 \cdot 1}{16} = \frac{5100}{16} = 318{,}8$$

$$\bar{y}_{общ} = \frac{250 \cdot 14 + 350 \cdot 11 + 450 \cdot 8 + 550 \cdot 5}{38} = \frac{13700}{38} = 360{,}5$$

$\sigma = 104{,}7$; $Z_{табл} = 0{,}3975$

$p = \frac{22}{38} = 0{,}58$; $q = 0{,}42$; $p \cdot \frac{q}{Z} = 0{,}58 \cdot \frac{0{,}42}{0{,}3975} = 0{,}61$

$$r = \frac{|318{,}8 - 390{,}9|}{104{,}7} \cdot 0{,}61 = 0{,}42.$$

Величина биссериального коэффициента корреляции также подтверждает умеренную тесноту связи между изучаемыми признаками.

8.7. Ранговые коэффициенты связи

В анализе социально-экономических явлений часто приходится прибегать к различным условным оценкам с помощью рангов, а взаимосвязь между отдельными признаками измерять с помощью непараметрических коэффициентов связи.

Ранжирование - это процедура упорядочения объектов изучения, которая выполняется на основе предпочтения. Ранг - это порядковый номер значений признака, расположенных в порядке возрастания или убывания их величин. Если значения признака имеют одинаковую количественную оценку, то ранг всех этих значений принимается равным средней арифметической из соответствующих номеров мест, которые они определяют. Данные ранги называются связными.

Среди непараметрических методов оценки тесноты связи наибольшее значение имеют ранговые коэффициенты Спирмена(ρ_{xy}) и Кендалла (τ_{xy}). Эти коэффициенты могут быть использованы для определения тесноты связи как между количественными, так и между качественными признаками (рейтинги, уровни образования, квалификации и т.п.).

Коэффициент корреляции рангов (коэффициент **Спирмена**) рассчитывается по формуле:

$$\rho_{xy} = 1 - \frac{6\sum d_i^2}{n(n^2 - 1)} \tag{8.18}$$

где d_i^2 - квадраты разности рангов;

n - число наблюдений (число пар рангов).

Коэффициент Спирмена принимает значения в интервале $[-1;\ 1]$.

Пример.

По данным о прибыли и объеме кредитных вложений 10 коммерческих банков одного из регионов Российской Федерации на 01.01.2004 г. определить с помощью коэффициента Спирмена зависимость между этими признаками.

Таблица 8.13.

Расчет коэффициента Спирмена

№ банка	Кредитные вложения, млн. руб., X	Прибыль, млн.руб., Y	Ранги		Разность рангов $d_i = R_x - R_y$	d_i^2
			R_x	R_y		
1	2	3	8	9	10	11
1	2887	557	9	7	2	4
2	1710	605	1	9	-8	64
3	3010	628	10	10	0	0
4	2472	488	6	5	1	1
5	2535	418	7	3	4	16
6	1897	397	4	2	2	4
7	2783	501	8	6	2	4
8	1862	589	3	8	-5	25
9	1800	269	2	1	1	1
10	2003	437	5	4	1	1
Итого	**-**	**-**	**-**	**-**	**-**	**120**

$\rho_{x/y} = 1 - \frac{6 \cdot 120}{10 \cdot 99} = 1 - \frac{720}{990} = 0{,}3$ **(связь слабая).**

Ранговый коэффициент корреляции Кендалла (τ_{xy}) также может использоваться для измерения взаимосвязи между качественными и количественными признаками, характеризующими однородные объекты и ранжированные по одному принципу. Расчет рангового коэффициента Кендалла осуществляется по формуле:

$$\tau = \frac{2S}{n(n-1)} \qquad (8.19)$$

где n - число наблюдений;

S - сумма разностей между числом последовательностей и числом инверсий по второму признаку.

Расчет данного коэффициента выполняется в следующей последовательности:

1. Значения X ранжируются в порядке возрастания или убывания;

2. Значения Y располагаются в порядке, соответствующем значениям X;

3. Для каждого ранга Y определяется число следующих за ним значений рангов, превышающих его величину. Суммируя таким образом числа определяется величина P, как мера соответствия последовательностей рангов по X и Y и учитывается со знаком (+);

4. Для каждого ранга Y определяется число следующих за ним значений рангов, меньших его величины. Суммарная величина обозначается через Q и фиксируется со знаком (-);

5. Определяется сумма баллов по всем членам ряда.

В приведенном примере (таблица 8.11)

$P = 1 + 8 + 1 + 6 + 4 + 3 + 3 + 2 + 1 = 29$

$Q = (-8) + 0 + (-6) + 0 + (-1) + (-1) + 0 + 0 + 0 = -16$

Таким образом:

$$\tau = \frac{2 \cdot (29 - 16)}{10 \cdot (10 - 1)} = 0{,}28$$

что свидетельствует о практическом отсутствии связи между рассматриваемыми признаками по данной совокупности коммерческих банков.

Как правило, коэффициент Кендалла меньше коэффициента Спирмена. При достаточно большом объеме совокупности значения данных коэффициентов имеют следующую зависимость:

$$\tau = \frac{2}{3} \rho_{x/y}$$

Связь между признаками признается статистически значимой, если значения коэффициентов ранговой корреляции Спирмена и Кендалла больше 0,5.

Для определения тесноты связи между произвольным числом ранжированных признаков применяется множественный коэффициент ранговой корреляции (коэффициент конкордации) W, который вычисляется по формуле:

$$W = \frac{12S}{m^2 \cdot (n^3 - n)} \tag{8.19}$$

где m - количество факторов

n - число наблюдений

S - отклонение суммы квадратов рангов от средней квадратов рангов.

Пример.

Определим тесноту связи между объемом реализованной продукции, прибылью и численностью работающих по 10 предприятиям отрасли.

Таблица 8.14.

Расчет коэффициента конкордации

№ пред-прия-тия	Уставной капитал, млн. руб., X	Число выставлен-ных акций, Y	Число занятых на предприя-тиях, Z	R_x	R_y	R_Z	Сумма строк	Квад-раты сумм
1	3069	871	320	9	7	1	17	289
2	1720	945	326	1	9	2	12	144
3	4217	1578	333	10	10	3	23	529
4	2465	697	342	6	5	4	15	225
5	2740	631	351	7	3	5	15	225
6	1910	510	366	4	2	6	12	144
7	2928	830	379	8	6	7	21	441
8	1866	873	382	3	8	8	19	361
9	1815	482	402	2	1	9	12	144
10	2379	676	405	5	4	10	19	361
Итого	**-**	**-**	**-**	**-**	**-**	**-**	**165**	**2863**

$$S = 2863 - \frac{(165)^2}{10} = 2863 - 2722{,}5 = 140{,}5$$

$$W = \frac{12S}{m^2(n^3 - n)} = \frac{12 \cdot 140{,}5}{9(1000 - 10)} = 0{,}19$$

,

что свидетельствует о слабой связи между рассматриваемыми признаками.

Ранговые коэффициенты Спирмена, Кендалла и конкордации имеют то преимущество, что с помощью их можно измерять и оценивать связи как между количественными так и между атрибутивными признаками, которые поддаются ранжированию.

Глава 9. Статистическое изучение динамики социально-экономических явлений

9.1. Понятие о рядах динамики и их виды

Процессы и явления социально-экономической жизни общества, являющиеся предметом изучения статистики, находятся в постоянном движении и изменении. Для того, чтобы выявить тенденции и закономерности социально-экономического развития явлений, статистика строит особые ряды статистических показателей, которые называются **рядами динамики** (иногда их называют временными рядами), то есть – это ряды изменяющихся во времени значений статистического показателя, расположенных в хронологическом порядке. В англоязычной литературе для временных рядов используется термин “time series”.

Ряды динамики получаются в результате сводки и обработки материалов периодического статистического наблюдения. Повторяющиеся во времени (по отчетным периодам) значения одноименных показателей в ходе статистической сводки (гл.2) систематизируются в хронологической последовательности. Значения показателя, составляющие ряд динамики, называются **уровнями ряда**.

Каждый ряд динамики характеризуется двумя параметрами: значениями времени и соответствующими им значениями уровней ряда. Уровни ряда обычно обозначаются “yt”: y1, y2 и т.д. В качестве показателя времени в рядах динамики могут указываться отдельные периоды (сутки, месяцы, кварталы, годы и т.д.) времени или определенные моменты (даты). Время в рядах динамики обозначается через “t”.

Ряды динамики, как правило, представляют в виде таблицы или графически.

Ряды динамики могут быть классифицированы по следующим признакам:

- В зависимости от способа выражения уровней ряды динамики подразделяются на ряды **абсолютных, относительных** и **средних** величин. При этом ряды динамики абсолютных величин рассматриваются как исходные, а ряды относительных и средних величин - как производные.

Ряды динамики абсолютных величин наиболее полно характеризуют развитие процесса или явления, например, грузооборота транспорта, инвестиций в основной капитал, добычи топлива, уставного капитала коммерческих банков и т.д.

Ряды относительных величин могут характеризовать во времени темпы роста (или снижения) определенного показателя; изменение удельного веса того или иного показателя в совокупности или изменение показателей интенсивности отдельных явлений, например, удельного веса приватизированных предприятий в той или иной отрасли; производства продукции на душу населения; структуры инвестиций в основной капитал по отраслям экономики, индекса потребительских цен и т.д.

Ряды динамики средних величин служат для характеристики изменения уровня явления, отнесенного к единице совокупности, например: данные о среднегодовой численности занятых в экономике; о средней урожайности отдельных сельскохозяйственных культур, о средней заработной плате в отдельных отраслях и т.д.

- В зависимости от характера временного параметра ряды динамики делятся на **моментные** и **интервальные**.

Уровни **моментных** рядов динамики характеризуют явление по состоянию на определенный момент времени.

Пример. Моментный ряд динамики, характеризующий численность персонала строительной фирмы на 1-е число каждого месяца за первое полугодие 2004 г., представлен в таблице 9.1.

Таблица 9.1.

Дата	1.01	1.02	1.03	1.04	1.05	1.06
Численность персонала, чел.	780	810	880	930	940	970

Следует помнить, что моментные ряды абсолютных величин нельзя суммировать. Бессмысленно, например, складывать численность персонала по состоянию на 1 января, 1 февраля и т.д. Полученная сумма ничего не выражает, так как в ней многократно повторяются одни и те же единицы совокупности.

Ряд, в котором уровни характеризуют результат, накопленный или вновь произведенный за определенный интервал времени, называется **интервальным**.

Пример. Интервальный ряд динамики, характеризующий динамику объема розничного товарооборота во всех каналах реализации в регионе, представлен в таблице 9.2.

Таблица 9.2.

Годы	2000	2001	2002	2003	2004
Товарооборот, млн.руб.	28,3	31,9	38,3	42,3	45,2

Важное аналитическое отличие моментных рядов от интервальных состоит в том, что сумма уровней интервального ряда вполне реальный показатель, например, общий объем розничного товарооборота за 2000-2004 г.г.

- В зависимости от расстояния между уровнями, ряды динамики подразделяются на ряды **с равноотстоящими уровнями** и **неравноотстоящими уровнями** во времени. Ряды динамики следующих друг за другом периодов или следующих через определенные промежутки дат называются равноотстоящими (табл. 9.1 и табл. 9.2). Если же в рядах даются прерывающиеся периоды или неравномерные промежутки между датами, то ряды называются не равноотстоящими (табл. 9.3).

Пример. Рядом динамики с не равноотстоящими уровнями во времени может служить объем экспорта продукции предприятия, представленный в таблице 9.3.

Таблица 9.3.

Годы	1993	1996	1998	2000	2004
Объем экспорта, млн.долл.	1110	1220	1320	1450	1640

- По числу показателей можно выделить **изолированные (одномерные)** и **комплексные (многомерные)** ряды динамики. Если ведется анализ во времени одного показателя ряда, то ряд динамики изолированный (например, данные о производстве газа по годам). В многомерном ряду представлена динамика нескольких показателей, характеризующих одно явление.

9.2. Сопоставимость уровней и смыкание рядов динамики

Важнейшим условием правильного построения рядов динамики является сопоставимость всех входящих в него уровней. Данное условие решается либо в процессе сбора и обработки данных, либо путем их пересчета.

Рассмотрим основные причины несопоставимости уровней ряда динамики.

Несопоставимость уровней ряда может возникнуть вследствие изменения **единиц измерения и единиц счета.** Нельзя сравнивать и анализировать цифры о производстве тканей, если за одни годы оно дано в погонных метрах, а за другие – в квадратных метрах.

На сопоставимость уровней ряда динамики непосредственно влияет **методология учета или расчета показателей.** Например, если в одни годы среднюю урожайность считали с засеянной площади, а в другие – с убранной, то такие уровни будут несопоставимы.

