РАЗДЕЛ 1 ТЕОРИЯ СТАТИСТИКИ

ГЛАВА 1 ПРЕДМЕТ, МЕТОД И ЗАДАЧИ СТАТИСТИКИ

1.1 Предмет и метод статистики

Слово «статистика» происходит от латинского слова status – состояние, положение вещей. Первоначально оно употреблялось в значении «политическое состояние».

Статистикой называют:

- отрасль знаний, объединяющей принципы и методы работы с числовыми данными, характеризующими массовые явления. В этом смысле статистика включает в себя: общую теорию статистики, теорию вероятностей, математическую статистику;
- отрасль практической деятельности, направленную на сбор, обработку, анализ и публикации статистических данных, отражающих явления и процессы общественной деятельности;
- собирание массовых данных, их обобщение, представление, анализ и интерпретация. Это особый метод, который используется в различных сферах деятельности, в решении разнообразных задач.

Предметом статистики являются массовые явления любой природы, в том числе и в экономике; статистика изучает количественную сторону этих явлений в неразрывной связи с их качественным содержанием в конкретных условиях места и времени.

Массовые явления – явления, повторяющиеся в пространстве и времени и отражающие некоторую статистическую закономерность.

Социально-экономическая жизнь общества проявляется в различного рода массовых явлениях: производство разного рода продукции, ее потребление, экспорт и импорт, перевозка грузов и пассажиров и другие явления экономической, культурной и политической жизни.

Статистическая закономерность – такая закономерность, когда в каждом отдельном явлении то, что присуще всей совокупности явлений (необходимое), проявляется в единстве с индивидуальным, присущим лишь этому конкретному явлению (случайным).

Специфика предмета статистики обусловливает специфику статистической методологии.

Статистическая методология представляет собой совокупность общих правил (принципов) и специальных приемов и методов, используемых в статистике.

Статистические методы включают как простые методы, которые могут быть понятны любому человеку, так и сложные математические процедуры, доступные специалистам.

Основные методы, используемые в статистике:

Группа статистических методов	Этап статистического исследования
Статистическое наблюдение	Сбор данных
Группировка Сводка	Обобщение данных
Статистические таблицы и графики	Представление данных
Выборочный метод Метод средних величин Вариационный анализ Корреляционный и регрессионный анализ Метод динамических рядов Индексный метод и т.п.	Анализ и интерпретация данных

Выбор методов анализа данных зависит от цели исследования и задач, служащих ее достижению.

1.2 Понятия и категории статистической науки

Статистическая совокупность – множество однокачественных варьирующихся явлений.

В данном явлении можно выделить два основных момента:

во-первых, это множество однородных явлений, т.е. явлений, объединенных общим качеством, представляющих собой проявление одной и той же закономерности (однородность множества (совокупности) устанавливается в каждом конкретном статистическом исследовании в соответствии с его целями и познавательными задачами);

во-вторых, это множество варьирующихся явлений, отличающихся друг от друга своими характеристиками. Именно это свойство вызывает необходимость изучения всего множества явлений одного вида. Если бы единицы совокупности были полностью тождественны друг другу, то достаточно было бы исследовать одно из них, е было бы потребности обращаться к их множеству.

Единица совокупности (элемент) – частный случай проявления изучаемой закономерности, это неделимый первичный элемент совокупности, выражающий ее качественную однородность.

Различают основную совокупность – полную совокупность изучаемых единиц и частную совокупность, включающую не все единицы, а лишь определенную часть (представляющую определенный тип явления).

Единицы статистической совокупности характеризуются свойствами, именуемыми в статистике признаками. Статистика изучает явления через их признаки. Чем более однородна совокупность, тем больше общих признаков имеют ее единицы и меньше варьируются их значения.

Классификация признаков

Основные классификации	Виды признаков
По отношению к цели исследования	Существенные (главные, выражающие содержательную сторону явлений) Несущественные (второстепенные)
По характеру выражения	Описательные (атрибутивные), выраженные словами (национальность человека, разновидность почв). Выделяют номинальные описательные признаки (нельзя ранжировать данные) и порядковые (можно ранжировать данные) - работники по мастерству, студенты по успеваемости. Количественные, выраженные числами

По характеру вариации	Альтернативные, которые могут принимать только два значения (пол человека, ходовая система трактора) Дискретные, которые могут принимать ограниченное количество значений в рамках данного диапазона (число членов семьи, количество этажей здания, комнат в квартире) Непрерывные, которые могут принимать бесконечное множество значений в рамках данного диапазона
По способу измерения	Первичные, которые непосредственно измеряются, учитывают (возраст, рост человека) Вторичные, которые рассчитываются через первичные по определенным формулам (средний балл, процент посещаемости занятий)
По отношению ко времени	Моментные, которые характеризую состояние объекта на какой-то момент (численность присутствующих на лекции) Интервальные (периодические), которые характеризуют результаты процесса за некоторый период (число занятий, пропущенных за семестр)

Объем (мера) совокупности (ее подмножества) – количество единиц, входящих в совокупность (ее подмножество).

1.3 Организация статистики в РФ

Официальная государственная статистика – система организаций, отвечающих за сбор социальных и экономических данных и их обобщение по определенной территории, за их точность и достоверность.

Главным органом государственной статистики является Федеральная служба государственной статистики – Росстат (до 2004 г. - Госкомстат России).

Росстат России включает управления:

- статистического планирования и организации статистического наблюдения;
- национальных счетов;

– статистики предприятия и структурных обследований;
– статистики труда;
– переписи населения и демографической статистики;
– статистики зарубежных стран и международного сотрудничества;
– статистик уровня жизни и обследований населения;
– статистики услуг, транспорта и связи;
– статистики основных фондов и строительства;
– статистики внутренней и внешней торговли;
– статистики окружающей среды и сельского хозяйства;
– статистики цен и финансов;
– сводной информации.

Федеральная служба государственной статистики входит в структуру федеральных органов исполнительной власти. Росстат является методологическим и организационным центром работы всех служб государственной статистики. Здесь разрабатываются федеральный план статистических работ на год и перспективу, методология расчета статистических показателей, сбора и разработки статистических данных, подводятся итоги работы государственной статистики за год.

Методологическая работа Росстата направлена на внедрение интегрированной системы учета и статистики, соответствующей международным стандартам, прежде всего на разработку системы национальных счетов РФ, позволяющей исследовать формирование основных пропорций экономики и рассчитывать важнейшие макроэкономические показатели, используемые в международной практике, а также на измерение инфляции и уровня жизни.

В каждом субъекте РФ имеется региональный комитет по статистике. В целом структура органов государственной статистики соответствует административному делению страны. Низовым звеном являются районные или городские отделы государственной статистики, которые имеются в административных районах краев, областей, а также в административных районах крупных городов (Москва, С.-Петербург).

Подразделения (отделы), выделяемые в Росстате, имеются и в региональных комитетах по статистике.

Основные функции всех статистических органов состоят в сборе, обработке, анализе и представлении данных в удобном для пользователя виде.

Статистическая служба оперативно предоставляет в установленном порядке официальную статистическую информацию Президенту РФ, Правительству РФ, Федеральному собранию, иным органам государственной власти и местного самоуправления, средствам массовой информации, организациям и гражданам, а также международным организациям; разрабатывает в установленном порядке официальную статистическую методологию; осуществляет подготовку, проведение и подведение итогов Всероссийской переписи населения и других статистических обследований в установленной сфере деятельности; разрабатывает и ведет общероссийские классификаторы технико-экономической и социальной статистики; взаимодействует с органами государственной власти иностранных государств и международными организациями в установленной сфере деятельности и т.д.

Официальная статистика в России является централизованной: руководство ею составляет функцию самостоятельного государственного учреждения - Федеральная служба государственной статистики. Также статистическую работу ведут все министерства и ведомства, т.е. существует ведомственная статистика.

Источником статистических сведений могут быть не только официальные органы, но и исследовательские группы («альтернативная статистика») – социологические институты и службы, организующие опросы населения, результаты которых доводятся как до граждан, так и до исполнительных и законодательных органов.

ГЛАВА 2 СТАТИСТИЧЕСКОЕ НАБЛЮДЕНИЕ

2.1 Статистическое наблюдение, его формы, виды и способы

Статистическое наблюдение – планомерное, научно организованное получение первичной статистической информации об изучаемом явлении и процессе.

Статистическое наблюдение должно отвечать следующим требованиям:

1. Наблюдаемое явление должно иметь научную или практическую ценность.
2. Должна обеспечиваться полнота фактов, относящихся к рассматриваемому вопросу. Если отсутствуют полные данные, анализ и выводы могут быть ошибочными.
3. Должна осуществляться тщательная и всесторонняя проверка (контроль) качества собираемых фактов для обеспечения достоверности статистических данных.
4. Наблюдение должно проводиться по заранее разработанному плану обеспечивающему научное решение программно-методологических и организационных вопросов наблюдения.

Составляющие статистического наблюдения:

Статистик – Инструментарий статистического наблюдения – объект наблюдения – данные наблюдения.

Статистическое наблюдение характеризуют формой, видом и способом.

Формы статистического наблюдения: отчетность, специально организованное наблюдение.

Отчетность - это организационная форма статистического наблюдения, при которой сведения поступают в виде обязательных отчетов в определенные сроки и по утвержденным формам. Отчетность должна вытекать из данных первичных документов.

Под первичным документом понимают систему унифицированных записей, фиксирующих осуществление хозяйственных операций (например, счет-фактура, накладная, путевой лист и т.д.).

Специально организованное статистическое наблюдение это сбор сведений посредством переписей, единовременных учетов и обследований. Например, перепись населения, социологические обследования, переписи промышленного оборудования и т.п.

Виды статистического наблюдения

По степени охвата единиц совокупности: сплошное и несплошное.

Сплошное наблюдение охватывает все без исключения единицы совокупности. Например, получение отчетности от предприятий и учреждений.

Несплошное наблюдение охватывает лишь часть изучаемой совокупности. Выделяют следующие виды несплошного наблюдения:

Выборочное наблюдение – наблюдение над частью совокупности, отобранной в случайном порядке. Данные выборочного наблюдения используются для характеристики всей совокупности. Это основной вид несплошного статистического наблюдения и находит применение в промышленности – для контроля качества продукции, в сельском хозяйстве – для выявления продуктивности скота, в торговле – для изучения степени удовлетворения спроса населения и т.д.

Наблюдение методом основного массива – это наблюдение за частью единиц совокупности, вносящих наибольший вклад в изучаемое явление. Например, вследствие концентрации производства в отрасли несколько крупных предприятий могут давать основной объем продукции, в то время как большая масса мелких предприятий выпускает ее незначительную часть. Поэтому достаточно обследовать эти несколько крупных предприятий, чтобы сделать вывод о положении в отрасли в целом.

Анкетное наблюдение – сбор данных, основанный на принципе добровольного заполнения адресатами анкет (опросных листов). Такой вид наблюдения может применяться в тех случаях, когда не требуется высокая точность сведений, а нужны приблизительные характеристики (при социологических опросах, при изучении спроса населения на отдельные товары и т.д.).

Монографическое наблюдение представляет собой подробное описание отдельных единиц в статистической совокупности. Данное наблюдение направлено на решение задач более глубокого исследования отдельных единиц совокупности, поэтому обычно проводится в отношении типичных единиц или характерных типов явлений. Программа монографического наблюдения предусматривает определенную свободу действий статистика. Это означает, что в процессе наблюдения не только даются ответы на поставленные вопросы, но и фиксируются признаки, стороны деятельности, которые могут представлять интерес для дальнейшего изучения и составления программы наблюдения уже для всей совокупности.

Преимущества несплошного наблюдения:

1) сокращение затрат (стоимостных, материальных, трудовых, временных);
2) более детальная программа наблюдения и более совершенный способ учета фактов.

По характеру регистрации данных во времени: непрерывное (текущее) и прерывное.

Непрерывное (текущее) статистическое наблюдение – такое систематическое наблюдение, при котором регистрация фактов производится по мере их совершения. При текущем наблюдении нельзя допускать значительного разрыва между моментом возникновения факта и моментом его регистрации. Например, учет произведенной продукции, учет отпуска материалов со склада и т.п.

Прерывное наблюдение проводится не постоянно по мере возникновения фактов, а либо через определенные промежутки времени (периодическое наблюдение, напр., перепись скота), либо время от времени, без соблюдения строгой периодичности или вообще единожды (единовремененное наблюдение, напр., перепись жилого фонда).

Выделяют три *способа наблюдения*:

Непосредственное наблюдение – это наблюдение, при котором факты устанавливаются и фиксируются регистратором путем замера, взвешивания или подсчета (при учете остатков товаров в торговле за основу берется их инвентаризация, при переписи оборудования сведения заносятся в формуляр на основе личного осмотра машин и т.д.).

Документальный учет – это способ наблюдения, при котором источником сведений служат соответствующие документы. Поскольку источником сведений при составлении первичных документов является непосредственное наблюдение, то при надлежащей организации первичного учета и правильной разработке на их основе форм статистической отчетности документальный способ наблюдения обеспечивает большую точность сведений.

Опрос – это наблюдение, при котором факты регистрируются со слов опрашиваемого. Применяют следующие основные способы опроса:

Экспедиционный (устный) способ заключается в том, что регистраторы сами фиксируют факты (заполняют формуляр наблюдения) со слов опрашиваемого

Способ саморегистрации (самоисчислении) – фиксация фактов (заполнение формуляра) производится самим опрашиваемым. Обязанность счетчиков состоит в раздаче бланков наблюдения опрашиваемым, инструктаже их, в сборе заполненных формуляров, которые при этом проверяются.

Корреспондентский способ заключается в том, что формуляры заполняются и отсылаются опрашиваемыми без участия регистраторов (сбор информации от покупателей о товаре). Этот способ не требует больших затрат, но не обеспечивает высокого качества материалов, так как проверить точность сообщаемых сведений непосредственно на месте не всегда возможно.

2.2 План статистического наблюдения

Программно-методологические задачи, содержащиеся в плане наблюдения:

1) определение цели наблюдения;
2) определение объекта наблюдения;
3) определение единицы наблюдения;
4) разработка программы наблюдения;
5) проектирование формуляров и текстов инструкций;
6) установление источников и способов сбора данных, формы и вида наблюдения.

Целью статистического наблюдения является получение достоверной информации, способствующей достижению цели исследования.

Объектом статистического наблюдения совокупность единиц изучаемого явления, о которых должны быть собраны статистические данные.

Минимальная (или максимальная) величина количественного признака, используемая для отграничивания объекта наблюдения, называется цензом наблюдения.

Единица наблюдения – это первичный элемент объекта статистического наблюдения, являющийся носителем признаков, подлежащих регистрации, и основой ведущегося при наблюдении счета.

Исходя из содержания объекта, цели и задач статистического наблюдения разрабатывается программа наблюдения. **Программой статистического наблюдения** называется перечень признаков (показателей), подлежащих изучению (при непосредственном наблюдении или документальном учете). При опросе программа – это перечень вопросов, на которые должны быть получены правдивые, достоверные ответы по каждой единице наблюдения.

Статистические формуляры – это основной инструмент наблюдения (бланки и отчетность). Обязательным элементом статистического формуляра является титульная и адресная его части.

Различают два вида носителей информации: индивидуальные и списочные формуляры. **Индивидуальный формуляр** содержит сведения об одной единице совокупности. В **списочном формуляре** содержатся сведения по нескольким единицам совокупности.

Организационные задачи, содержащиеся в плане статистического наблюдения:

1) выбор органа наблюдения;
2) выбор времени наблюдения;
3) выбор места проведения наблюдения;
4) определение предварительных списков обследуемых единиц;
5) расстановка и подготовка кадров и др.

Наблюдение может проводиться собственными силами (внутренний орган наблюдения) либо внешними организациями (внешний орган наблюдения).

Вопрос о времени проведения наблюдения включает выбор сезона наблюдения, установление срока (периода) и критического момента (даты) наблюдения.

Сезон (время года) для наблюдения следует выбирать такой, в котором изучаемый объект пребывает в обычном для него состоянии.

Период проведения наблюдения – время начала и окончания сбора сведений.

Критической называют дату, по состоянию на которую сообщаются сведения. Критическим моментом наблюдения выбирают полночь, момент окончания одних суток и начала других.

Территория проведения наблюдения охватывает все места нахождения единиц наблюдения; ее границы зависят от определения единицы наблюдения.

Наиболее существенный этап подготовительной работы – составление списка обследуемых единиц. Этот список необходим для проверки полноты и своевременности поступивших сведений, так и для определения объема работ и расчета необходимого количества работников для проведения статистического наблюдения.

В целях успешного осуществления статистического наблюдения важное значение имеют подготовка статистического инструментария (бланки, инструкции и т.д.), его размножение и своевременное снабжение им персонала, проводящего наблюдение.

2.3 Ошибки статистического наблюдения и меры борьбы с ними

К данным, собираемым в ходе статистического наблюдения, предъявляется два важных требования: достоверность и сопоставимость.

Достоверность (адекватность, истинность) – это степень приближения, соответствия данных тому, что есть на самом деле. Достоверность данных связана со всеми составляющими статистического наблюдения.

Расхождение между фактическим значением и величиной, полученной в результате наблюдения, будет погрешностью (ошибкой) статистического наблюдения. Выделяют ошибки регистрации и ошибки репрезентативности (представительности).

Ошибки регистрации возникают вследствие неправильного установления фактов в процессе наблюдения или неправильной их записи. Они подразделяются на случайные и систематические и могут быть как при сплошном, так и несплошном наблюдении.

Случайные ошибки – это ошибки, которые не имеют какой-либо направленности и могут быть допущены как опрашиваемыми в их ответах, так и регистраторами при заполнении бланков. Случайные ошибки всегда непреднамеренные.

Систематические ошибки имеют определенную направленность и могут быть преднамеренными (сознательными) и непреднамеренными.

Ошибки репрезентативности (представительности) свойственны несплошному наблюдению. Они возникают в результате того, что состав отобранной для обследования части единиц совокупности недостаточно полно отображает состав всей изучаемой совокупности, хотя регистрация сведений по каждой отобранной для обследования единице была проведена точно.

Пример. Пусть имеется полная (общая) совокупность из 10 единиц, характеризуемых признаком X. Значения признака X приведены ниже.

i	1	2	3	4	5	6	7	8	9	10
X	40	40	40	40	40	40	36	36	36	36

Из данной совокупности были извлечены три частичные (неполные) совокупности: А, Б, В, каждая объемом 5 единиц.

Случай А:

i	1	2	3	4	5
X	40	40	40	40	40

Данная частичная совокупность включает только единицы с большими значениями признака (т.е. ее структура не соответствует структуре общей совокупности).