В процессе развития во времени, прежде всего, происходят количественные изменения явлений, а затем на определенных ступенях совершаются качественные скачки, приводящие к изменению закономерностей явления. Поэтому научный подход к изучению рядов динамики заключается в том, чтобы ряды, охватывающие большие периоды времени, разделять на такие, которые бы объединяли лишь однокачественные периоды развития совокупности, характеризующейся одной закономерностью развития.

Процесс выделения однородных этапов развития рядов динамики носит название **периодизации динамики.** Вопрос о том, какие этапы развития прошло то или иное явление за определенный исторический отрезок времени, решается теорией той науки, к области которой относится изучаемая совокупность явлений.

Важно также, чтобы в ряду динамики **интервалы или моменты**, по которым определены уровни, имели **одинаковый экономический смысл**. Скажем, при изучении роста поголовья скота бессмысленно сравнивать цифры поголовья по состоянию на 1 октября с данными 1 января, так как первая цифра включает не только скот, оставшийся на зимовку, но и предназначенный к убою, а вторая цифра включает только скот, оставленный на зимовку.

Уровни ряда динамики могут оказаться несопоставимыми **по кругу охватываемых объектов** вследствие перехода ряда объектов из одного подчинения в другое.

Несопоставимость уровней ряда может возникнуть вследствие изменений **территориальных границ** областей, районов и так далее.

Следовательно, прежде чем анализировать динамический ряд, надо, исходя из цели исследования, убедиться в сопоставимости уровней ряда и, если последняя отсутствует, добиться ее дополнительными расчетами. Для того, чтобы привести уровни ряда динамики к сопоставимому виду, иногда приходится прибегать к приему, который носит название **смыкание рядов динамики**. Под смыканием понимают объединение в один ряд (более длинный) двух или нескольких рядов динамики, уровни которых являются несопоставимыми. Для осуществления смыкания необходимо, чтобы для одного из периодов (переходного) имелись данные, исчисленные по разной методологии (или в разных границах).

Пример. Предположим, что в N-ом регионе имеются данные об общем объеме оборота розничной торговли за 1999-2001гг. в фактически действующих ценах, а за 2001-2004гг. – в сопоставимых ценах (табл. 9.4.).

Таблица 9.4.

Динамика общего объема оборота розничной торговли

Годы	1999	2000	2001	2002	2003	2004
Оборот розничной торговли, млрд. руб. (в фактически действующих ценах)	19,7	20	21,2	-	-	-
Оборот розничной торговли, млрд. руб. (в сопоставимых ценах)	-	-	22,8	24,6	25,2	26,1
Сомкнутый ряд абсолютных величин (в сопоставимых ценах; млрд. руб.)	21,3	21,5	22,8	24,6	25,2	26,1
Сопоставимый ряд относительных величин (в % к 2001 г.)	92,9	94,3	100	107,9	110,5	114,5

Решение. Чтобы проанализировать динамику общего объема розничной торговли за 1999-2004 гг., необходимо сомкнуть (объединить) приведенные выше два ряда в один. А чтобы уровни нового ряда были сопоставимы, необходимо пересчитать данные 1999-2001 гг. в сопоставимые цены. Для этого на основе данных об объеме розничной торговли за 2001 г. в фактических и сопоставимых ценах находим соотношение между ними: 22,8:21,2 = 1,08. Умножая на полученный коэффициент данные за 1999-2001 гг., приводим их, таким образом, к сопоставимому виду с последующими уровнями. Сомкнутый (сопоставимый) ряд динамики показан в предпоследней строке таблицы 9.4.

Другой способ смыкания рядов заключается в том, что уровни года, в котором произошли изменения (в нашем примере – уровни 2001 г.), как до изменений, так и после изменений (для нашего примера – в фактических и сопоставимых ценах, т.е. 21,2 и 22,8) принимаются за 100%, а остальные пересчитываются в процентах по отношению к этим уровням соответственно (в нашем примере в фактических ценах – по отношению к 21,2, в сопоставимых ценах – к 22,8). В результате получаем сомкнутый ряд динамики, который показан в последней строке таблицы 9.4.

Та же проблема приведения к сопоставимому виду возникает и при параллельном анализе развития во времени экономических показателей отдельных стран, административных и территориальных районов. Это, во-первых, вопрос о сопоставимости цен сравниваемых стран, во-вторых, вопрос о сопоставимости методики расчета сравниваемых показателей. В таких случаях ряды динамики **приводятся к одному основанию**, то есть к одному и тому же периоду или моменту времени, уровень которого принимается за базу сравнения, а все остальные уровни выражаются в виде коэффициентов или в процентах по отношению к нему.

9.3. Аналитические показатели ряда динамики

На практике для количественной оценки динамики явлений широко применяется ряд основных аналитических показателей. К таким показателям относятся: абсолютный прирост, темп роста и прироста, абсолютное значение одного процента прироста. При этом принято сравниваемый уровень называть отчетным, а уровень, с которым происходит сравнение – базисным.

Абсолютный прирост (Δ) характеризует размер увеличения (или уменьшения) уровня ряда за определенный промежуток времени. Он равен разности двух сравниваемых уровней и выражает абсолютную скорость роста. В общем случае абсолютный прирост может быть представлен в виде:

$$\Delta i = yi - yi\text{-}k, \qquad (9.1)$$

где yi – текущий уровень ряда динамики; i = 2,3,…,n; k = 1,2,…,n-1.

При k = 1 от текущего уровня yi вычитается предыдущий уровень yi-1, и получается формула для расчета цепного абсолютного прироста:

$$\Delta ц = yi - yi\text{-}1 \qquad (9.2)$$

При k = i-1 из формулы (9.1) вытекает выражение для базисного абсолютного прироста, определяемого относительно начального уровня ряда:

$$\Delta б = yi - y1 \qquad (9.3)$$

Для записи формулы базисного абсолютного прироста в более общем виде уровень y1 в формуле (9.3) может быть заменен на уровень ряда динамики, принятый за базу сравнения – y0:

$$\Delta б = yi - y0 \qquad (9.4)$$

Показатель интенсивности изменения уровня ряда – в зависимости от того, выражается ли он в виде коэффициента или в процентах, принято называть **коэффициентом роста или темпом роста**. Иными словами, коэффициент роста и темп роста представляют собой две формы выражения интенсивности изменения уровня. Разница между ними заключается только в единице измерения.

Коэффициент роста показывает, во сколько раз данный уровень ряда больше базисного уровня (если этот коэффициент больше единицы) или какую часть базисного уровня составляет уровень текущего периода за некоторый промежуток времени (если он меньше единицы).

Темпы роста характеризуют отношение двух сравниваемых уровней ряда в виде:

$$Tp = \frac{y_i}{y_{i-k}} \cdot 100\% \qquad (9.5)$$

где yi – текущий уровень ряда динамики; i = 2,3,…,n; k = 1,2,…,n-1.

Отметим, что индекс уровня yi-k, находящийся в знаменателе, определяется так же, как и в случае абсолютного прироста. Следовательно, из выражения формулы (9.6) в зависимости от значений индекса k получаются формулы для расчета цепных и базисных темпов роста.

Цепной темп роста будет равен:

$$Tp_{ц} = \frac{y_i}{y_{i-1}} \cdot 100\% \qquad (9.6)$$

Базисный темп роста может быть представлен в виде :

$$Tp_{б} = \frac{y_i}{y_1} \cdot 100\% \qquad (9.7)$$

где y1 – уровень ряда динамики, принятый за базу сравнения.

Темп роста всегда число положительное. Если темп роста равен100%, то значение уровня не изменилось, если больше 100%, то значение уровня повысилось, а если меньше 100% - понизилось.

Темп прироста характеризует абсолютный прирост в относительных величинах. Определенный в процентах темп прироста показывает, на сколько процентов изменился сравниваемый уровень по отношению к уровню, принятому за базу сравнения. Темп прироста рассчитывается как отношение абсолютного прироста к уровню, принятому за базу сравнения:

$$Tпр = \frac{y_i - y_{i-k}}{y_{i-k}} \cdot 100\% \qquad (9.8)$$

где yi – текущий уровень ряда динамики; i = 2,3,…,n; k = 1,2,…,n-1.

Если темп роста всегда положительное число, то темп прироста может быть положительным, отрицательным и равным нулю.

При k = 1 получаем цепной темп прироста:

$$Тпр_ц = \frac{y_i - y_{i-1}}{y_{i-1}} \cdot 100\% \qquad (9.9)$$

Преобразовав выражение формулы (9.9), можно показать зависимость цепного темпа прироста от соответствующего темпа роста:

$$Тпр_ц = \frac{y_i}{y_{i-1}} \cdot 100\% - 100\% = Тр_ц - 100\% \qquad (9.10)$$

где Трц – цепной темп роста.

Базисный темп прироста равен отношению базисного абсолютного прироста к уровню ряда, принятому за базу сравнения:

$$Тпр_б = \frac{y_i - y_1}{y_1} \cdot 100\% \qquad (9.11)$$

По аналогии с формулой (9.11) получаем:

$$Тпр_б = \frac{y_i}{y_1} \cdot 100\% - 100\% = Тр_б - 100\% \qquad (9.12)$$

где Трб – базисный темп роста.

Сравнение абсолютного прироста и темпа прироста за одни и те же периоды времени показывает, что в реальных экономических процессах замедление темпов прироста не всегда сопровождается уменьшением абсолютных приростов. Поэтому на практике часто проводят сопоставление этих показателей. Для этого рассчитывают **абсолютное значение одного процента прироста**. Оно представляет собой одну сотую часть базисного уровня и в то же время – отношение абсолютного прироста к соответствующему темпу прироста:

$$|\%| = \frac{\Delta_ц}{Тпр_ц\ y\%} = \frac{y_i - y_{i-1}}{\frac{y_i - y_{i-1}}{y_{i-1}} \cdot 100} = \frac{y_{i-1}}{100} = 0{,}01 \cdot y_{i-1} \qquad (9.13)$$

Таким образом, базисные показатели динамики характеризуют окончательный результат всех изменений в уровнях ряда от периода, к которому относится базисный уровень, до данного (i-го) периода. Цепные показатели динамики характеризуют интенсивность изменения уровня от периода к периоду (или от даты к дате) в пределах изучаемого промежутка времени (рис. 9.1).

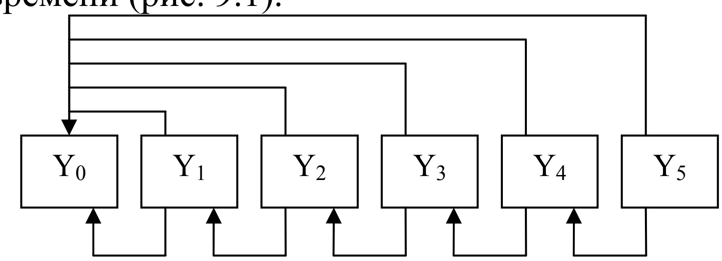

Рис. 9.1. Построение цепных и базисных аналитических показателей динамики

Пример. По данным о числе проданных квартир в N-ом регионе рассчитаем аналитические показатели ряда динамики (табл.9.5).

Таблица 9.5.

Динамика числа проданных квартир в N-ом регионе за 2000-2004 гг.

Годы	Число проданных кваритир, тыс.ед	Абсолютный прирост, тыс. ед.		Темп роста, %		Темп прироста, %		Абсолютное значение одного процента прироста, тыс.ед.
		по сравнению с преды-дущим годом	по сравне-нию с 2000 г.	по сравне-нию с преды-дущим годом	по сравне-нию с 2000 г.	по сравне-нию с преды-дущим годом	по сравне-нию с 2000 г.	
А	1	2	3	4	5	6	7	8
2000	108	…	-	…	100,0	…	-	…
2001	107	-1	-1	99,1	99,1	-0,9	-0,9	1,08
2002	110	+3	+2	102,8	101,9	+2,8	+1,9	1,07
2003	111	+1	+3	100,9	102,8	+0,9	+2,8	1,10

Решение.

- Рассчитаем цепные и базисные абсолютные приросты (формулы 9.2 и 9.3):

Цепные: Δ01/00 = 107-108 = -1 тыс.ед.

Δ02/01 = 110-107 = +3 тыс.ед.

и т.д. (см. табл. 9.5 гр.2)

Базисные: Δ01/00 = 107-108 = -1 тыс.ед.

Δ02/00 = 110 – 108 = +2 тыс.ед.

Δ03/00 = 111-108 = +3 тыс.ед.