Случай Б:

i	6	7	8	9	10
X	40	36	36	36	36

Структура данной частичной совокупности также не соответствует структуре общей совокупности, так как в ней преобладают единицы с малыми значениями признака.

Случай В:

i	1	5	6	9	10
X	40	40	40	36	36

Структура данной частичной совокупности полностью соответствует структуре общей совокупности.

Необходимо по данным частичным совокупностям оценить среднее значение *X* для общей совокупности.

Искомое значение равно:

$$\bar{\tilde{O}} = \frac{40+40+40+40+40+40+36+36+36+36}{10} = 38{,}4$$

Оценка данной величины для случая А:

$$\tilde{\tilde{O}} = \frac{40+40+40+40+40}{5} = 40$$

что дает завышенную оценку среднего.

Оценка данной величины для случая Б:

$$\tilde{\tilde{O}} = \frac{40+36+36+36+36}{5} = 36{,}8$$

что дает заниженную оценку среднего.

Оценка данной величины для случая В:

$$\tilde{\tilde{O}} = \frac{40+40+40+36+36}{5} = 38{,}4$$

что совпадает с истинным значением.

Для выявления и устранения ошибок регистрации может применяться счетный и логический контроль собранного материала.

Счетный контроль основан на детерминированной связи между признаками и заключается в проверке точности арифметических расчетов, применявшихся при составлении отчетности или заполнении формуляров обследования.

Логический контроль основан на логической взаимосвязи между признаками и заключается в проверке ответов на вопросы программы наблюдения путем логического осмысления или путем сравнения полученных данных с другими источниками по этому же вопросу.

ГЛАВА 3 СВОДКА И ГРУППИРОВКА СТАТИСТИЧЕСКИХ ДАННЫХ

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 объекта в целом при помощи обобщающих показателей, что достигается путем сводки и группировки первичного статистического материала.

Метод группировок в единстве с другими статистическими методами является важным средством социально-экономического познания, а также ведущим звеном в статистическом исследовании. Можно собрать прекрасный статистический материал, но испортить его неумелой сводкой и группировкой. Рассмотрим основные понятия и категории.

Сводка – это комплекс последовательных операций по обобщению конкретных единичных фактов, образующих совокупность, для 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.

Сложная сводка – это комплекс операций, включающих группировку единиц наблюдения, подсчет итогов по каждой группе и по всему объекту и представление результатов группировки и сводки в виде статистических таблиц.

Проведение сложной сводки необходимо осуществлять по следующим этапам:

- 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.

3.2 Понятие, последовательность и приемы проведения статистической группировки. Виды статистических группировок

Группировкой называется разделение единиц изучаемой совокупности на однородные группы по определенным, существенным для них признакам. Группировка в статистическом анализе выполняет следующие определенные функции:

- выделение социально-экономических типов явлений;
- изучение структуры и структурных сдвигов, происходящих в социально-экономических явлениях;
- анализ взаимосвязей между явлениями.

Виды статистических группировок. В соответствии с функциями группировки, различают следующие ее виды: типологическая, структурная, аналитическая.

Типологическая группировка – это разделение качественно неоднородной совокупности на отдельные качественно однородные группы и выявление на этой основе экономических типов явлений. Таким образом, основная задача такой группировки – это идентификация типов социально-экономических явлений, поэтому важное значение при ее построении должно уделяться выбору группировочного признака.

Структурная группировка – это выявление закономерностей распределения единиц однородной совокупности по варьирующим значениям исследуемого признака. Она позволяет изучить структуру совокупности и происходящих в ней сдвигов. Надобность в таких группировках возникает потому, что однородность однокачественных явлений, элементов, входящих в статистическую совокупность, отнюдь не означает их тождественности.

Структурные группировки отличаются от типологических не столько по внешнему виду, сколько по целям, т.е. отличаются по уровню качественных различий между группами.

Аналитическая группировка – это исследование взаимосвязей варьирующих признаков в пределах однородной совокупности. При ее построении можно установить взаимосвязи между двумя признаками и более. При этом один признак будет результативным, а другой – факторным.

Факторными называются признаки, оказывающие влияние на изменение результативных. Результативными называются признаки, изменяющиеся под влиянием факторных.

Основные этапы построения аналитической группировки следующие:

- обоснование и выбор факторного и результативного признаков;
- группировка единиц совокупности по факторному признаку;
- подсчет числа единиц в каждой из образованных групп и определение объема варьирующих признаков в созданных группах;
- исчисление средних размеров результативного показателя (признака) по каждой из образованных групп;
- оформление результатов группировки в таблице;
- сопоставление изменения значений факторного и результативного признаков, определяющее характер связи между ними, т.е. выявление взаимосвязи между признаками, когда с возрастанием значения факторного признака систематически возрастает или убывает значение результативного признака.

Принципы построения статистических группировок и классификаций. Построение группировки начинается с определения состава группировочных признаков.

Выбор группировочного признака, т.е. признака, по которому производится объединение единиц исследуемой совокупности в группы – один из самых существенных и сложных вопросов теории группировки и статистического исследования. От правильного выбора группировочного признака зависят выводы статистического исследования. В качестве основания группировки необходимо использовать существенные обоснованные признаки.

Группированным признаком называется признак, по которому проводится разбиение единиц совокупности на отдельные группы. В основание группировки могут быть положены как количественные, так и атрибутивные признаки. Первые имеют числовое выражение (объем торгов, возраст человека, доход семьи и т.д.), а вторые отражают состояние единицы совокупности (пол человека, семейное положение, отраслевую принадлежность предприятия, его форму собственности и т.д.).

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

Если группировка строится по атрибутивному признаку, то число групп, как правило, будет таким, сколько имеется градаций, видов состояний у этого признака. Например, группировка предприятий по формам собственности учитывает муниципальную, федеральную и собственность субъектов Федерации.

Если группировка проводится по количественному признаку, то число групп зависит от числа единиц исследуемого объекта и степени колеблемости группировочного признака, в каждом отдельном случае его необходимо обосновать.

Определение числа групп можно осуществить и математическим путем с использованием формулы Стерджесса

$$n = 1 + 3{,}322\ lgN,$$

где n – число групп;

N – число единиц совокупности.

Согласно формуле, выбор числа групп зависит от объема совокупности.

Недостаток формулы состоит в том, что ее применение дает хорошие результаты, если совокупность состоит из большого числа единиц и распределение единиц по признаку, положенному в основание группировки, близко к нормальному.

Для определения числа групп можно воспользоваться данными таблицы 1, где количество групп задается, не так жестко, как в предыдущей формуле.

Таблица 1 – Определение числа групп в зависимости от численности совокупности

Численность совокупности, ед.	Рекомендуемое число групп
До 40	3-4
40-60	4-5

60-100	5-6
100-300	6-8
Свыше 300	8-10

Когда определено число групп, то следует определить интервалы группировки.

Интервал – это значение варьирующего признака, лежащее в определенных границах. Каждый интервал имеет свою величину, верхнюю и нижнюю границы или хотя бы одну из них. Нижней границей интервала называется наименьшее значение признака в интервале, а верхней границей – наибольшее значение признака в интервале. Величина интервала представляет собой разность между верхней и нижней границами интервала.

Интервалы группировки в зависимости от их величины бывают *равные и неравные*.

Если вариация признака проявляется в сравнительно узких границах и распределение носит равномерный характер, то строят группировку с равными интервалами.

Величина равного интервала определяется по следующей формуле:

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

где X_{max} и X_{min} – максимальное и минимальное значения признака в совокупности;

n – число групп.

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

Существуют нижеследующие правила записи числа шага интервала. Если величина интервала, рассчитанная по формуле, представляет собой величину, которая имеет один знак до запятой (например, 0,88; 1,585; 4,71), то в этом случае полученные значения целесообразно округлить до десятых и их использовать в качестве шага интервала. В приведенном выше примере это будут соответственно значения: 0,9; 1,6; 4,7. Если рассчитанная величина интервала имеет две значащие цифры до запятой и несколько знаков после запятой (например, 15,985), то это значение необходимо округлить до целого числа (до 16). В случае, когда рассчитанная величина интервала представляет собой трехзначное, четырехзначное и так далее число, эту величину следует округлить до ближайшего числа, кратного 100 или 50. Например, 557 следует округлить до 600.

Если размах вариации признака велик и его значения варьируются неравномерно, то необходимо использовать группировку с неравными интервалами.

Неравные интервалы могут быть прогрессивно-возрастающими или прогрессивно-убывающими в арифметической или геометрической прогрессии. Величина интервалов, изменяющихся в арифметической прогрессии, определяется следующим образом:

$$h_{i+1} = h_i + a,$$

а в геометрической прогрессии:

$$h_{i+1} = h_i \times q,$$

где a – константа, имеющая для прогрессивно-возрастающих интервалов знак «+», а для прогрессивно-убывающих интервалов знак «–»;

q – константа (для прогрессивно-убывающих интервалов $q >$ 1; в другом случае – $q < 1$).

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

Например, при построении группировки промышленных предприятий строительного комплекса по показателю численности работающих, который варьирует от 400 до 2800 чел., целесообразно рассматривать неравные интервалы. Поэтому следует образовывать неравные интервалы: 400–800; 800–1600; 1600–2800, т.е. величина каждого последующего интервала больше предыдущего на 400 чел. и увеличивается в арифметической профессии.

Интервалы группировок могут быть закрытыми и открытыми.