и т.д. (см. табл. 9.5 гр.3)

- Рассчитаем цепные и базисные темпы роста (формулы 9.6 и 9.7):

Цепные:

$$Tp_{01/00} = \frac{107}{108} \cdot 100 = 99{,}1\%$$

$$Tp_{02/01} = \frac{110}{107} \cdot 100 = 102{,}8\%$$

и т.д. (см. табл. 9.5 гр.4)

Базисные:

$$Tp_{01/00} = \frac{107}{108} \cdot 100 = 99{,}1\%$$

$$Tp_{02/00} = \frac{110}{108} \cdot 100 = 101{,}9\%$$

$$Tp_{03/00} = \frac{111}{108} \cdot 100 = 102{,}8\%$$

и т.д. (см. табл. 9.5 гр.5)

- Рассчитаем цепные и базисные темпы прироста (формулы 9.10 и 9.12):

Цепные: Тпр01/00 = 99,1% - 100% = -0,9%
Тпр02/01 = 102,8% - 100% = +2,8%
и т.д. (см. табл. 9.5 гр.6)

Базисные: Тпр01/00 = 99,1% - 100% = -0,9%
Тпр02/00 = 101,9% - 100% = +1,9%
Тпр03/00 = 102,8% - 100% = +2,8%
и т.д. (см. табл. 9.5 гр.7)

- Рассчитаем абсолютное значение одного процента прироста (формула 9.13):

|%|2001 = 108*0,01 = 1,08 тыс.ед.
|%|2002 = 107*0,01 = 1,07 тыс.ед.
и т.д. (см. табл. 9.5 гр.8)

9.4. Средние показатели в рядах динамики и методы их исчисления

Каждый ряд динамики можно рассматривать как некую совокупность m меняющихся во времени показателей, которые можно обобщать в виде средних величин. Для обобщения данных по рядам динамики рассчитываются: средний уровень ряда; средний абсолютный прирост; средний темп роста и прироста.

Средний уровень ряда динамики ($\bar{y}$) рассчитывается по средней хронологической. **Средней хронологической** называется средняя, исчисленная из значений, изменяющихся во времени. Такие средние обобщают хронологическую вариацию. В хронологической средней отражается совокупность тех условий, в которых развивалось изучаемое явление в данном промежутке времени.

Средний уровень ряда определяется по-разному для моментных и интервальных рядов.

- Для интервальных равноотстоящих рядов средней уровень находится по формуле простой средней арифметической:

$$\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n} \qquad (9.14)$$

где n – число уровней или длина ряда.

- Для интервальных неравноотстоящих рядов средний уровень находится по формуле взвешенной средней арифметической:

$$y = \frac{\sum_{i=1}^{n} y_i t_i}{\sum_{i=1}^{n} t_i} \qquad (9.15)$$

где ti – продолжительность интервалов времени между уровнями (число периодов времени, при которых значение уровня не изменяется).

Пример. В таблице 9.7. приведен интервальный ряд динамики с равноотстоящими уровнями. По этим данным можно рассчитать среднегодовой уровень числа проданных квартир за 2000-2004 гг. Он будет равен 347 тыс.ед. ($\bar{y}$ = 1735/5), то есть в среднем ежегодно число проданных квартир в регионе за 2000-2004 гг. составило полученное значение.

- Средний уровень моментного равноотстоящего ряда динамики находится по формуле средней хронологической простой:

$$\bar{y} = \frac{\frac{y_1 + y_2}{2} + \frac{y_2 + y_3}{2} + \frac{y_3 + y_4}{2} + \ldots + \frac{y_{n-1} + y_n}{2}}{n-1} =$$

$$= \frac{\frac{y_1}{2} + y_2 + y_3 + \ldots + y_{n-1} + \frac{y_n}{2}}{n-1} \text{ или } \bar{y} = \frac{\frac{y_1 + y_n}{2} + \sum_{i=2}^{n-1} y_i}{n-1} \qquad (9.16)$$

- Средний уровень моментных рядов динамики с неравноотстоящими уровнями определяется по формуле средней хронологической взвешенной:

$$\bar{y} = \frac{(y_1 + y_2)t_1 + (y_2 + y_3)t_2 + \ldots + (y_{n-1} + y_n)t_{n-1}}{2(t_1 + t_2 + \ldots + t_{n-1})} =$$

$$= \frac{\sum_{\mathrm{i=1}}^{\mathrm{n-1}} (\mathrm{y_i} + y_{i+1})t_i}{2\sum_{i=1}^{n-1} t_i} \qquad (9.17)$$

где t – продолжительность интервала времени между соседними уровнями.

Пример. Покажем расчет среднего уровня моментного ряда динамики с равноотстоящими уровнями по данным о численности работников фирмы на 1-е число каждого месяца 2004 г. (чел.):

1/I	1/II	1/III	1/IV
347	350	349	351

Среднемесячная численность работников фирмы за 1 квартал (по формуле 9.16) составит: $\bar{y} = \frac{347/2 + 350 + 349 + 351/2}{3} = \frac{1048}{3} \approx 349 чел$

Пример. Известна списочная численность рабочих организаций на некоторые даты 2004 г. (чел.). Ряд динамики имеет не равноотстоящие уровни во времени:

1/I	1/III	1/VI	1/IX	1/I-1995
530	570	520	430	550

Среднегодовая численность работников за 1994 г. (по формуле 9.17) составит:

$$\bar{y} = \frac{(530 + 570)2 + (570 + 520)3 + (520 + 430)3 + (430 + 550)4}{2(2 + 3 + 3 + 4)} = \frac{12240}{12} = 510 чел$$

Обобщающим показателем абсолютной скорости изменения явления во времени является **средний абсолютный прирост** за весь период, ограничивающий ряд динамики. Скоростью в данном случае будем называть прирост (уменьшение) в единицу времени. Для его определения используется формула средней арифметической простой:

$$\bar{\Delta} = \frac{\sum \Delta_ц}{n - 1} \tag{9.18}$$

Подставив в числитель выражение для цепных абсолютных приростов, получим более удобную форму записи для среднего абсолютного прироста:

$$\bar{\Delta} = \frac{y_2 - y_1 + y_3 - y_2 + \ldots + (y_n - y_{n-1})}{n - 1} = \frac{y_n - y_1}{n - 1} \tag{9.19}$$

где yn и y1 - соответственно конечный и начальный уровни ряда динамики.

Пример. По данным таблицы 9.7 определим средний абсолютный прирост числа проданных квартир за период 2000-2004 гг. Он будет равен 1,0 тыс.ед. [(112-108) : 4].

Сводной обобщающей характеристикой интенсивности изменения уровней ряда динамики служит **средний темп роста**. Он показывает, сколько в среднем процентов последующий уровень составляет от предыдущего в течение всего периода наблюдения.

Средний темп (коэффициент) роста рассчитывается по формуле средней геометрической из цепных коэффициентов роста:

$$\overline{Tp} = \sqrt[n-1]{K_{2/1}K_{3/2}...K_{n/n-1}} = \sqrt[n-1]{\Pi K_{ц}} \tag{9.20}$$

Выразив цепные коэффициенты (темпы) роста через соответствующие уровни ряда, получим:

$$\overline{Tp} = \sqrt[n-1]{\frac{y_2}{y_1}\cdot\frac{y_3}{y_2}\cdot\frac{y_4}{y_3}\cdots\frac{y_n}{y_{n-1}}}\cdot 100\% = \sqrt[n-1]{\frac{y_n}{y_1}}\cdot 100\% \tag{9.21}$$

Пример. По данным таблицы 9.7 рассчитаем средний темп роста числа проданных квартир за период 2000-2004 гг. по формуле 9.21:

$$\overline{Tp} = \sqrt[4]{0{,}991\cdot 1{,}028\cdot 1{,}009\cdot 1{,}009} = \sqrt[4]{1{,}037} = 1{,}009 \text{ или } 100{,}9\%$$

или по формуле $\overline{Tp} = \sqrt[4]{\frac{112}{108}} = \sqrt[4]{1{,}037} = 1{,}009$ или 100,9%

Когда приходится производить расчет средних темпов роста по периодам различной продолжительности (не равноотстоящие уровни), то используют среднюю геометрическую, взвешенную по продолжительности периодов. Формула средней геометрической взвешенной будет иметь вид:

$$\overline{Tp} = \sqrt[\sum t]{(K_{2/1})^{t_1}\ (K_{3/2})^{t_2}.....(K_{n/n-1})^{t_{n-1}}} \tag{9.22}$$

где t – интервал времени, в течение которого сохраняется данный темп роста.

Средний темп прироста не может быть определен непосредственно на основании последовательных темпов прироста или показателей среднего абсолютного прироста. Для его вычисления необходимо сначала найти средний темп роста, а затем его уменьшить на единицу или на 100%:

$$\overline{T}пр = \overline{T}p - 100\% \tag{9.23}$$

Пример. По данным таблицы 9.7 был рассчитан средний темп роста числа проданных квартир за 2000-2004 гг. равный 100,9%, отсюда средний темп прироста будет равен:

$$\overline{T}p = 100{,}9\% - 100\% = 0{,}9\%.$$

9.5. Методы анализа основной тенденции (тренда) в рядах динамики

Важной задачей статистики при анализе рядов динамики является определение основной тенденции развития, присущей тому или иному ряду динамики.

Под основной тенденцией развития ряда динамики понимают изменение, определяющее общее направление развития. Это - систематическая составляющая долговременного действия. В некоторых случаях общая тенденция ясно прослеживается в динамике рассматриваемого показателя, в других случаях она может не просматриваться из-за ощутимых случайных колебаний. Например, в отдельные моменты времени сильные колебания розничных цен могут заслонить наличие тенденции к росту или снижению этого показателя. Поэтому для выявления основной тенденции развития в статистике применяются 2 группы методов:

- сглаживание или механическое выравнивание отдельных уровней ряда динамики с использованием фактических значений соседних уровней;
- выравнивание с применением кривой, проведенной между конкретными уровнями таким образом, чтобы она отражала тенденцию, присущую ряду и одновременно освободила его от незначительных колебаний.

Рассмотрим методы каждой группы.

Метод укрупнения интервалов основан на укрупнении периодов времени, к которым относятся уровни. Например, ряд недельных данных можно преобразовать в ряд помесячной динамики, ряд квартальных данных заменить годовыми уровнями. Уровни нового ряда могут быть получены путем суммирования уровней исходного ряда, либо могут представлять средние уровни.

Распространенным приемом при выявлении тенденции развития является **сглаживание ряда динамики**. Суть различных приемов сглаживания сводится к замене фактических уровней ряда расчетными уровнями, которые в меньшей степени подвержены колебаниям. Это способствует более четкому проявлению тенденции развития.

Метод простой скользящей средней. Сглаживание ряда динамики с помощью скользящей средней заключается в том, что вычисляется средний уровень из определенного числа первых по порядку уровней ряда, затем средний уровень из такого же числа уровней, начиная со второго, далее – начиная с третьего и т.д. Таким образом, при расчете средних уровней они как бы «скользят» по ряду динамики от его начала к концу, каждый раз отбрасывая один уровень вначале и добавляя один следующий. Отсюда название – **скользящая средняя**.

Каждое звено скользящей средней – это средний уровень за соответствующий период, который относится к **середине выбранного периода**, если число уровней ряда динамики нечетное.

Нахождение скользящей средней по четному числу членов рядов динамики несколько сложнее, так как средняя может быть отнесена только к середине между двумя датами, находящимся в середине интервала сглаживания. Например, средняя, найденная для четырех уровней, относится к середине между вторым и третьим, третьим и четвертым уровнями и так далее. Чтобы ликвидировать такой сдвиг, применяют так называемый **способ центрирования. Центрирование** заключается в нахождении средней из двух смежных скользящих средних для отнесения полученного уровня к определенной дате. При центрировании необходимо находить скользящие суммы, скользящие средние нецентрированные по этим суммам и средние из двух смежных нецентрированных скользящих средних.

Пример. Покажем расчет скользящей средней за 3 и 4 месяца по данным, представленным в таблице 9.6.

Таблица 9.6.

Динамика продажи магнитофонов в торговой сети за 2004 год

Месяц	Продано магнитофонов, тыс.шт.	Трехуровневые скользящие суммы	Трехуровневые скользящие средние	Четырехуровневые скользящие суммы	Четырехуровневые скользящие средние нецентрированные	Четырехуровневые скользящие средние центрированные
А	1	2	3	4	5	6
январь	23	-	-	-		-
февраль	25	-	23	-		-
март	21	69	24	-		24,4
апрель	26	72	25	95		24,9
май	28	75	26	100		25,8
июнь	24	78	27	99		27,0
июль	29	81	27	107		27,5
август	28	81	29	109		28,4
сентябрь	30	87	29	111		29,3
октябрь	29	87	30	116		30,1
ноябрь	31	90	31	118		-
декабрь	33	93	-	123		-

Недостаток метода простой скользящей средней состоит в том, что сглаженный ряд динамики сокращается ввиду невозможности получить сглаженные уровни для начала и конца ряда. Этот недостаток устраняется применением метода аналитического выравнивания для анализа основной тенденции.

Аналитическое выравнивание предполагает представление уровней данного ряда динамики в виде функции времени – y = f(t).

При таком подходе изменение явления связывают лишь с течением времени, считается, что влияние других факторов несущественно или косвенно сказывается через фактор времени. Правильно построенная модель должна соответствовать характеру изменения тенденции исследуемого явления. Выбранная функция позволяет получить выровненные или теоретические значения уровней ряда динамики.

Для отображения основной тенденции развития явлений во времени применяются различные функции: полиномы разной степени, экспоненты, логистические кривые и другие виды.

Полиномы имеют следующий вид:

полином первой степени $\overline{y}_t = a_0 + a_1 t$

полином второй степени $\overline{y}_t = a_0 + a_1 t + a_2 t^2$ (9.24)

полином третьей степени $\overline{y}_t = a_0 + a_1 t + a_2 t^2 + a_3 t^3$

полином n-ой степени $\overline{y}_t = a_0 + a_1 t + a_2 t^2 + ... + a_n t^n$

Здесь a0; a1; a2; … an – параметры полиномов, t – условное обозначение времени. В статистической практике параметры полиномов невысокой степени иногда имеют конкретную интерпретацию характеристик динамического ряда. Так, например, параметр a0 характеризует средние условия развития ряда динамики, параметр a1 – скорость роста, параметр a2 – ускорение роста, параметр a3 – изменение ускорения.

Оценка параметров в моделях (9.24) находится методом наименьших квадратов. Как известно, суть его состоит в определении таких параметров (коэффициентов), при которых сумма квадратов отклонений расчетных значений уровней от фактических значений была бы минимальной. Таким образом, эти оценки находятся в результате минимизации выражения:

$$\sum_{t=1}^{n}\left(y_t - \overline{y_t}\right)^2 \rightarrow \min \qquad (9.25)$$

где yt - фактическое значение уровня ряда динамики; $\overline{y_t}$ - расчетное значение; n – длина ряда динамики.

В результате минимизации выражения (9.25) получается система нормальных уравнений:

$$\begin{cases} na_0 + a_1\sum t + a_2\sum t^2 + ... + a_p\sum t^p = \sum y \\ a_0\sum t + a_1\sum t^2 + a_2\sum t^3 + ... + a_p\sum t^{p+1} = \sum yt \\ \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \\ a_0\sum t^p + a_1\sum t^{p+1} + a_2\sum t^{p+2} + ... + a_p\sum t^{2p} = \sum yt^p \end{cases} \qquad (9.26)$$

где n - число членов в ряду динамики, t=1,2,...,n

Система 9.26, состоящая из “p” уравнений, содержит в качестве известных величин $\sum y,\ \sum yt, \ldots, \sum yt^p$, то есть суммы наблюдаемых значений уровней динамического ряда, умноженные на показатели времени в степени 1,2,...,p и неизвестных величин aj. Решение этой системы относительно a0, a1,...,ap и дает искомые значения параметров.

Системы для расчета параметров полиномов невысоких степеней намного проще. Обозначим последовательные параметры полиномов как a0, a1, a2. Тогда системы нормальных уравнений для оценивания параметров прямой $\bar{y}_t = a_0 + a_1 t$ примет вид:

$$\begin{cases} na_0 + a_1 \sum t = \sum y \\ a_0 \sum t + a_1 \sum t^2 = \sum yt \end{cases} \qquad (9.27)$$

для параболы второго порядка (yt=a0+a1t+a2t2):

$$\begin{cases} na_0 + a_1 \sum t + a_2 \sum t^2 = \sum y \\ a_0 \sum t + a_1 \sum t^2 + a_2 \sum t^3 = \sum yt \\ a_0 \sum t^2 + a_1 \sum t^3 + a_2 \sum t^4 = \sum yt^2 \end{cases} \qquad (9.28)$$

Составление нормальных уравнений можно упростить, воспользовавшись тем, что величины Σt, Σt2 и т.д. не зависят от конкретных уровней ряда. Эти суммы являются функциями только числа членов в динамическом ряду. Для них получены следующие формулы:

$\sum t = \frac{n(n+1)}{2}$; $\sum t^2 = \frac{n(n+1)(2n+1)}{6}$; $\sum t^3 = \frac{n^2(n+1)^2}{4}$;

$\sum t^4 = \frac{n(n+1)(2n+1)(3n^2+3n-1)}{30}$

(суммирование по t = 1+n).

Другой подход к упрощению расчетов заключается в переносе начала координат в середину ряда динамики. В этом случае упрощаются сами нормальные уравнения, а так же уменьшаются абсолютные значения величин, участвующих в расчете. Если до переноса начала координат t было равно 1,2,3,...,n, то после переноса:

- для нечетного числа уровней ряда t = …; -3; -2; -1; 0; 1; 2; 3; …
- для четного числа уровней ряда t = …; -5; -3; -1; 1; 3; 5; …

Следовательно, Σt и все Σtp , у которых “p” – нечетное число, равны 0. Таким образом, все члены уравнений, содержащие Σt с такими степенями, могут быть исключены. Системы нормальных уравнений теперь упрощаются для прямой:

$$\begin{cases} a_0 n = \sum y \\ a_1 \sum t^2 = \sum ty \end{cases} \quad (9.29)$$

для параболы второго порядка:

$$\begin{cases} a_0 n + a_2 \sum t^2 = \sum y \\ a_1 \sum t^2 = \sum ty \\ a_0 \sum t^2 + a_2 \sum t^4 = \sum t^2 y \end{cases} \quad (9.30)$$

Решая системы (9.29) и (9.30), получим величины параметров соответствующих полиномов.

При сглаживании ряда динамики по показательной кривой (yt=a0a1t) для определения параметров применяется также метод наименьших квадратов, но только к логарифмам исходных данных. Так, для нахождения параметров показательной функции необходимо решить следующую систему уравнений:

$$\begin{cases} \sum \lg y = n \lg a_0 + \lg a_1 \sum t \\ \sum t \lg y = n \lg a_0 \sum t + \lg a_1 \sum t^2 \end{cases} \quad (9.31)$$

Если $\sum t = 0$, то параметры уравнения lg a0 и lg a1 находим по формулам:

$$\lg a_0 = \frac{\sum \lg y}{n}; \quad \lg a_1 = \frac{\sum t \lg y}{\sum t^2}.$$

Пример. Необходимо определить основную тенденцию ряда динамики числа проданных квартир в N-ом регионе за 2000-2004 гг.

Таблица 9.7.

Таблица исходных и расчетных данных

Годы	Число проданных квартир, тыс.ед.	t	t2	yt	$\bar{y}_t$
А	1	2	3	4	5
2000	108	-2	4	-216	107,2
2001	107	-1	1	-107	108,4
2002	110	0	0	0	109,6
2003	111	+1	1	+111	110,8
2004	112	+2	4	+224	112,0
Итого	548	0	10	+12	548,0

Первые две графы – ряд динамики, подвергаемый выравниванию, дополняются графой 2, в которой показана система отсчета времени “t”. Причем эта система выбирается таким образом, чтобы Σt = 0. В качестве функции выравнивания выбрано уравнение прямой линии: $\bar{y}_t = a_0 + a_1 t$, параметры данного уравнения находим по упрощенным формулам:

$$\left.\begin{aligned} a_0 &= \frac{\sum y}{n} \\ a_1 &= \frac{\sum ty}{\sum t^2} \end{aligned}\right\}$$

Затем в графах 3 и 4 проводим необходимые расчеты и находим: Σy = 548; Σyt = 12; Σt2 = 10. Отсюда $a_0 = \frac{548}{5} = 109{,}6$; $a_1 = \frac{12}{10} = 1{,}2$.

Уравнение прямой будет иметь вид: $\bar{y}_t$ = 109,6 + 1,2 t.

На основе этого уравнения находятся выровненные годовые уровни путем подстановки в него соответствующих значений “t” (графа 5 таблицы 9.7).

Полученное уравнение показывает, что численность проданных квартир в регионе растет в среднем на 1,2 тысяч единиц в год. Таким образом, величина параметра a1 в уравнении прямой показывает среднюю величину абсолютного прироста выровненного ряда динамики.

Сумма уровней эмпирического ряда ($\sum y_i$) полностью совпала с суммой расчетных значений выровненного ряда ($\sum \bar{y}_e$).

Результаты произведенного аналитического выравнивания ряда динамики проданных квартир за 2000-2004 гг. и фактические данные отражены на рисунке 9.2

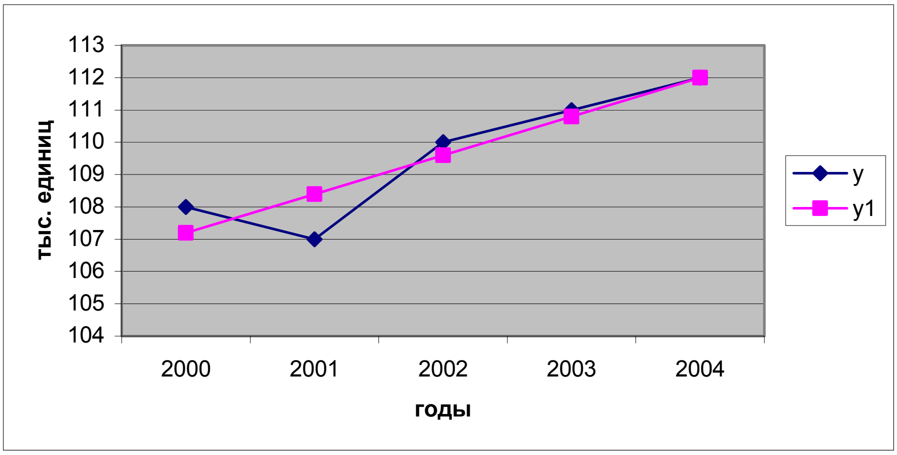

Рис. 9.2. Динамика численности проданных квартир в N-ом регионе за 2000-2004 гг.

9.6. Методы выявления сезонной компоненты

При рассмотрении квартальных или месячных данных многих социально-экономических явлений часто обнаруживаются определенные, постоянно повторяющиеся колебания, которые существенно не изменяются за длительный период времени. Они являются результатом влияния природно-климатических условий, общих экономических факторов, а также ряда многочисленных разнообразных факторов, которые частично являются регулируемыми. В статистике периодические колебания, которые имеют определенный и постоянный период, равный годовому промежутку, носят название «сезонных колебаний» или «сезонных волн», а динамический ряд в этом случае называют **тренд-сезонным**, или просто **сезонным рядом динамики**.

Сезонные колебания характеризуются специальным показателями, которые называются **индексами сезонности** (Is). Совокупность этих показателей отражает сезонную волну. Индексами сезонности являются процентные отношения фактических внутригодовых уровней к постоянной или переменной средней.

Для выявления сезонных колебаний обычно берут данные за несколько лет, распределенные по месяцам или кварталам. Данные за несколько лет (обычно не менее трех) берутся для того, чтобы выявить устойчивую сезонную волну, на которой не отражались бы случайные условия одного года.

Если ряд динамики не содержит ярко выраженной тенденции в развитии, то индексы сезонности вычисляются непосредственно по фактическим данным без их предварительного выравнивания.

Для каждого месяца определяется средняя величина уровня, например, за три года ($\bar{y}_i$), затем из них рассчитывается среднемесячный уровень для всего ряда ($\bar{y}$) и в заключение определяется процентное отношение средних для каждого месяца к общему среднемесячному уровню ряда, то есть:

$$I_S = \frac{\bar{y}_i}{y}100\% \tag{9.32}$$

Пример. За 2002-2004гг. по месяцам имеются данные о числе зарегистрированных браков населением N-го города. Рассчитать индексы сезонности методом постоянной средней (табл. 9.8).

Рассчитанные индексы сезонности характеризуют сезонную волну числа зарегистрированных браков населения во внутригодовой динамике, где пик регистрации приходится на январь месяц.

Таблица 9.8.

Динамика зарегистрированных браков населения N-го города за 2002-2004 гг.

Месяцы	Зарегистрировано браков, шт.				Индекс сезонности %
	2002	2003	2004	В среднем за три года	
январь	190	155	145	163,3	120,7
февраль	165	140	135	146,7	108,4
март	150	153	135	146,0	107,9
апрель	135	140	146	140,3	103,0
май	135	136	131	134,0	99,0
июнь	123	130	136	129,7	95,9
июль	125	128	125	126,0	93,1
август	120	125	124	123,0	90,9
сентябрь	118	118	120	118,7	87,7
октябрь	126	130	128	128,0	94,6
ноябрь	130	131	135	132,0	97,6
декабрь	138	131	139	136,0	100,5
Средний уровень ряда	137,9	134,8	133,3	135,3	100,0

Если же ряд динамики содержит определенную тенденцию в развитии, то прежде чем вычислить сезонную волну, фактические данные должны быть обработаны так, чтобы была выявлена общая тенденция. Обычно для этого прибегают к аналитическому выравниванию ряда динамики.