Закрытыми называются интервалы, у которых имеются верхняя и нижняя границы. У *открытых* интервалов указана только одна граница: верхняя – у первого, нижняя – у последнего.

Например, группы коммерческих банков по уровню дохода работающих в них сотрудников (тыс. руб.): до 10; 10–20; 20–30; 30–40; 40 и более.

При группировке единиц совокупности по количественному признаку границы интервалов могут быть обозначены по разному, в зависимости от того, непрерывный это признак или прерывный.

Если основанием группировки служит непрерывный признак (например, группы строительных фирм по объему работ (млн. руб.): 12–14, 14–16, 16–18, 18–20), то одно и то же значение признака выступает и верхней, и нижней границами двух смежных интервалов. В данном случае объем работ 14 млн. руб. составляет верхнюю границу первого и нижнюю границу второго интервалов; 16 млн. руб. – соответственно второго и третьего и т.д., т.е. верхняя граница i-го интервала равна нижней границе $(i + 1)$ интервала.

При таком обозначении границ может возникнуть вопрос, в какую группу включать единицы объекта, значения признака у которых совпадают с границами интервалов. Например, во вторую или третью группу должна войти строительная фирма с объемом работ 16 млн. руб. Если нижняя граница формируется по принципу «включительно», а верхняя – по принципу «исключительно», то фирма должна быть отнесена к третьей группе, в противном случае – ко второй.

3.3 Ряды распределения

Разновидностью группировок являются ряды распределения.

Ряды распределения – это упорядоченное распределение единиц совокупности по определенному признаку.

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

Атрибутивным называют ряд распределения, построенный по качественным признакам, например, распределение студентов группы по полу.

Вариационным рядом называют ряд распределения, построенный по количественному признаку. В зависимости от характера вариации признака различают дискретные и интервальные ряды.

Дискретный вариационный ряд характеризует распределение единиц совокупности по признаку, принимающему только целые значения.

Построение *интервального ряда* целесообразно при непрерывной вариации признака.

Анализ рядов распределения проводится на основе их графического изображения.

Полигон используется при изображении дискретных вариационных рядов, *гистограмма* применяется для изображения интервального вариационного ряда.

Кумулятивная кривая строится по накопленным частотам, которые откладываются по оси ординат, а по оси абсцисс откладываются варианты ряда.

3.4 Статистические таблицы, правила построения и применения. Виды статистических таблиц

Результаты сводки и группировки материалов статистического наблюдения, как правило, оформляют в виде таблиц.

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

Статистическая таблица имеет подлежащее и сказуемое. ***Подлежащее*** статистической таблицы характеризует объект исследования. В подлежащем дается перечень единиц совокупности либо групп исследуемого объекта по существенным признакам. ***Сказуемое*** статистической таблицы - система показателей, которыми характеризуется объект изучения, т.е. подлежащее таблицы.

Обычно подлежащее таблицы располагается слева и составляет содержание строк. Сказуемое, как правило, представлено справа и составляет содержание граф.

Выбор места для подлежащего и сказуемого таблиц часто зависит от обозримости статистического материала, изложенного в таблице. Поэтому расположение подлежащего и сказуемого таблицы может меняться местами.

В зависимости от структуры подлежащего различают статистические таблицы ***простые***, в подлежащем которых дается простой перечень единиц совокупности, и ***сложные***, подлежащее которых содержит группы единиц совокупности по одному (групповые) или нескольким (комбинационные) количественным либо атрибутивным признакам.

Простые таблицы подразделяются на *перечневые и монографические*. Перечневые таблицы в свою очередь могут быть видовыми, территориальными, временными и др.

В зависимости от структурного строения сказуемого различают статистические таблицы с его простой и сложной разработкой. Простая разработка сказуемого означает последовательное перечисление показателей, характеризующих подлежащее. При сложной разработке сказуемого признаки, характеризующие подлежащее, берутся в сочетании, комбинации.

Основа статистической таблицы

Содержание строк	**Наименование граф (верхние заголовки)**				**Итоговая графа**
А	**1**	**2**	**…**	**n-1**	**n**
Наименование строк (боковые заголовки)					
Итоговая строка					

Статистические таблицы необходимо правильно и грамотно составлять и оформлять, только тогда они становятся средством обобщения и систематизации статистических данных.

Основные правила построения и оформления таблиц

1. Таблица должна иметь общий заголовок, который должен кратко и четко выражать основное содержание таблицы. В заголовке таблицы должны найти отражение объект, признак, время и

место совершения события. Если таблица содержит данные о размерах определенного социально-экономического явления, выраженного в одинаковых единицах измерения, то эту единицу измерения необходимо вынести в заголовок таблицы, поставив ее в скобки или отделив запятой от общего заголовка.

2. Все строки и графы таблицы должны иметь названия. Повторяющиеся термины, имеющие единый смысл, объединяются в общие заголовки. Все слова в заголовках подлежащего и сказуемого таблиц должны по возможности писаться полностью.

3. Графы и строки могут нумероваться, если их число велико или по их данным производятся вычисления. Графы, содержащие подлежащее, нумеруются заглавными буквами алфавита; графы, содержащие сказуемое, - арабскими цифрами в порядке возрастания.

4. В графах и строках должны быть указаны единицы измерения, соответствующие показателям, содержащимся в подлежащем и сказуемом, при этом следует использовать общепринятые сокращения единиц измерения (руб., чел., м^2 и т.д.)

5. Все данные одной строки (графы) следует представлять с одинаковой степенью точности.

6. Информация, расположенная в строках (графах) таблицы, завершается итогом. Причем итоговые строки могут располагаться как в первых, так и в последних строках таблицы. В сложных таблицах итоговые графы и строки необходимы.

7. Если в таблице приводятся взаимосвязанные данные (например, абсолютные данные о расчетном задании и выполнении расчетного задания, число банков и удельный вес банков (% к итогу), абсолютный прирост, темп роста, темп прироста и т.д.), целесообразно их располагать в рядом стоящих графах.

8. Все клетки таблицы должны быть заполнены. Отсутствие данных об анализируемом явлении может быть обусловлено различными причинами, поэтому в статистике при заполнении таблиц используют следующие условные обозначения:

- если численное значение признака неизвестно, то ставится многоточие «...» или пишется «нет сведений»;

- если данная позиция не имеет осмысленного содержания, бессмысленна, то ставится крестик «Х»;

- если явление отсутствует, то ставится прочерк, тире «-»;

- если явление существует, но значение его показателя очень мало, то используют обозначения (0,0) или (0,00), предполагающие возможность наличия малых чисел.

9. В ряде случаев к таблице даются примечания, сноски, в которых приводятся необходимые разъяснения (например, данные, рассчитанные по методологии, отличной от методологии расчета остальных данных). Обычно такие сноски располагаются ниже таблицы.

10. Необходимо указывать источники данных, приведенных в таблицах (название публикаций, обследования или указание на условность данных).

3.5 Статистические графики, их виды, правила построения и использования

Несмотря на многообразие видов графических изображений, каждый график должен включать следующие элементы: графический образ, поле графика, масштабные ориентиры и систему координат.

Графический образ – геометрические знаки, совокупность точек, линий, фигур, с помощью которых изображаются статистические величины.

Поле графика представляет собой пространство, в котором размещаются геометрические знаки.

Масштабные ориентиры определяются масштабом и масштабной шкалой.

Масштаб – это мера перевода числовой величины в графическую, а масштабная шкала – линия, определенные точки которой могут быть прочитаны как определенные числа.

Шкала состоит из линии (носителя шкалы) и ряда намеченных на ней точек, расположенных в определенном порядке. Носитель шкалы может быть представлен прямой или кривой линией.

Шкалы могут быть равномерными и неравномерными.

Для размещения геометрических знаков в поле графика необходима система координат. Наиболее распространенной является система прямоугольных координат.

Масштабом равномерной шкалы называется длина отрезка, принятого за единицу и измеренного в каких-либо мерах.

Рассмотрим построение основных видов диаграмм на конкретных числовых примерах.

На столбиковых диаграммах статистические данные изображаются в виде вытянутых по вертикали прямоугольников.

При построении столбиковых диаграмм необходимо выполнять следующие требования:

1) шкала, по которой устанавливается высота столбика, должна начинаться с нуля;

2) шкала должна быть непрерывной;

3) основания столбиков должны быть равны между собой.

Пример. Изобразим графически данные о числе дошкольных учреждений в Пензенской области (на конец года), ед.: 1991 – 904, 1998 – 502, 2005 – 360, 2009 г. – 355.

На горизонтальной оси откладываем произвольные основания четырех столбиков с произвольным расстоянием между основанием. Масштаб на вертикальной оси – 1 деление = 200 единиц (рисунок 2).