При использовании способа аналитического выравнивания ход вычислений индексов сезонности следующий:

- по соответствующему полиному вычисляются для каждого месяца (квартала) выровненные уровни на момент времени (t);

- вычисляются отношения фактических месячных (квартальных) данных (yi) к соответствующим выровненным данным ($\bar{y}$t) в процентах $\left[I_s = (y_i : \bar{y}_t) \times 100\right]$;

- находятся средние арифметические из процентных отношений, рассчитанных по одноименным периодам Ii=(I1+I2+I3+...+In):n, где n - число одноименных периодов.

В общем виде формулу расчета индекса сезонности данным способом можно записать так:

$$I_S = \left[\sum \frac{y_i}{y_t} \right] : n \times 100\% \qquad (9.33)$$

Расчет заканчивается проверкой правильности вычислений индексов, так как средний индекс сезонности для всех месяцев (кварталов) должен быть 100 процентов, то сумма полученных индексов по месячным данным равна 1200, а сумма по четырем кварталам - 400.

9.7. Элементы прогнозирования и интерполяции

Анализ динамики социально-экономических явлений, выявление и характеристика основной тенденции развития дают основание для прогнозирования - определения будущих размеров уровня экономического явления.

Процесс прогнозирования предполагает, что закономерность развития, действующая в прошлом (внутри ряда динамики), сохранится и в прогнозируемом будущем, то есть прогноз основан на **экстраполяции**. Экстраполяция, проводимая в будущее, называется **перспективой,** и в прошлое - **ретроспективой**. Обычно, говоря об экстраполяции рядов динамики, подразумевают чаще всего перспективную экстраполяцию. Первоначальные прогнозы, как правило, сводятся к экстраполяции тенденции. При этом могут использоваться разные методы, в зависимости от исходной информации. Можно выделить следующие элементарные методы экстраполяции: на основе среднего абсолютного прироста, среднего темпа роста и экстраполяция на основе применения метода наименьших квадратов и представления развития явлений во времени в виде уравнения тренда, т.е. математической функции уровней ряда (y) от фактора времени (t).

Прогнозирование по **среднему абсолютному приросту** может быть выполнено в том случае, если есть уверенность считать общую тенденцию линейной, то есть метод основан на предположении о равномерном изменении уровня (под равномерностью понимается стабильность абсолютных приростов).

В этом случае, чтобы получить прогноз на “i” шагов вперед (i – период упреждения), достаточно воспользоваться следующей формулой:

$$\hat{y}_{n+1} = y_n + i \cdot \overline{\Delta} \qquad (9.34)$$

где yn – фактическое значение в последней n-ой точке ряда (конечный уровень ряда); $\hat{y}_{n+1}$ - прогнозная оценка значения (n+1) уровня ряда; $\overline{\Delta}$ - значение среднего абсолютного прироста, рассчитанное для ряда динамики y1; y2; y3; …; yn.

Прогнозирование **по среднему темпу роста** можно осуществлять в случае, когда есть основание считать, что общая тенденция ряда характеризуется показательной (экспоненциальной) кривой. Для нахождения прогнозного значения на “i” шагов вперед необходимо использовать следующую формулу:

$$\hat{y}_{n+1} = y_n \cdot \overline{K_p^i} \qquad (9.35)$$

где $\overline{K_p}$ - средний коэффициент роста, рассчитанный для ряда y1; y2; y3; …; yn.

К недостаткам рассмотренных методов следует отнести то, что они учитывают лишь конечный и начальный уровень ряда, исключая влияние промежуточных уровней. Тем не менее, методы среднего абсолютного прироста и среднего темпа роста имеют весьма широкую область применения, что объясняется простотой их вычисления. Они могут быть использованы как приближенные, простейшие способы прогнозирования, предшествующие более глубокому количественно-качественному анализу.

Наиболее распространенным методом прогнозирования является **аналитическое выражение тренда**. При этом для выхода за границы исследуемого периода достаточно продолжить значения независимой переменной времени (t).

При таком подходе к прогнозированию предполагается, что размер уровня, характеризирующего явление, формируется под воздействием множества факторов, причем не представляется возможным выделить отдельно их влияние. В связи с этим ход развития связывается не с какими-либо конкретными факторами, а с течением времени. На практике для описания тенденции развития явления широко используются модели кривых роста, представляющие собой различные функции времени y = f(t).

Процедура разработки прогноза с использованием кривых роста включает в себя следующие этапы: 1) выбор одной или нескольких кривых, форма которых соответствует характеру изменения ряда динамики; 2) оценка параметров выбранных кривых; 3) проверка адекватности выбранных кривых прогнозируемому процессу и окончательный выбор кривой роста; 4) расчет точечного и интервального прогнозов.

Остановимся на величине доверительного интервала прогноза, который определяется по формуле:

$$\overline{y}_{t+1} \pm t_\alpha \sigma_t \qquad (9.36)$$

где σ - средняя квадратическая ошибка тренда;

$\overline{y}_{t+1}$ - расчетное значение уровня;

$t\alpha$ - доверительная величина, определяемая на основе t-критерия Стьюдента.

Вместо $t\alpha$ - критерия удобно использовать коэффициент (К*).

Например, необходимо провести прогноз на 2005-2006гг. по данным таблицы (9.5) количества проданных квартир в N-ом регионе.

Для экстраполяции используем уравнение тренда, полученное по прямой: $\overline{y}_t = 39{,}7 + 0{,}25t$. Подставив соответствующее значение t в наше уравнение, получим точечные прогнозы на 2005-2006гг. (графа 2 таблицы 9.9). Для построения интервальных прогнозов рассчитаем среднеквадратическую ошибку тренда (σt=0,56) и используем значения K[1]).

Результаты прогноза представлены в таблице (9.9):

Таблица 9.9.

Прогнозные значения численности проданных квартир в N-ом регионе на 2005-2006гг.

Годы	t	$\hat{y}_{n+1}$	K	σtK*	$\hat{y}_{n+1} \pm \sigma \cdot K*$
A	1	2	3	4	5
2005	3	113,2	2,374	1,33	111,9-114,5
2006	4	114,4	2,741	1,53	112,9-115,9

При анализе рядов динамики иногда приходится прибегать к определению некоторых неизвестных уровней внутри данного ряда динамики, то есть к **интерполяции**.

Как и экстраполяция, интерполяция может производится на основе среднего абсолютного прироста, среднего темпа роста, а также с помощью аналитического выравнивания. При интерполяции предполагается, что ни выявленная тенденция, ни ее характер не претерпели существенных изменений в том промежутке времени, уровень (уровни) которого нам неизвестны.

[1] Значения K* взяты из книги Е.М. Четыркина «Статистические методы прогнозирования». М., 1975 г., стр. 183

Глава 10. Статистический анализ структуры

10.1. Понятие структуры и основные направления ее исследования

Изучаемые статистикой процессы и явления в сфере промышленного или сельскохозяйственного производства, финансов, коммерции, демографии, в социальной и политической областях, как правило, характеризуются внутренней структурой, которая с течением времени может изменяться. Динамика структуры вызывает изменение внутреннего содержания исследуемых объектов и их экономической интерпретации, приводит к изменению установившихся причинно-следственных связей. Именно поэтому изучение структуры и структурных сдвигов занимает важное место в экономико-статистическом анализе.

В статистике под **структурой** понимают совокупность элементов социально-экономических явлений, обладающих определенной устойчивостью внутригрупповых связей при сохранении основных свойств, характеризующих эту совокупность как целое. В качестве примеров можно привести структуру населения региона по возрасту или уровню доходов, структуру предприятий отрасли по численности промышленно-производственного персонала или стоимости основных фондов и другие.

Классификация структур прежде всего предполагает их разделение на два основных вида по временному фактору. **Моментные** структуры характеризуют строение социально-экономических явлений по состоянию на определенные моменты времени и отображаются посредством моментных относительных показателей , как правило, на начало или на конец периода (например, структура парка транспортных средств). **Интервальные** структуры характеризуют строение социально-экономических явлений за определенные периоды времени - дни, недели, месяцы, кварталы, годы (например, структура экспорта и импорта).

Статистика имеет дело как с фактическими, реально существующими структурами, так и со структурами перспективными, прогнозными, оптимальными и стандартизованными. Последние представляют собой какие-либо условные или фактические структуры, принятые в качестве эталонных для расчета и сравнения стандартизованных показателей. Например, для сравнения уровней рождаемости, смертности, заболеваемости и т.п. по двум или более регионам рассчитывают стандартизованные коэффициенты на основе некоторой стандартизованной структуры, в качестве которой может использоваться возрастная структура населения в целом по стране.

Основные направления статистического изучения структуры включают:

а) характеристику структурных сдвигов отдельных частей совокупности за два и более периодов;

б) обобщающую характеристику структурных сдвигов в целом по совокупности;

в) оценку степени концентрации и централизации.

Рассмотрим последовательно эти три направления исследования.

10.1.1. Частные показатели структурных сдвигов

Анализ структуры и ее изменений базируется на относительных показателях структуры - долях или удельных весах, представляющих собой соотношения размеров частей и целого. При этом как частные, так и обобщающие показатели структурных сдвигов могут отражать либо «абсолютное» изменение структуры в процентных пунктах или долях единицы (кавычки показывают, что данные показатели являются абсолютными по методологии расчета, но не по единицам измерения), либо ее относительное изменение в процентах или коэффициентах.

«Абсолютный» прирост удельного веса i-ой части совокупности показывает, на сколько процентных пунктов возросла или уменьшилась данная структурная часть в j-ый период по сравнению с (j-1) периодом:[1]

$$Д_{d_i} = d_{i\,j} - d_{i\,j-1}, \tag{10.1}$$

где d_{ij} - удельный вес (доля) i-ой части совокупности в j-ый период;

d_{ij-1} - удельный вес (доля) i-ой части совокупности в (j-1)-ый период.

Знак прироста показывает направление изменения удельного веса данной структуры части («+» - увеличение, «-» - уменьшение), а его значение - конкретную величину этого изменения.

Темп роста удельного веса представляет собой отношение удельного веса i-ой части в j-ый период времени к удельному весу той же части в предшествующий период:

$$Тр_{d_i} = \frac{d_{i\,j}}{d_{i\,j-1}} \cdot 100 \tag{10.2}$$

Темпы роста удельного веса выражаются в процентах и всегда являются положительными величинами. Однако, если в совокупности

[1] Здесь и далее при исследовании моментных структур под периодами будут подразумеваться моменты времени.

имели место какие-либо структурные изменения, часть темпов роста будет больше 100%, а часть - меньше.

Рассчитаем частные показатели структурных сдвигов по данным о величине зарегистрированного уставного капитала действующих в РФ кредитных организаций (табл. 10.1.):

Таблица 10.1.

Группы кредитных организаций по величине уставного капитала (млн.руб.)	Число кредитных организаций		Удельный вес, в % к итогу		Прирост удельного веса, проц. пунктов Δ_{d_i}	Темп роста удельного веса, % Tp_{d_i}
	1.01.00	1.01.03	1.01.00 d_{i0}	1.01.03 d_{i1}		
А	1	2	3	4	5(гр.4-гр.3)	6(гр.4:гр.3)100
до 10	595	294	44,1	22,1	-22,0	50,1
10 -30	313	291	23,2	21,9	-1,3	94,4
30 - 60	253	253	18,8	19,0	0,2	101,1
60 – 150	93	198	6,9	14,9	8,0	215,9
150-300	43	123	3,2	9,3	6,1	290,6
300 и более	52	170	3,8	12,8	9,0	336,8
Итого	1349	1329	100,0	100,0	0	Х

Как следует из данных таблицы 10.1, наиболее существенно в «абсолютном» выражении изменился удельный вес кредитных организаций с уставным капиталом до 10 млн. руб. - снизился на 22 процентных пункта. В относительном выражении наиболее сильно (почти в 3,4 раза) выросла доля кредитных организаций с уставным капиталом свыше 300 млн. руб.

Мы рассмотрели показатели структурных сдвигов за один интервал между двумя периодами. Если же изучаемая структура представлена данными за три и более периодов, появляется необходимость в динамическом осреднении приведенных выше показателей, т.е. в расчете средних показателей структурных сдвигов.

Средний «абсолютный» прирост удельного веса i-ой структурной части показывает, на сколько процентных пунктов в среднем за какой-либо период (день, неделю, месяц, год и т.п.) изменяется данная структурная часть:

$$\overline{Д}_{d_i} = \frac{d_{in} - d_{i1}}{n-1}, \qquad (10.3)$$

где n - число осредняемых периодов.

Сумма средних «абсолютных» приростов удельных весов всех k структурных частей совокупности, также как и сумма их приростов за один временной интервал, должна быть равна нулю.