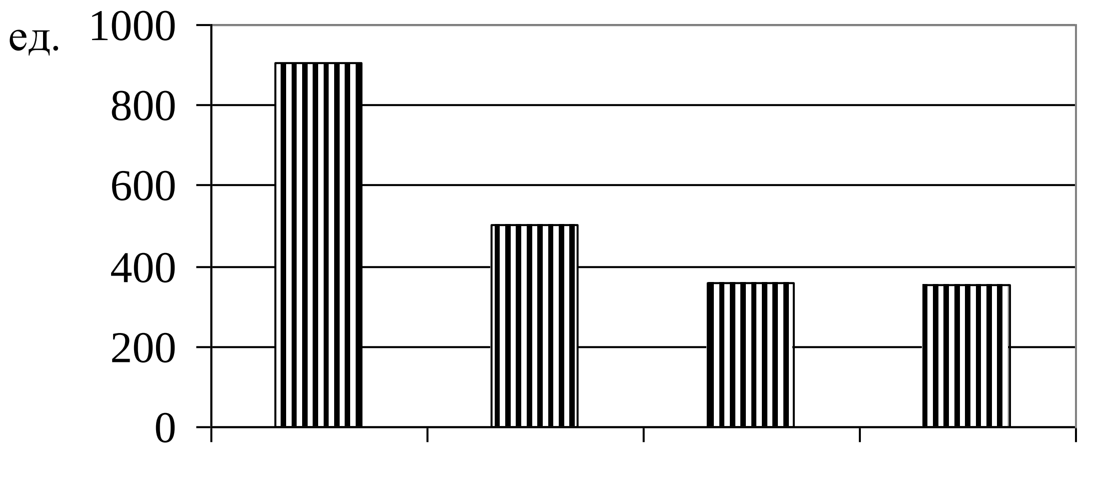

Рисунок 2 – Динамика дошкольных учреждений в Пензенской области (на конец года)

На столбиковой диаграмме изображаемые величины пропорциональны высоте столбиков.

Принципы построения полосовых или ленточных диаграмм абсолютно аналогичен. Разница заключается в том, что масштабной шкалой будет горизонтальная ось.

В круговых и квадратных диаграммах величина изображаемого явления выражается размером площади. Чтобы построить квадратную диаграмму, нужно из сравниваемых статистических величин извлечь квадратные корни, а затем построить квадраты со сторонами, пропорциональными полученным результатам.

Пример. Построим квадратную диаграмму о численности студентов в негосударственных средних специальных учебных заведениях в Пензенской области на начало учебного года: 2005/2006 – 411 чел., 2009/2010 – 112 чел.

Извлечем квадратные корни из показателей численности студентов. Это составит соответственно 20,2 и 10,6 чел. Выбираем масштаб. Примем 1 см за четыре человека. Тогда стороны первого квадрата составят 5,1 (20,2 : 4), а второго − 2,7 (10,6 : 4) (рисунок 3).

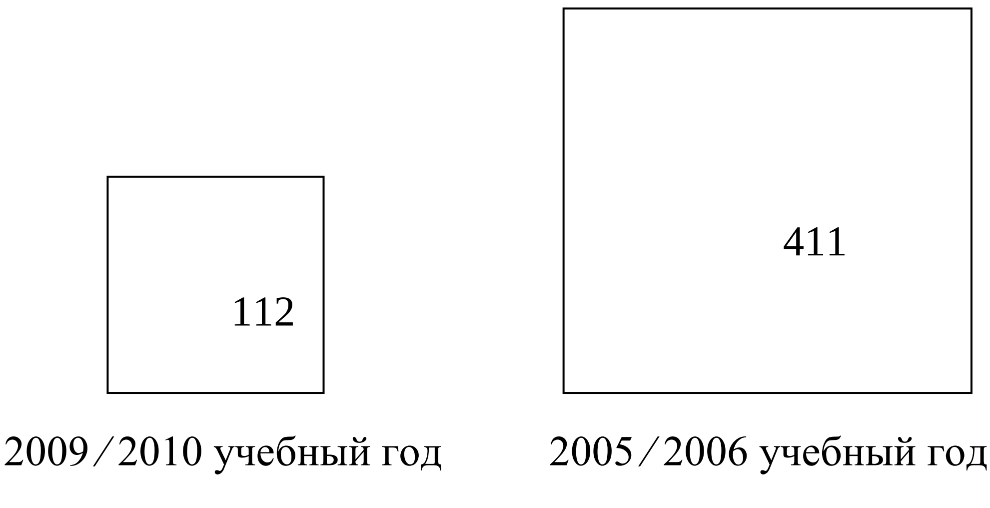

Рисунок 3 – Численность студентов в негосударственных средних специальных учебных заведениях в Пензенской области (на конец года, чел.)

Круговые диаграммы строятся аналогично. Разница в том, что на графике вычеркиваются круги, площади которых пропорциональны квадратным корням из изображаемых величин. Основываясь на данных предыдущего примера, необходимо построить два круга с радиусами 5,16 и 2,7 см.

Секторные диаграммы удобно стоить следующим образом: вся величина явления принимается за 100 %, рассчитываются доли отдельных его частей в процентах. Круг разбивается на секторы пропорционально частям изображаемого целого. Таким образом, на 1 % приходится 3,6°. Для получения центральных углов секторов, изображающих доли частей целого, необходимо их процентное выражение умножить на 3,6°.

Пример. Изобразим с помощью секторной диаграммы состав студентов государственных и муниципальных высших и средних специальных учебных заведений в Пензенской области в начале 2009/2010 учебного года по формам обучения. По очной форме обучения обучалось 51 %, по очно-заочной форме – 3 %, по заочной – 44 %, экстернат – 2 %. Построим круг произвольного радиуса. По данным о доле студентов по каждой форме обучения определим величину центральных углов: для очной формы – 183,6° (51:3,6) для очно-заочной – 10,8° (3:3,6), для заочной – 158,4° (44:3,6), для экстерната – 7,2° (2:3,6).

При помощи транспортира разделим круг на соответствующие сектора (рисунок 4).

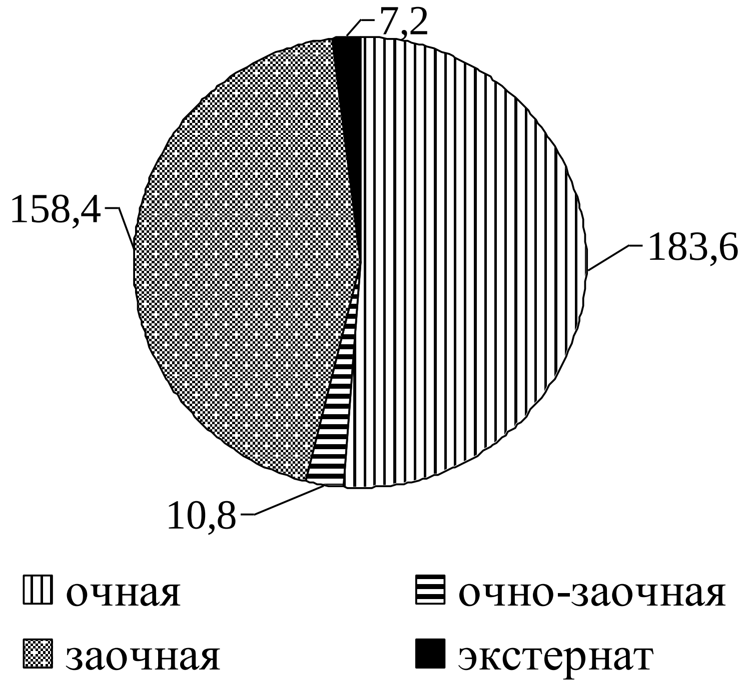

Рисунок 4 – Структура форм обучения студентов в государственных и муниципальных высших учебных заведениях в Пензенской области на начало 2007/2008 учебного года, проц.

Для одновременного сопоставления трех величин, связанных между собой таким образом, что одна величина является произведением двух других применяют диаграммы, называемые «знак Варзара».

Знак Варзара представляет собой прямоугольник, у которого один сомножитель принят за основание, другой – за высоту, а вся площадь равна произведению.

Пример. Имеются данные по Пензенской области о производстве зерновых культур в 2009 г. Урожайность составила 20,3 ц с 1 га, а посевная площадь 771,0 тыс. гектар.

Положим в основание прямоугольника урожайность, а в качестве высоты возьмем показатель посевной площади. Тогда площадь полученного прямоугольника будет пропорциональна валовому сбору зерновых культур (рисунок 5).

Линейные диаграммы широко применяются для характеристики изменений явлений во времени, выполнения плановых заданий, а также для изучения рядов распределения, выявления связи между явлениями. Линейные диаграммы строятся на координатной сетке. Геометрическими знаками в линейных диаграммах служат точки и последовательно соединяющие их отрезки прямой, которые складываются в ломанные кривые.

Рисунок 5 – Зависимость валового сбора зерновых культур в Пензенской области от урожайности *и посевной площади в 2009 г.*

Пример. Изобразим динамику урожайности зерновых культур в Пензенской области по данным таблицы 17.

Таблица 17 – Динамика урожайности зерновых культур за 2000-2009 гг. в хозяйствах всех категорий в Пензенской области, ц с 1 га

2000 г.	2001 г.	2002 г.	2003 г.	2004 г.	2005 г.	2006 г.	2007 г.	2008 г.	2009 г.
10,0	13,8	14,8	13,7	13,4	13,5	15,0	12,8	19,3	20,3

В прямоугольной системе координат на ось ординат нанесем данные об урожайности зерновых культур, на ось абсцисс – данные о периодах времени (рисунок 6).