Средний темп роста удельного веса характеризует среднее относительное изменение удельного веса i-ой структурной части за n периодов, и рассчитывается по формуле средней геометрической:

$$\overline{T}p_{d_i} = \sqrt[n-1]{Tp_{d\,i1} \cdot Tp_{d\,i2} \cdot Tp_{d\,i3} \cdot \ldots \cdot Tp_{d\,in-1}} \qquad (10.4)$$

Подкоренное выражение этой формулы представляет собой последовательное произведение цепных темпов роста удельного веса за все временные интервалы. После проведения несложных алгебраических преобразований данная формула примет следующий вид:

$$\overline{T}p_{d_i} = \sqrt[n-1]{\frac{d_{in}}{d_{i1}}} \cdot 100 \qquad (10.5)$$

Для иллюстрации этих формул воспользуемся приведенным выше примером (таблица 10.1). Рассчитаем средний годовой прирост (в данном случае - снижение) удельного веса кредитных организаций 1-ой группы (число уровней ряда *n* на рассматриваемом интервале равно 4 – 2000, 2001, 2002, 2003гг.):

$$\overline{\Delta}_{d_1} = \frac{22{,}1 - 44{,}1}{4 - 1} = -7{,}3$$ проц. пункта.

Таким образом можно заключить, что удельный вес кредитных организаций с маленьким уставным капиталом ежегодно снижался в среднем на 7,3 процентного пункта.

По последней группе определим средний месячный темп роста удельного веса:

$$\overline{T}p_{d_1} = \sqrt[3]{\frac{12{,}8}{3{,}8}} \cdot 100 = 149{,}9\%$$

Мы получили, что удельный вес кредитных организаций данной группы в среднем ежегодно возрастал почти в полтора раза.

При анализе структуры исследуемого объекта или явления за ряд периодов также можно определить **средний удельный вес** каждой i-ой части за весь рассматриваемый временной интервал. Однако для его расчета одних лишь относительных данных об удельных весах структурных частей недостаточно, необходимо располагать еще и информацией о размерах этих частей в абсолютном выражении. Используя эти данные, средний удельный вес любой i-ой структурной части можно определить по формуле:

$$\overline{d_i} = \frac{\sum_{j=1}^{n} x_{ij}}{\sum_{j=1}^{n} \sum_{i=1}^{k} x_{ij}} \cdot 100 \qquad (10.6)$$

где x_{ij} - величина i-ой структурной части в j- период времени в абсолютном выражении.

Проиллюстрируем эту формулу следующим примером. По итогам биржевых торгов на ММВБ корпоративными ценными бумагами определим средний удельный вес ценных бумаг каждого вида в общем объеме выручки от их реализации (табл. 10.2.):

Таблица 10.2.

Вид ценных бумаг	Объем выручки от продажи			
	200	2001	2002	Итого
Акции, млрд.руб. (x_{1j}) в % к итогу (d_{1j})	472,0 93,1	707,5 92,4	1144,5 90,5	2324,0 ...
Облигации, млрд.руб. (x_{2j}) в % к итогу (d_{2j})	35,1 6,9	58,1 7,6	120,0 9,5	213,2 ...
Всего, млрд.руб.	507,1	765,6	1264,5	2537,2

Определим средний удельный вес выручки от продажи акций в общем объеме выручки от реализации корпоративных ценных бумаг:

$$\bar{d} = \frac{2324,0}{2537,2} \cdot 100 = 91,6\% .$$

Рассчитаем средний удельный вес выручки от продажи облигаций:

$$\bar{d} = \frac{213,2}{2537,2} \cdot 100 = 8,4\% .$$

Итак, в 2000 – 2002 гг. на долю акций в среднем ежегодно приходилось 91,6% общего объема выручки от реализации корпоративных ценных бумаг, а на долю облигаций - только 8,4%. Отметим, что если бы для расчета этих средних показателей мы воспользовались лишь исходными данными в процентах, результаты были бы иными – удельный вес выручки от продажи облигаций был бы заниженным.

10.1.2. Обобщающие показатели структурных сдвигов

В отдельных случаях исследователю необходимо в целом оценить структурные изменения в изучаемом социально-экономическом явлении за определенный временной интервал, которые характеризуют подвижность, или наоборот, стабильность, устойчивость данной структуры. Как правило, это требуется для сравнения динамики одной и той же структуры в различные периоды или нескольких структур, относящихся к разным объектам. Во втором случае число структурных частей у разных объектов необязательно должно совпадать.

Среди применяемых для этой цели обобщающих показателей наиболее распространен **линейный коэффициент «абсолютных» структурных сдвигов**, представляющий собой сумму приростов удельных весов, взятых по модулю, деленную на число структурных частей:

$$\overline{\Delta}_{d_1-d_0} = \frac{\sum_{i=1}^{k} \left| d_{i\,j} - d_{i\,j-1} \right|}{k} \qquad (10.7)$$

Этот показатель отражает то среднее изменение удельного веса (в процентных пунктах), которое имело место за рассматриваемый временной интервал в целом по всем структурным частям совокупности.

Для решения данной задачи также применяют **квадратический коэффициент «абсолютных» структурных сдвигов**, который рассчитывается по формуле:

$$y_{d_1-d_0} = \sqrt{\frac{\sum_{i=1}^{k} (d_{i\,j} - d_{i\,j-1})^2}{k}} \qquad (10.8)$$

Линейный и квадратический коэффициенты «абсолютных» структурных сдвигов позволяют получить сводную оценку скорости изменения удельных весов отдельных частей совокупности. Для сводной характеристики интенсивности изменения удельных весов используется **квадратический коэффициент относительных структурных сдвигов**:

$$y_{\frac{d_1}{d_0}} = \sqrt{\sum_{i=1}^{k} \frac{(d_{i\,j} - d_{i\,j-1})^2}{d_{i\,j-1}} \cdot 100} \qquad (10.9)$$

Данный показатель отражает тот средний относительный прирост удельного веса (в процентах), который наблюдался за рассматриваемый период.

По данным таблицы 10.3 рассчитаем обобщающие показатели структурных сдвигов.

Таблица 10.3.

Структура использования денежных доходов населения РФ
в 1995 - 2001 гг.

N п/п	Направление использования доходов	Удельный вес, в % к итогу			Расчетные графы						
		1995 (d_{i1})	1988 (d_{i2})	2001 (d_{i3})	$\lvert d_{i2}-d_{i1}\rvert$	$(d_{i2}-d_{i1})^2$	$\frac{(d_{i2}-d_{i1})^2}{d_{i1}}$	$\lvert d_{i3}-d_{i2}\rvert$	$(d_{i3}-d_{i2})^2$	$\frac{(d_{i3}-d_{i2})^2}{d_{i2}}$	$\lvert d_{i3}-d_{i1}\rvert$
А	Б	1	2	3	4	5	6	7	8	9	10
1	Покупка товаров и оплата услуг	70,4	77,6	78,6	7,2	51,84	0,74	1,0	1,00	0,01	8,2
2	Оплата обязательных платежей и взносов	5,8	6,1	9,2	0,3	0,09	0,02	3,1	9,61	1,58	3,4
3	Накопления сбережений во вкладах и ценных бумагах	5,4	2,5	4,1	2,9	8,41	1,56	1,6	2,56	1,02	1,3
4	Покупка валюты	14,8	12,1	6,0	2,7	7,29	0,49	6,1	37,21	3,08	8,8
5	Прирост денег на руках	3,6	1,7	2,1	1,9	3,61	1,00	0,4	0,16	0,09	1,5
	Все доходы	100,0	100,0	100,0	15,0	71,24	3,81	12,2	50,54	5,78	23,2

Для расчета линейного коэффициента «абсолютных» структурных сдвигов за первый период (с 1995 по 1998 гг.) и за второй период (с 1998 по 2001 гг.) соответственно воспользуемся данными итогов гр.4 и гр.7 таблицы 10.3:

$$\overline{\Delta}^{I}_{d_1-d_0} = \frac{15,0}{5} = 3,0$$ проц. пункта

$$\overline{\Delta}^{II}_{d_1-d_0} = \frac{12,2}{5} = 2,4$$ проц. пункта

Итак, с 1995г. по 1998г. удельный вес отдельных направлений использования доходов населения изменился в среднем на 3,0 процентного пункта. С 1998г. по 2001г. «абсолютные» структурные сдвиги несколько уменьшились. Этот вывод подтверждается квадратическими коэффициентами «абсолютных» структурных сдвигов (необходимые промежуточные расчеты выполнены в гр.5 и гр.8 таблицы 10.3):

$$\sigma^{I}_{d_1-d_0} = \sqrt{\frac{71,24}{5}} = 3,8$$ проц. пункта,

$$\sigma^{II}_{d_1-d_0} = \sqrt{\frac{50,54}{5}} = 3,2$$ проц. пункта.

Далее определим величину квадратических коэффициентов относительных структурных сдвигов, воспользовавшись итоговыми данными гр.6 и гр.9:

$$\sigma^{I}_{\frac{d_1}{d_0}} = \sqrt{3,81 \cdot 100} = 19,5\%$$

$$\sigma^{II}_{\frac{d_1}{d_0}} = \sqrt{5,78 \cdot 100} = 24,0\%$$

Расчеты показывают, что в относительном выражении за первые три года удельный вес каждой статьи расходов в среднем изменился примерно на 1/5 своей величины; в последующие три года относительные структурные сдвиги заметно усилились.

Для сводной оценки структурных изменений в исследуемой совокупности в целом за рассматриваемый временной интервал, охватывающий несколько недель, месяцев, кварталов или лет, наиболее удобным является **линейный коэффициент «абсолютных» структурных сдвигов за n периодов**:

$$\overline{Д}^{(n)}_{d_1-d_0} = \frac{\sum_{i=1}^{k}\left|d_{in} - d_{i1}\right|}{k(n-1)} \qquad (10.10)$$

Используя итоговые данные гр.10 таблицы 10.3 и учитывая, что n равно 7 годам, получим:

$$\overline{\Delta}^{(n)}_{d_1-d_0} = \frac{23{,}8}{5 \cdot 6} = 0{,}8$$ проц.пункта

Таким образом, за рассматриваемый период среднегодовое изменение по всем направлениям использования доходов составило 0,8 процентного пункта.

Необходимо отметить, что последний показатель может использоваться как для сравнения динамики двух и более структур, так и для анализа динамики одной и той же структуры за разные по продолжительности периоды времени.

10.1.3. Показатели концентрации и централизации

Одна из задач статистического анализа структуры заключается в определении степени концентрации изучаемого признака по единицам совокупности или в оценке неравномерности его распределения. Такая неравномерность может иметь место в распределении доходов по группам населения, жилой площади по группам семей, прибыли по группам предприятий и т.д. При исследовании неравномерности распределения изучаемого признака по территории понятие «концентрация» обычно заменяется понятием «локализация».

Оценка степени концентрации наиболее часто осуществляется по **кривой концентрации (Лоренца)** и рассчитываемым на ее основе характеристикам. Для этого необходимо иметь частотное распределение единиц исследуемой совокупности и взаимосвязанное с ним частотное распределение изучаемого признака. Для удобства вычислений и повышения аналитичности данных единицы совокупности, как правило, разбиваются на равные группы - 10 групп по 10% единиц в каждой, 5 групп по 20% единиц и так далее.

Наиболее известным показателем концентрации является **коэффициент Джини**, обычно используемый как мера дифференциации или социального расслоения:

$$G = 1 - 2\sum_{i=1}^{k} d_{xi}\, d_{yi}^{H} + \sum_{i=1}^{k} d_{xi}\, d_{yi}\ , \qquad (10.11)$$

где d_{xi} - доля i-ой группы в общем объеме совокупности;

d_{yi} - доля i-ой группы в общем объеме признака;

d_{yi}^{H} - накопленная доля i-ой группы в общем объеме признака

Если доли выражены в процентах, данную формулу можно преобразовать:

для 10%-го распределения -

$$G = 110 - 0{,}2\sum_{i=1}^{k} d_{yi}^{H} \qquad (10.12)$$

для 20%-го распределения -

$$G = 120 - 0{,}4\sum_{i=1}^{k} d_{yi}^{H} \qquad (10.13)$$

Чем ближе к 1 (100%) значение данного признака, тем выше уровень концентрации; при нуле мы имеем равномерное распределение признака по всем единицам совокупности.

Оценка степени концентрации также может быть получена на основе **коэффициента Лоренца**:

$$L = \frac{\sum_{i=1}^{k} \left| d_{xi} - d_{yi} \right|}{2} \qquad (10.14)$$

При использовании данного коэффициента можно оперировать как долями единицы, так и процентами. Коэффициент Лоренца изменяется в тех же границах, что и коэффициент Джини.

Определим степень концентрации доходов населения по данным таблицы 10.4.:

Таблица 10.4.

Распределение доходов населения России в 2002г.