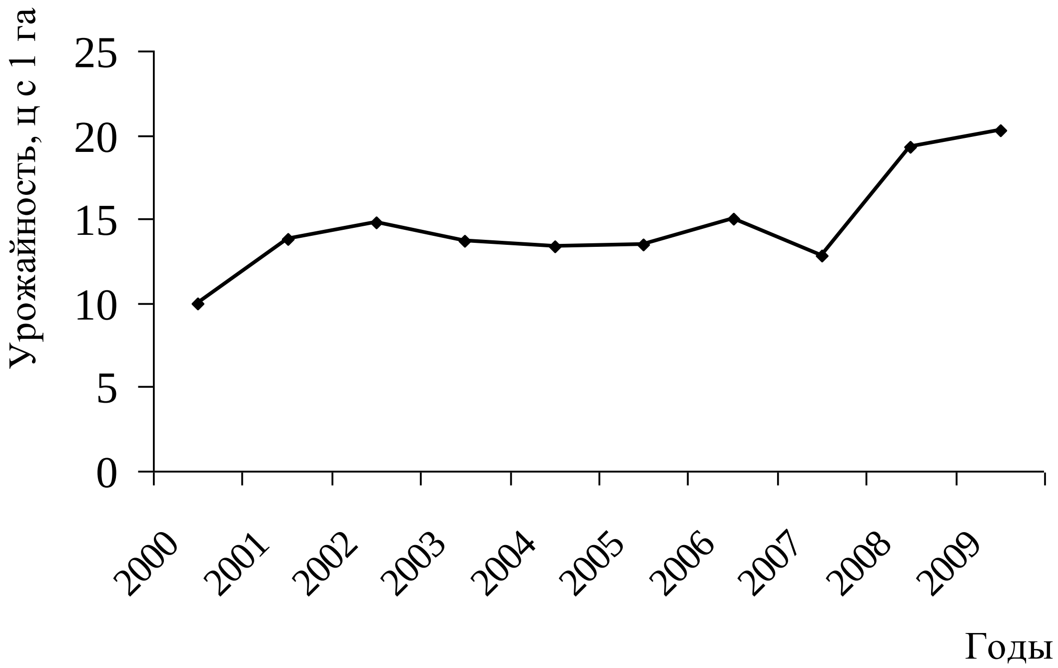

Рисунок 6 – Динамика урожайности зерновых культур в Пензенской области в хозяйствах всех категорий

Ряды распределения чаще всего изображаются в виде полигона или гистограммы. Полигон строят в основном для изображения дискретных рядов (рисунок 7). При его построении на оси абсцисс откладываются значения варьирующего признака, а на оси ординат – абсолютные или относительные численности единиц совокупности (частоты или частости).

Полигон на рисунке 7 построен на основании (условных) данных о распределении семей по числу детей.

Гистограмма распределения применяется чаще всего для изображения интервальных рядов. Для ее построения по оси абсцисс откладываются интервалы признака, а по оси ординат – численности единиц совокупности. На отрезках, изображающих интервалы, стоят прямоугольники, площади которых пропорциональны численностям единиц (рисунок 8).

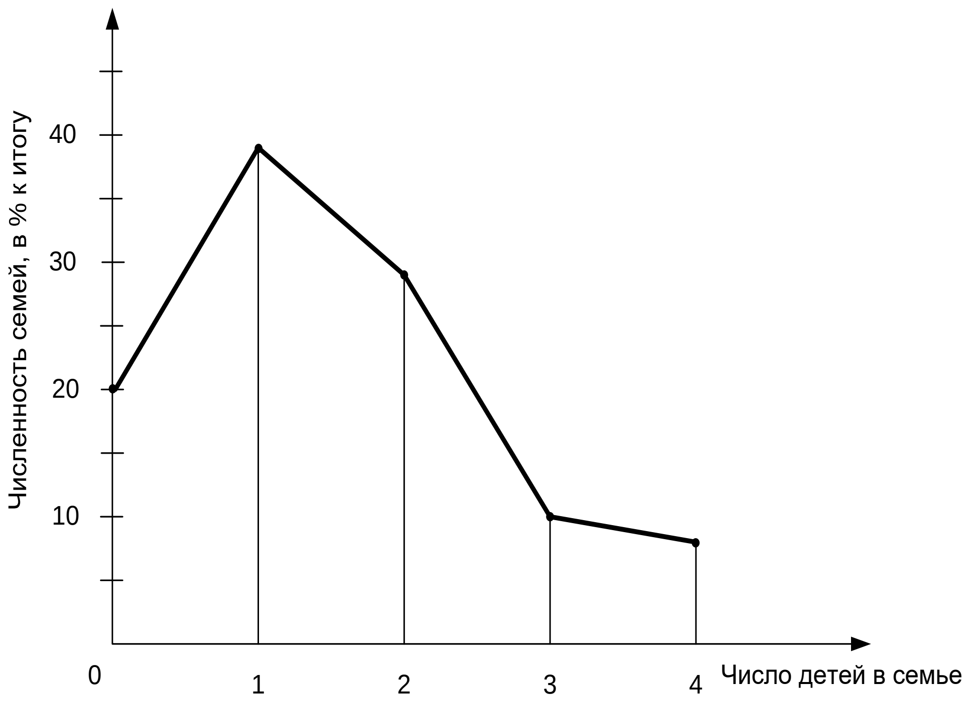

Рисунок 7 – Полигон распределения семей по числу детей в одном из регионов в 2010 г.

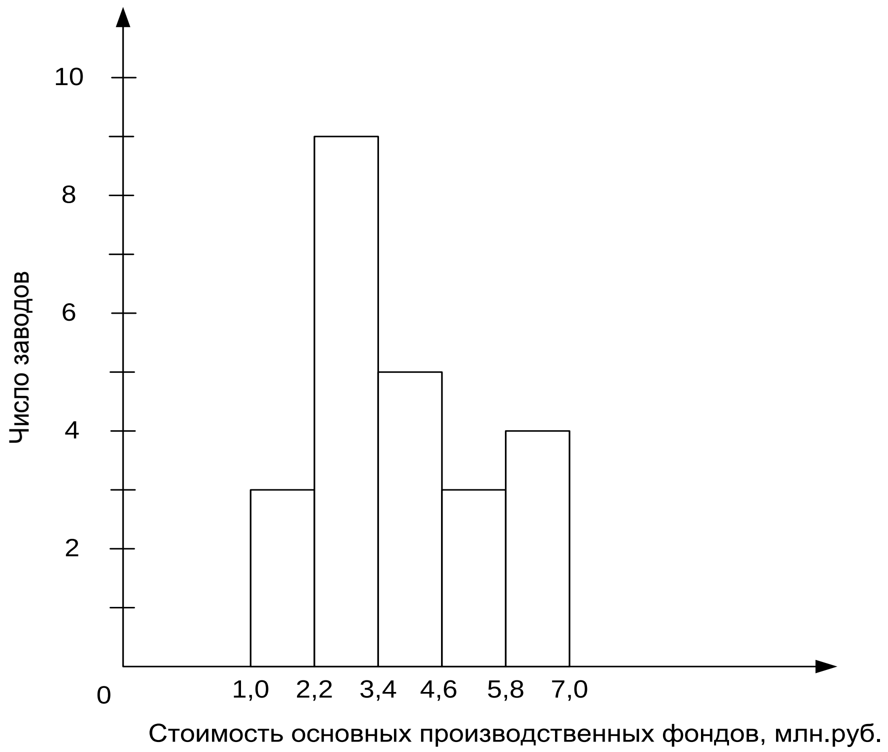

Рисунок 8 – Гистограмма распределения фирм в одной из отраслей по стоимости основных производственных фондов в 2010 г.

Пример. Имеются следующие условные данные.

Таблица 18 – Продажа моркови на рынках сельхозпродуктов города в 2010 г., ц

Месяц	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	Итого за год
Морковь	36	42	44	54	43	70	41	43	39	37	37	34	520

Определяем среднемесячную продажу моркови. Она составляет 43,3 ц. Вычертим круг радиусом, равным среднемесячному показателю (R = 43,3 ц). На горизонтальном диаметре построим шкалу, взяв длину радиуса, равную 2,7 см. Следовательно, 1 см = 43,3 / 2,7 ≈ 16 ц. Затем весь круг разделим на 12 радиусов (соответственно числу месяцев в году). На радиусе сделаем отметку согласно масштабу исходя из приведенных данных за каждый месяц. Данные, которые превысили среднемесячный уровень, отмечаются за пределами окружности на продолжении радиуса. Отметки различных месяцев соединяются между собой (рисунок 9).

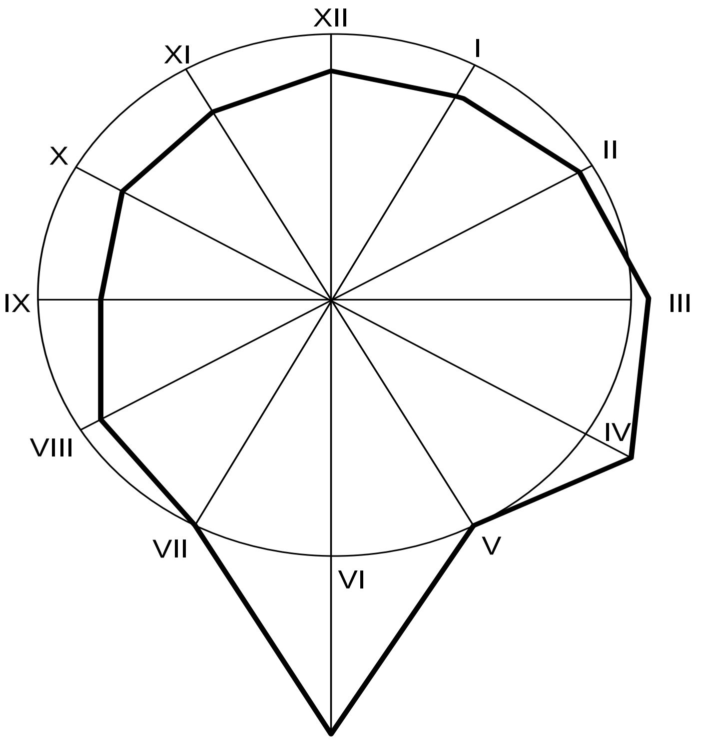

Рисунок 9 – Продажа моркови на рынках сельхозпродуктов в одном из городов в 2010 г., ц

Статистические карты представляют собой вид графических изображений статистических данных на схематический географической карте и характеризуют уровень или степень распространения того или иного явления на определенной территории. При этом различают картограммы и картодиаграммы.