20%-ные группы населения	Объем денежных доходов		d_{xi}	$d_{xi}\ d_{yi}$	d_{yi}^{H}	$d_{xi}\ d_{yi}^{H}$	$\lvert d_{xi} - d_{yi}\rvert$
	% к итогу	d_{yi}					
А	1	2	3	4	5	6	7
Первая (с наименьшими доходами)	5,6	0,056	0,2	0,0112	0,056	0,0112	0,144
Вторая	10,4	0,104	0,2	0,0208	0,160	0,032	0,096
Третья	15,4	0,154	0,2	0,0308	0,314	0,0628	0,046
Четвертая	22,8	0,228	0,2	0,0456	0,542	0,1084	0,028
Пятая (с наивысшими доходами)	45,8	0,458	0,2	0,0916	1,000	0,2000	0,258
Итого	100,0	1,0	1,0	0,2000	X	0,4144	0,572

Для расчета коэффициента Джини воспользуемся итоговыми данными граф 4 и 6 таблицы 10.4 :
$G = 1 - 2 \cdot 0{,}4144 + 0{,}2 = 0{,}371$ или 37,1%.

Такой же результат мы получим, выполнив расчеты в процентах:
$G = 120 - 0{,}4\ (5{,}6 + 16{,}0 + 31{,}4 + 54{,}2 + 100{,}0) = 37{,}1\%$.

Второй способ расчета проще, однако, исходная формула незаменима в тех случаях, когда имеются неравные группы по объему совокупности (в нашем примере - по численности населения).

Для сравнения отметим, что наибольшей величины за последние годы коэффициент Джини, рассчитанный по данным о распределении общего объема денежных доходов населения РФ, достигал в 1999г. – 40,0%.

Используя данные графы 7 таблицы 10.4 определим коэффициент Лоренца:

$$L = \frac{0{,}572}{2} = 0{,}286 \text{ или } 28{,}6\%.$$

Оба коэффициента указывают на относительно высокую степень концентрации доходов населения.

Если под концентрацией понимается степень неравномерности распределения изучаемого признака, не связанная ни с объемом совокупности, ни с численностью отдельных групп, то централизация означает сосредоточение объема признака у отдельных единиц (объема продукции данного вида на отдельных предприятиях, капитала в отдельных банках и т.п.). **Обобщающий показатель централизации** имеет следующий вид:

$$I_Z = \sum_{i=1}^{k} \left(\frac{m_i}{M} \right)^2, \qquad (10.15)$$

где m_i - значение признака i-ой единицы совокупности;

М - объем признака всей совокупности.

Максимальное значение, равное 1, данный коэффициент достигает лишь в том случае, когда совокупность состоит только из одной единицы, обладающей всем объемом признака. Минимальное значение коэффициента приближается к нулю, но никогда его не достигает.

Рассмотрим следующий пример. Предположим, выпуск продукции А сконцентрирован на 5 предприятиях, расположенных в трех районах области (табл. 10.5.):

Таблица 10.5.

Район	Число предприятий	Объем производства, млн.руб.		Доля одного предприятия в общем объеме продукции, (гр.3: Итог гр.2)
		всего	в среднем на 1 предприятие (гр.2:гр.1)	
А	1	2	3	4
А	1	5374	5374	0,584
Б	1	1225	1225	0,133
В	3	2610	870	0,094
Итого	5	9209	Х	Х

Вычислим показатель централизации производства данного вида продукции:

$I_z = 0{,}584^2 + 0{,}133^2 + 3 \cdot 0{,}094^2 = 0{,}39$

Рассчитанная величина свидетельствует о высокой степени централизации. Отметим, что аналитическая ценность показателей концентрации и централизации повышается при проведении сравнений во временном или территориальном аспектах.

Глава 11. Индексы

11.1. Общие понятия об индексах

«Индекс» в переводе с латинского - указатель или показатель. В статистике индексом называют показатель относительного изменения данного уровня исследуемого явления по сравнению с другим его уровнем, принятым за базу сравнения. В качестве такой базы может быть использован или уровень за какой-либо прошлый период времени (динамический индекс), или уровень того же явления по другой территории (территориальный индекс). Индексы являются незаменимым инструментом исследования в тех случаях, когда необходимо сравнить во времени или пространстве две совокупности, элементы которых непосредственно суммировать нельзя.

В целом, индексный метод направлен на решение следующих задач:

1) характеристика общего изменения уровня сложного социально-экономического явления;

2) анализ влияния каждого из факторов на изменение индексируемой величины путем элиминирования воздействия прочих факторов;

3) анализ влияния структурных сдвигов на изменение индексируемой величины.

В дальнейшем изложении индексного метода будут использоваться следующие общепринятые обозначения:

i - индивидуальный индекс;

I - сводный индекс;

p - цена;

q - количество;

1 - текущий период;

0 - базисный период.

Простейшим показателем, используемым в индексном анализе, является **индивидуальный индекс**, который характеризует изменение во времени экономических величин, относящихся к одному объекту:

$i_p = \dfrac{p_1}{p_0}$ - индекс цены,

где p_1 - цена товара в текущем периоде;

p_0 - цена товара в базисном периоде;

Изменение физической массы проданного товара в натуральном выражении измеряется индивидуальным индексом физического объема реализации:

$$i_q = \frac{q_1}{q_0}.$$

Изменение стоимостного объема товарооборота по данному товару отразится в значении индивидуального индекса товарооборота. Для его расчета товарооборот текущего периода (произведение цены на количество проданного товара) сравнивается с товарооборотом предшествующего периода:

$$i_{pq} = \frac{p_1 q_1}{p_0 q_0}.$$

Данный индекс также может быть получен как произведение индивидуального индекса цены и индивидуального индекса физического объема реализации.

Индивидуальные индексы, в сущности, представляют собой относительные показатели динамики или темпы роста, и по данным за несколько периодов времени могут рассчитываться в цепной или базисной формах.

В отличие от индексов индивидуальных, сводные индексы позволяют обобщить показатели по нескольким товарам. Исходной формой сводного индекса является агрегатная форма.

Агрегатная форма индекса позволяет найти для разнородной совокупности такой общий показатель, в котором можно объединить все ее элементы. При анализе динамики цен индивидуальные цены различных товаров складывать неправомерно, но суммировать товарооборот по этим товарам вполне допустимо. В текущем периоде такой товарооборот по *n* товарам составит:

$$p_1^1 q_1^1 + p_1^2 q_1^2 + p_1^3 q_1^3 + \ldots + p_1^n q_1^n = \sum p_1 q_1$$

Аналогично получим для базисного периода:

$$p_0^1 q_0^1 + p_0^2 q_0^2 + p_0^3 q_0^3 + \ldots + p_0^n q_0^n = \sum p_0 q_0$$

Если мы сравним товарооборот в текущем периоде с его величиной в базисном периоде, то получим **сводный индекс товарооборота**:

$$\mathrm{I}_{pq} = \frac{\sum p_1 q_1}{\sum p_0 q_0} \qquad (11.1)$$

Для иллюстрации этого и последующих индексов воспользуемся следующими условными данными (табл. 11.1.):

Таблица 11.1.

Цены и объем реализации трех товаров

Товар	Январь		Февраль	
	цена, руб.	продано, тыс.шт.	цена, руб.	продано, тыс.шт.
А	20	9	22	8
Б	60	15	65	13
В	30	7	35	11

Рассчитаем индекс товарооборота:

$$\mathrm{I}_{pq} = \frac{22 \cdot 8 + 65 \cdot 13 + 35 \cdot 11}{20 \cdot 9 + 60 \cdot 15 + 30 \cdot 7} = 1{,}089$$

Рассчитанное значение индекса позволяет заключить, что товарооборот в целом по данной товарной группе в текущем периоде по сравнению с базисным возрос на 8,9% /108,9% - 100,0%/. Отметим, что размер товарной группы, единицы измерения товаров при расчете этого и последующих индексов значения не имеют.

Величина индекса товарооборота формируется под воздействием двух факторов – на нее оказывает влияние как изменение цен на товары, так и изменение объемов их реализации. Для того, чтобы оценить изменение только цен (индексируемой величины), необходимо количество проданных товаров (веса индекса) зафиксировать на каком-либо постоянном уровне. При исследовании динамики таких показателей как цена и себестоимость физический объем реализации обычно фиксируют на уровне текущего периода. Таким способом получают **сводный индекс цен** (по методу Пааше):

$$\mathrm{I}_{p} = \frac{p_1^1 q_1^1 + p_1^2 q_1^2 + ... + p_1^n q_1^n}{p_0^1 q_1^1 + p_0^2 q_1^2 + ... + p_0^n q_1^n} = \frac{\sum p_1 q_1}{\sum p_0 q_1} \qquad (11.2)$$

Для рассматриваемого примера получим:

$$\mathrm{I}_{p} = \frac{22 \cdot 8 + 65 \cdot 13 + 35 \cdot 11}{20 \cdot 8 + 60 \cdot 13 + 30 \cdot 11} = 1{,}107.$$

Таким образом, по данной товарной группе цены в феврале по сравнению с январем в среднем возросли на 10,7%. При построении данного индекса цена выступает в качестве индексируемой величины, а количество проданного товара - в качестве веса.

Рассмотрим сводный индекс цен более подробно. Числитель данного индекса содержит фактический товарооборот текущего периода. Знаменатель же представляет собой условную величину, показывающую каким был бы товарооборот в текущем периоде при условии сохранения цен на базисном уровне. Поэтому соотношение этих двух категорий и отражает имевшее место изменение цен.

Числитель и знаменатель сводного индекса цен также можно интерпретировать и по-другому. Числитель представляет собой сумму денег, фактически уплаченных покупателями за товары в текущем периоде. Знаменатель же показывает, какую сумму покупатели заплатили бы за те же товары, если бы цены не изменились. Разность числителя и знаменателя будет отражать величину экономии (если знак «-») или перерасхода («+») покупателей региона от изменения цен:

$$\text{E} = \sum p_1 q_1 - \sum p_0 q_1 = 1406 - 1270 = 136 \textit{тыс.руб.}$$

Необходимо отметить, что в статистической практике также используется сводный индекс цен, построенный по методу Ласпейреса, когда веса или объемы продаж фиксируются на уровне базисного, а не текущего периода:

$$\text{I}_p = \frac{\sum p_1 q_0}{\sum p_0 q_0} \tag{11.3}$$

Третьим индексом в рассматриваемой индексной системе (включающий индекс цен, рассчитанный по методу Паше) является **сводный индекс физического объема реализации**. Он характеризует изменение количества проданных товаров не в денежных, а в физических единицах измерения. Весами в данном случае выступают цены, которые фиксируются на базисном уровне:

$$\text{I}_q = \frac{\sum q_1 p_0}{\sum q_0 p_0} \tag{11.4}$$

В нашем случае индекс составит:

$$\text{I}_q = \frac{8 \cdot 20 + 13 \cdot 60 + 11 \cdot 30}{9 \cdot 20 + 15 \cdot 60 + 7 \cdot 30} = 0{,}984$$

Физический объем реализации (товарооборота) сократился на 1,6% (98,4%-100,0%).

Между рассчитанными индексами существует следующая взаимосвязь:

$$\text{I}_p \cdot \text{I}_q = \text{I}_{pq}$$

или

$$1{,}107 \cdot 0{,}984 = 1{,}089$$

На основе данной взаимосвязи по значениям двух известных индексов всегда можно определить неизвестное значение третьего индекса.

11.2. Средние формы сводных индексов

На практике при расчете индексов часть необходимой информации может отсутствовать или базироваться на результатах выборочных обследований. В подобных случаях вместо индексов в агрегатной форме удобнее использовать средние арифметические и средние гармонические индексы. Любой сводный индекс можно представить как среднюю взвешенную из индивидуальных индексов. Однако при этом форму средней нужно выбрать таким образом, чтобы полученный средний индекс был тождественен исходному агрегатному индексу.

Предположим, мы располагаем данными о стоимости проданной продукции в текущем периоде и индивидуальными индексами цен, полученными, например, в результате выборочного наблюдения. Тогда при расчете сводного индекса цен по методу Пааше можно использовать следующую замену:

$$p_0 q_1 = \frac{1}{i_p} p_1 q_1$$

В целом же **сводный индекс цен** в данном случае будет выражен **в форме средней гармонической**:

$$\mathrm{I}_p = \frac{\sum p_1 q_1}{\sum \frac{1}{i_p} p_1 q_1} \tag{11.5}$$

Рассмотрим следующий условный пример (табл. 11.2.):

Таблица 11.2.

Данные о реализации и ценах по товарной группе

Товар	Реализация в текущем периоде, руб.	Изменение цен в текущем периоде по сравнению с базисным, %
А	44000	-1,3
Б	56000	+4,2
В	31000	+2,5

Последняя графа таблицы содержит информацию об изменениях индивидуальных индексов цен или их приростах. С учетом этих приростов несложно определить первоначальные значения индексов, которые по товарам А, Б и В соответственно составляют 0,987, 1,042 и 1,025.

Рассчитаем значение сводного индекса:

$$\mathrm{I}_p = \frac{44000 + 56000 + 31000}{\frac{44000}{0{,}987} + \frac{56000}{1{,}042} + \frac{31000}{1{,}025}} = 1{,}019.$$

Произведенный расчет позволяет заключить, что цены по данной товарной группе в среднем возросли на 1,9%.