Картограмма – это схематическая карта, на которой штриховкой различной густоты, точками или окраской определенной степени насыщенности показывается сравнительная интенсивность какого-либо показателя в пределах каждой единицы нанесенного на карту территориального деления (например, плотность населения по области, распределение районов по урожайности и т.п.).

Картодиаграмма представляет собой сочетание диаграммы с географической картой. В качестве изобразительных знаков в картограмме используют 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.

Основные правила построения статистических графиков

1. Важно использовать такие способы изображения, которые наилучшим образом отвечают содержанию и логической природе отображаемых показателей;

2. Вид графического изображения выбирается в зависимости от цели использования;

3. При построении графика необходима максимальная точность, построение в соответствии с масштабом использованием масштабных шкал (равномерных или неравномерных);

4. Должна быть выбрана система координат;

5. График обязательно должен сопровождаться заголовком, в котором указывают, какой изображен показатель, по какой территории, за какое время (период времени), в каких единицах измерения;

6. График не следует перегружать материалом, также он должен быть достаточно наглядным;

7. Факторные признаки размещаются на горизонтальной шкале графика и их значения читаются слева направо, результативный признак - по вертикальной шкале и читается снизу вверх.

ГЛАВА 4 СТАТИСТИЧЕСКИЕ ПОКАЗАТЕЛИ

4.1 Абсолютные показатели и их значение

Статистический показатель – это обобщающая количественная характеристика какого-то свойства статистической совокупности или ее части. *Этим он отличается от признака (свойства, присущего единице совокупности) (средний возраст студента в группе на начало учебного года – статистический показатель, а возраст конкретного студента данной группы - признак)*.

Выделяют следующие атрибуты статистического показателя:

1. **Качественная сторона** (основание или содержание), которая отражает сущности изучаемого свойства статистической совокупности без указания места, времени и возможности определения числового значения. Определяется она понятиями, входящими в наименование показателя, и связана с функцией, которую выполняет показатель.

Основные **функции** статистического показателя:

плановая (показатель в плане, норматив для данного показателя);

отчетная (показатель в отчете);

прогностическая, т.е. роль статистических показателей в предвидении будущего;

оценочная – заключается в том, что на основе статистических показателей люди, общество, государство оценивают деятельность предприятий, организаций, трудовых коллективов, правительства и т.д.;

познавательная и информационная;

рекламно-пропагандистская.

Понятия, входящие в наименование показателя, можно разделить на 2 группы:

а) понятия чисто статистические;

б) понятия, являющиеся предметом изучения других областей знания.

Например, в показателе: средняя ожидаемая продолжительность жизни – содержание определяется понятиями:

средняя величина (статистическое понятие) и продолжительность жизни (понятие демографии).

2. Количественная сторона: методология расчета (формула), величина (числовое значение) и единица измерения.

3. Пространственные границы: территориальные, отраслевые и иные границы статистического показателя.

4. Границы во времени: интервал или момент.

Статистический показатель является инструментом познания изучаемых явлений и процессов. *С его помощью обогащается и уточняется понятие явления. Однако при построении показателей почти всегда приходится дополнительно упрощать, схематизировать реальные явления, а потому статистические показатели лишь с известной степенью приближения отражают объективную реальность. Они дают приближенное, неточное и неполное отображение свойств изучаемого объекта, доступное при имеющемся уровне знаний и возможностях учета, измерения, сбора и передачи информации.*

Методика исчисления статистических показателей постоянно совершенствуется: от исчисления некоторых показателей отказываются, в то же время появляются новые, более точные.

Абсолютные показатели. Данные показатели отражают физические размеры изучаемых статистикой процессов и явлений, а именно их массу, площадь, объем, протяженность, временные характеристики, а также могут представлять объем совокупности, т.е. число составляющих единиц. К абсолютным показателям, например, относятся площадь территории страны, объем промышленного производства, эксплуатационная длина железнодорожных путей сообщения, число предприятия отрасли 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, километры, мили, литры, баррели, штуки и т.д.

В группу натуральных также входят условно-натуральные измерители, которые используются в тех случаях, когда какой-либо продукт имеет несколько разновидностей и общий объем можно определить только исходя из общего для всех разновидностей потребительского свойства.

В условиях рыночной экономики особое значение имеют *стоимостные единицы измерения*, позволяющие дать денежную оценку социально-экономическим объектам и явлениям.

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.

4.2 Сущность относительных величин и единицы измерения. Виды относительных величин

В статистической практике для аналитических целей широко применяются относительные показатели.

Относительные показатели. Они представляют собой результат деления одного абсолютного показателя на другой и выражают 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.

При расчете относительного показателя абсолютный показатель, находящийся в числителе получаемого отношения, называется текущим, или сравнительным. Показатель же, с которым производится сравнение и который находится в знаменателе, называется основанием, или базой сравнения. Таким образом, рассчитываемый относительный показатель указывает, во сколько раз сравниваемый абсолютный показатель больше базисного, или какую долю он составляет от базисного показателя, или сколько единиц первого приходится на 1, 100, 1000 и т.д. единиц второго. Относительный показатель может выражаться в коэффициентах, процентах, промилле, продецимилле или быть именованным числом.

Все используемые на практике относительные статистические показатели можно подразделить на следующие виды: показатели динамики, плана, реализации плана, структуры, координации, интенсивности, уровня экономического развития, сравнения.

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

$$ОПД = \frac{Текущий\ уровень}{Предшествующий\ или\ базисный\ уровень}$$

Рассчитанная таким образом величина показывает, во сколько раз текущий уровень превышает предшествующий (базисный) или какую долю от последнего он 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.

В таблице 33 приведено производство сахара-песка в Пензенской области.

Таблица 33 – Производство сахара-песка в Пензенской области, тыс. т

2005 г.	2006 г.	2007 г.	2008 г.	2009 г.
287	212	224	207	181

Рассчитаем относительные показатели динамики с переменной и постоянной базой сравнения и отразим результаты в таблице 34.

Таблица 34 – Расчет показателей динамики производства сахара-песка в Пензенской области

Переменная база сравнения (цепные показатели)	Постоянная база сравнения (базисные показатели)
$\frac{212}{287}\times 100 = 73{,}9\ \%$ $\frac{224}{212}\times 100 = 105{,}7\ \%$ $\frac{207}{224}\times 100 = 92{,}4\ \%$ $\frac{181}{207}\times 100 = 87{,}4\ \%$	$\frac{212}{287}\times 100 = 73{,}9\ \%$ $\frac{224}{287}\times 100 = 78{,}0\ \%$ $\frac{207}{287}\times 100 = 72{,}1\ \%$ $\frac{181}{287}\times 100 = 63{,}1\ \%$

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 получим:

$$0{,}739 \times 1{,}057 \times 0{,}924 \times 0{,}874 = 0{,}631$$

Относительные показатели выполнения планового задания. Все субъекты финансово-хозяйственной деятельности (от малых предприятий и до крупных корпораций) в той или иной степени осуществляют как текущее, так и стратегическое планирование, а также сравнивают реально достигнутые результаты с ранее намеченными. Для этой цели используются относительные показатели выполнения плана (ОПВП) и планового задания (ОППЗ).

$$ОПВП = \frac{Фактический\ уровень\ текущего\ периода}{Плановый\ уровень\ на\ этот\ же\ период}$$

$$ОППЗ = \frac{Плановый\ уровень\ на\ текущий\ период}{Фактически\ достигнутый\ уровень\ в\ базисном\ периоде}$$

Первый показатель отражает фактический объем производства в процентах или коэффициентах по сравнению с плановым уровнем, второй – какой объем производства был запланирован от уровня, фактически произведенного в сравниваемом периоде. Например, оборот торговой фирмы в 2009 г. составил 2,0 млн. руб., на следующий год руководство фирмы считает реальным довести оборот до 2,8 млн. рублей. Фактический оборот фирмы в 2010 г. составил 2,6 млн. рублей. Тогда относительный показатель выполнения плана составил 92,9 % $\left(\frac{2,6}{2,8}\times 100\right)$, а относительный показатель планового задания 140 % $\left(\frac{2,8}{2,0}\times 100\right)$.

Между рассмотренными относительными величинами существует следующая зависимость:

$$ОПД = ОПВД \times ОППЗ$$

Относительный показатель структуры (ОПС) 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.

В таблице 35 представлены исходные данные и расчет относительных показателей структуры.

Таблица 35 – Поголовье крупного рогатого скота по формам собственности в Пензенской области (на 1 января 2010 г.)