Мы получили значение сводного индекса цен в среднегармонической форме, соответствующее сводному индексу Пааше в агрегатной форме. Для получения значения, соответствующего индексу Ласпейреса, индекс цен необходимо представить в среднеарифметической форме. При этом используется следующая замена:

$$p_1 q_0 = i_p p_0 q_0$$

С учетом этой замены **сводный индекс цен в среднеарифметической форме** можно представить следующим образом:

$$\mathrm{I}_p = \frac{\sum i_p p_0 q_0}{\sum p_0 q_0} \qquad (11.6)$$

Среднеарифметическая форма также может использоваться при расчете сводного индекса физического объема товарооборота. При этом производится замена:

$$q_1 p_0 = i_q q_0 p_0$$

Тогда сводный индекс физического объема товарооборота имеет вид:

$$\mathrm{I}_q = \frac{\sum i_q q_0 p_0}{\sum q_0 p_0} \qquad (11.7)$$

Для иллюстрации этой формы расчета воспользуемся следующим примером (табл. 11.3.):

Таблица 11.3.

Данные о реализации трех товаров в натуральном и стоимостном выражении

Товар	Стоимостной объем реализации в базисном периоде, руб.	Изменение физического объема реализации в текущем периоде по сравнению с базисным, %
А	87000	+3,4
Б	54000	-12,0
В	73000	-8,5

Индивидуальные индексы физического объема соответственно будут равны 1,034; 0,880; 0,915. С учетом этого рассчитаем среднеарифметический индекс:

$$\mathrm{I}_q = \frac{1{,}034 \cdot 87000 + 0{,}880 \cdot 54000 + 0{,}915 \cdot 73000}{87000 + 54000 + 73000} = 0{,}955.$$

В результате расчета мы получили, что физический объем реализации товаров рассматриваемой товарной группы в среднем снизился на 4,5%.

11.3. Расчет сводных индексов за последовательные периоды

На практике, как правило, расчет индексов не является разовой акцией. Индексы позволяют получать сводную оценку изучаемых процессов постоянно, месяц за месяцем, год за годом. Однако при этом для достижения сопоставимости они должны рассчитываться по единой методологии. Такая методология или схема расчета индексов за несколько последовательных временных периодов называется системой индексов.

В зависимости от информационной базы и целей исследования индексная система может строится по-разному. Рассмотрим некоторые варианты ее построения их на примере сводного индекса цен, рассчитываемого за n периодов.

Если сравнивать цены каждого периода с ценами периода предшествующего получаемая индексная система будет включать цепные индексы, отражающие изменение цен за каждый из периодов рассматриваемого временного интервала. При этом в качестве весов можно использовать объемы реализации каждого конкретного периода или же постоянные объемы какого-либо периода, принятого в качестве базисного. Тогда индексная система будет включать индексы, соответственно, с переменными или с постоянными весами. Цепные индексы цен с переменными весами имеют следующий вид:

$$\mathrm{I}_{p\,1/0}=\frac{\sum p_1q_1}{\sum p_0q_1};\quad \mathrm{I}_{p\,2/1}=\frac{\sum p_2q_2}{\sum p_1q_2};\quad \mathrm{I}_{p\,3/2}=\frac{\sum p_3q_3}{\sum p_2q_3};\quad \ldots\quad \mathrm{I}_{p\,n/n-1}=\frac{\sum p_nq_n}{\sum p_{n-1}q_n}.$$

При использовании постоянных весов система преобразуется:

$$\mathrm{I}_{p\,1/0}=\frac{\sum p_1q_0}{\sum p_0q_0};\quad \mathrm{I}_{p\,2/1}=\frac{\sum p_2q_0}{\sum p_1q_0};\quad \mathrm{I}_{p\,3/2}=\frac{\sum p_3q_0}{\sum p_2q_0};\quad \ldots\quad \mathrm{I}_{p\,n/n-1}=\frac{\sum p_nq_0}{\sum p_{n-1}q_0}.$$

Отметим, что использование постоянных весов более предпочтительно, так как рассчитываемые таким образом индексы мультипликативны, т.е. их можно последовательно перемножать и получать величину показателя за более продолжительный период. Так, например, располагая индексами цен за три последовательных месяца можно получить сводную оценку изменения цены в целом за квартал и т.п. Индексы с переменными весами такой возможности не предоставляют.

Если сравнивать цены каждого периода с ценами какого-либо базисного периода (как правило – начального) получаемая индексная система будет включать базисные индексы, отражающие изменение цен накопленным итогом, т.е. с начала рассматриваемого временного интервала. Например, изменение цен в январе по сравнению с декабрем предшествующего года, в феврале – по сравнению с тем же декабрем и т.д. При этом в качестве весов также можно использовать объемы реализации каждого конкретного периода или же постоянные объемы периода, принятого в качестве базисного. Система базисных индексов с переменными весами имеет следующий вид:

$$I_{p\,1/0}=\frac{\sum p_1q_1}{\sum p_0q_1};\quad I_{p\,2/0}=\frac{\sum p_2q_2}{\sum p_0q_2};\quad I_{p\,3/0}=\frac{\sum p_3q_3}{\sum p_0q_3};\quad \ldots\quad I_{p\,n/0}=\frac{\sum p_nq_n}{\sum p_0q_n}.$$

Базисные индексы цен с постоянными весами рассчитываются по формулам:

$$I_{p\,1/0}=\frac{\sum p_1q_0}{\sum p_0q_0};\quad I_{p\,2/0}=\frac{\sum p_2q_0}{\sum p_0q_0};\quad I_{p\,3/0}=\frac{\sum p_3q_0}{\sum p_0q_0};\quad \ldots\quad I_{p\,n/0}=\frac{\sum p_nq_0}{\sum p_0q_0}.$$

Отметим, что использование постоянных весов приводит базисные индексы, так же как и индексы цепные, к сопоставимому виду.

11.4. Индексный анализ влияния структурных изменений

Индексы позволяют оценить динамику показателей, характеризующих разнородные в качественном отношении совокупности, как правило, товарные группы. Однако, даже если рассматриваемая совокупность однородна (товар или вид продукции одного вида) на величине результативного показателя будет отражаться влияние структурных изменений, например, изменений в структуре производства или реализации данного товара по территориям. Рассмотрим случай, когда один товар или вид продукции реализуется или производится в нескольких местах (табл. 11.4.):

Таблица 11.4.

Данные о ценах и объемах реализации товара «Х» в двух регионах

Регион	**2003**		**2004**	
	цена, тыс.руб.	**продано, шт.**	**цена, тыс.руб.**	**продано, шт**
1	7	36000	8	10000
2	5	12000	6	34000

Проведем анализ изменения цен на данный товар. Из таблицы видно, что цена в каждом регионе возросла. Для сводной оценки этого роста воспользуемся средними показателями. Так как в данном случае реализуется один и тот же товар, вполне правомерно рассчитать его среднюю цену за июнь и за июль. **Индекс цен переменного состава** представляет собой соотношение средних значений за два рассматриваемые периода:

$$\mathrm{I}_p^{пс} = \frac{\sum p_1 q_1}{\sum q_1} : \frac{\sum p_0 q_0}{\sum q_0} =$$
$$= \frac{8 \cdot 10000 + 6 \cdot 34000}{10000 + 34000} : \frac{7 \cdot 36000 + 5 \cdot 12000}{36000 + 12000} =$$
$$= 6{,}45 : 6{,}50 = 0{,}992. \qquad (11.8)$$

Рассчитанное значение индекса указывает на снижение средней цены данного товара на 0,8%, т.е. с 6,50 тыс. руб. до 6,45 тыс. руб. В то же время, из приведенной выше таблицы видно, что цена в каждом регионе в 2003 г. по сравнению с 2002 г. возросла. Данное несоответствие объясняется влиянием изменения структуры реализации товаров по регионам: в 2002 г. по более высокой цене продали товара втрое больше, а в 2003 г. ситуация принципиально изменилась (в данном условном примере для наглядности числа подобраны таким образом, чтобы это различие в структуре продаж было очевидным). Иными словами, на динамике средней цены данного товара отразились структурные сдвиги в рассматриваемой совокупности. Оценить воздействие этого фактора можно с помощью **индекса структурных сдвигов**:

$$\mathrm{I}^{стр} = \frac{\sum p_0 q_1}{\sum q_1} : \frac{\sum p_0 q_0}{\sum q_0} =$$
$$= \frac{7 \cdot 10000 + 5 \cdot 34000}{44000} : \frac{7 \cdot 36000 + 5 \cdot 12000}{48000} = 0{,}839 \qquad (11.9)$$

Первая формула в этом индексе позволяет ответить на вопрос, какой была бы средняя цена в 2003 г., если бы цены в каждом регионе сохранились на уровне предыдущего года. Вторая часть формулы отражает фактическую среднюю цену 2002 г. В целом по полученному значению индекса мы можем сделать вывод, что за счет структурных сдвигов цены снизились на 16,1%.

Последним в данной системе является **индекс цен фиксированного состава**, который не учитывает влияние структуры:

$$\mathrm{I}_p^{фс} = \frac{\sum p_1 q_1}{\sum p_0 q_1} = 1{,}183. \qquad (11.10)$$

Полученное значение индекса позволяет сделать вывод о том, что если бы структура реализации товара «Х» по регионам не изменилась, средняя цена возросла бы на 18,3%. Однако, влияние на среднюю цену фактора структурных изменений оказалось сильнее и в итоге цена даже несколько снизилась. Данное взаимодействие рассматриваемых факторов отражается в следующей взаимосвязи:

$$\mathrm{I}_p^{фс} \cdot \mathrm{I}^{ст\,\mathrm{р}} = \mathrm{I}_p^{пс}$$

$1{,}183 \cdot 0{,}839 = 0{,}992.$

Аналогично строятся индексы структурных сдвигов, переменного и фиксированного состава для анализа изменения себестоимости, урожайности и других показателей.

Заключение

Мы рассмотрели приемы сбора, обработки и анализа статистических данных, которые являются методологическим базисом любой статистической работы. В то же время, необходимо отметить, что статистическое наблюдение не является обязательным этапом статистического исследования. Во многих случаях экономист-аналитик имеет дело с материалом, полученным из баз данных, бюллетеней информационных агентств, статистических сборников и других источников. Тогда работа должна начинаться с проверки полноты и качества данных, их группировки, а при отсутствии необходимости в этих этапах - с расчета индивидуальных и обобщающих показателей.

Рассмотренные приемы и методы с успехом могут использоваться не только в практике статистического анализа. Статистическая методология исследования в настоящее время заняла прочные позиции во многих областях знания. Статистические формулы находят применение в макро- и микроэкономике, оценке бизнеса и недвижимости, финансовом анализе, техническом анализе товарных и финансовых рынков.

Более того, подвергающийся статистической обработке материал не обязательно должен относиться к экономической области. В большинстве случаев, описанные приемы и показатели будут работоспособны и эффективны при обобщении и анализе технической, биологической, медицинской, демографической и социологической информации.

Рассматриваемые в пособии методы в большинстве случаев иллюстрированы практическими примерами. Подобные вычисления при небольших объемах совокупности или коротких динамических рядах не очень трудоемки. При работе же с большими массивами статистической информации необходимо использовать прикладное программное обеспечение, существенно ускоряющее и упрощающее все расчеты. Среди наиболее распространенных современных программных продуктов следует отметить пакеты Мезозавр, ОЛИМП, САНИ, Эвриста, STATISTICA, STATGRAPHICS и SPSS.

Литература

1. Елисеева И.И., Юзбашев М.М. Общая теория статистики: Учебник / Под. ред. чл.-корр. РАН И.И.Елисеевой. - М.: Финансы и статистика, 2004.
2. Ефимова М.Р., Петрова Е.В., Румянцев В.Н. Общая теория статистики: Учебник. - М.: ИНФРА-М, 2004.
3. Общая теория статистики: Статистическая методология в изучении коммерческой деятельности: Учебник / Под ред. А.А.Спирина, О.Э.Башиной. - М.: Финансы и статистика, 2004.
4. Практикум по теории статистики: Учеб. пособие /Под.ред. проф. Р.А. Шмойловой. – М.: Финансы и статистика, 2004.
5. Статистический словарь / Гл. ред. М.А.Королев. - М.: Финансы и статистика, 1989.
6. Шмойлова Р.А., Минашкин В.Г., Садовникова Н.А., Шувалова Е.Б. Теория статистики: Учебник / Под ред. проф. Шмойловой Р.А. - М.: Финансы и статистика, 2004.
7. Тюрин Ю.Н., Макаров А.А. Статистический анализ данных на компьютере / Под ред. В.Э.Фигурнова. - М.: ИНФРА-М, Финансы и статистика, 1998.