Показатель	Тыс. гол.	% к итогу

Все формы собственности	297,9	100
в том числе: государственная	9,0	$\frac{9,0}{297,9}\times 100 = 3,02$
муниципальная	2,4	$\frac{2,4}{297,9}\times 100 = 0,81$
частная	286,5	$\frac{286,5}{297,2}\times 100 = 96,17$

Относительный показатель координации (*ОПК*) представляет собой отношение одной части к другой части этой же совокупности:

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

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 одну единицу (иногда – на 100, 1000 и т.д. единиц) базисной структурной части.

По данным предыдущей таблицы вычислим, что на каждую голову скота, находящуюся в государственной форме собственности приходилось 32 головы скота частной формы собственности (286,5: 9).

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

$$ОПИ = \frac{Показатель,\ характеризующий\ явление\ А}{\begin{array}{c}Показатель,\ характеризующий \\ среду\ распространения\ явлений\ А\end{array}}$$

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

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 км^2.

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. Так как объемные показатели производства по своей природе являются интервальными, а показатель численности населения – моментным, в расчете используют среднюю за период численности населения (например, среднегодовую).

К относительным показателям уровня экономического развития относятся показатели, характеризующие производство угля, стали, нефти, газа на душу населения, производство зерна, молока, мяса на душу населения и т.п.

Относительный показатель сравнения (*ОПС*) представляет собой соотношение одного и того же абсолютного показателя, характеризующего разные объекты (предприятия, фирмы, районы, области, страны и т.п.):

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

Например, имеются данные о среднедушевых ежемесячных доходах населения за 2009 г., руб.: Пензенская область – 11535, Самарская область – 18175, Саратовская область – 10356, Ульяновская область – 7450. Таким образом, среднедушевой доход населения в Самарской области превышал аналогичный показатель по Пензенской области – в 1,58 раза $\left(\frac{18175}{11535}\right)$, в Саратовской области – в 1,76 раза $\left(\frac{18175}{10356}\right)$, в Ульяновской области – в 2,44 раза $\left(\frac{18175}{7450}\right)$.

4.3 Понятие средних величин и их значение в статистике

Наиболее распространенной формой статистических показателей, используемой в социально-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 незаменимыми в анализе явлений и процессов общественной жизни.

Важнейшее свойство средней заключается в том, что она отражает то общее, что присуще всем единицам исследуемой совокупности. Значения признака отдельных единиц совокупности варьируют под влиянием множества факторов, среди которых могут быть как основные, так и случайные.

Сущность средней в том и заключается, что в ней взаимопогашаются те отклонения значений признака, которые обусловлены 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.

Для вычисления средних величин используется формула степенной средней:

$$\bar{x} = \sqrt[k]{\frac{\sum x_i^k \times f_i}{\sum f_i}},$$

где $\bar{x}$ – средняя величина исследуемого явления;
x_i – i-й вариант осредняемого признака ($i = 1n$);
f_i – вес i-го варианта.

Формулы средних величин, используемых в статистике, представлены в таблице 36.

Для вычисления среднего значения качественного признака (средняя урожайность, средняя себестоимость и т.д.), как правило используется средняя арифметическая или средняя гармоническая. Чтобы правильно выбрать формулу расчета, необходимо составить логическую формулу средней или исходное соотношение средней (ИСС):

$$ИСС = \frac{Объем\ осредняемого\ признака}{Объем\ совокупности}$$

Таблица 36 – Формулы средних величин, используемых в статистике

Значение k	Наименование средней	Формула средней	
		простая	взвешенная
-1	Средняя гармоническая	$\bar{x} = \frac{n}{\sum \frac{1}{x}}$	$\bar{x} = \frac{\sum f}{\sum \frac{f}{x}}$
0	Средняя геометрическая	$\bar{x} = \sqrt[n]{x_1 \times x_2 \times x_3 \times \ldots x_n}$	$\bar{x} = \sqrt[\sum f]{x_1^{f1} \times x_2^{f2} \times x_3^{f3} \times \ldots x_n^{fn}}$
1	Средняя арифметическая	$\bar{x} = \frac{\sum x}{n}$	$\bar{x} = \frac{\sum xf}{\sum f}$
2	Средняя квадратическая	$\bar{x} = \frac{\sum x^2}{n}$	$\bar{x} = \sqrt{\frac{\sum x^2 f}{\sum f}}$

Свойство степенных средних возрастать с повышением показателя степени определяющей функции называется в статистике ***правилом мажорантности средних***.

Вид средней выбирается в каждом отдельном случае путем конкретного анализа изучаемой совокупности, он определяется материальным содержанием изучаемого явления, а также принципами суммирования и взвешивания. Характер имеющихся данных определяет существование только одного истинного среднего значения показателя.

Кроме степенных средних в статистической практике используются структурные средние величины, в качестве которых рассматриваются ***мода и медиана***.

4.4 Средняя арифметическая, ее свойства и способы вычисления

Средняя арифметическая, представляет собой частное от деления суммы индивидуальных значений признака на их количество. Средняя арифметическая бывает ***простой и взвешенной***.

Средняя арифметическая простая применяется в тех случаях, когда каждое индивидуальное значение признака встречается один или одинаковое число раз, то есть когда средняя рассчитывается по группировочным единицам совокупности.

Расчет средней арифметической простой:

$$\overline{x_{ap}} = \frac{x_1 + x_2 + ... + x_n}{n} = \frac{\sum x}{n}$$

где $x_1, x_2, ...x_n$ – индивидуальные значения варьирующего признака; n – число единиц совокупности.

Пример. Имеется информация о стаже пяти рабочих, при этом стаж первого рабочего составил 5 лет, второго – 7, третьего – 4, четвертого – 10, пятого – 12 лет.

Определить средний стаж работы.

Решение. Поскольку в исходных данных значение каждого варианта встречалось только один раз, для определения среднего стажа одного рабочего следует применить формулу простой средней арифметической. В нашем примере средний стаж будет равен:

$$\overline{x_{ap}} = \frac{5+7+4+10+12}{5} = 7{,}6$$

Среднюю арифметическую взвешенную рассчитывают в тех случаях, когда отдельные значения исследуемой совокупности встречаются не один, а много, причем неодинаковое число раз, то есть представляют собой ряд распределения.

Если индивидуальные значения признака (варианты) обозначить $x_1, x_2, ...x_n$, а числа, показывающие, сколько раз повторятся варианты (частоты), – $f_1, f_2, f_3, ..., f_n$, то средняя арифметическая взвешенная будет равна:

$$\overline{x_{ap}} = \frac{x_1 f_1 + x_2 f_2 + ... + x_n f_n}{f_1 + f_2 + ... + f_n} = \frac{\sum x_i f_i}{\sum f_i}$$

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

Пример. Рассчитать среднюю заработную плату работников в бригаде из 20 человек, оплата труда которых варьируется от 2800 до 4400 руб., где x_i – варианты осредняемого признака, f_i – частота, показывающая, сколько раз встречается i-е значение в совокупности.

Заработная плата работников бригады

Заработная плата, руб. (x_i)	2800	3200	3600	4000	4400	Итого
Число работников, чел. (f_i)	2	11	5	1	1	20

Решение. Применяя формулу средней арифметической взвешенной, получаем:

$$\overline{x_{ap}} = \frac{2800*2 + 3200*11 + 3600*5 + 4000*1 + 4400*1}{2 + 11 + 5 + 1 + 1} = 3225 \text{руб.}$$

В отдельных случаях веса могут быть представлены не абсолютными величинами, а относительными (в процентах или долях единицы). Тогда формула средней арифметической взвешенной будет иметь вид:

$$\overline{x_{ap}} = \frac{\sum xd}{\sum d}$$

где $d = \frac{f}{\sum f}$ – частность, то есть доля каждой частоты в общей сумме всех частот.

Если частоты подсчитывают в долях (коэффициентах), то $\sum d = 1$ и формула средней арифметической взвешенной имеет вид:

$$\overline{x_{ap}} = \sum xd$$

Вычисление средней арифметической интервального ряда

Иногда среднюю арифметическую величину исчисляют по данным интервального вариационного ряда (когда варианты признака, по которому определяется средняя, представлены в виде интервалов «от – до»). В этом случае в качестве значений признаков в группах принимают середины интервалов, в результате чего образуется дискретный ряд.

Пример. Определить средний размер капитальных затрат на одно предприятие по следующим данным:

Исходные данные		**Расчетные значения**	
Группы предприятий по размеру капитальных затрат, тыс.руб.	**Число предприятий, *f***	**Середина интервала, тыс.руб., *x***	***x*f***
До 3000	5	2900	14500
3000 – 3200	15	3100	46500
3200 – 3400	20	3300	66000
3400 – 3600	30	3500	105000
3600 – 3800	16	3700	59200
3800 и более	14	3900	54600
Итого	100	-	345800

Решение. Для исчисления средней в интервальном ряду необходимо прежде всего перейти от интервального ряда к дискретному путем замены интервальных значений их средними значениями (простая средняя между верхней и нижней границами каждого интервала) (см. расчет в табл., гр. 3).

Если имеются интервалы с так называемыми открытыми границами (в нашем примере первый интервал – до 3000 тыс. руб. или последний интервал – 3800 тыс. руб.), то для расчета средней условно определяют неизвестную границу интервала. Обычно в этих условиях берут значение последующего интервала (для первого) или предыдущего (для последнего).

После того как найдено среднее значение интервалов, расчет производится по средней арифметической взвешенной.

В нашем примере средний размер капитальных затрат на одно предприятие составит:

$$\overline{x_{ap}} = \frac{\sum xf}{\sum f} = \frac{345800}{100} = 3458 \text{тыс.руб.}$$

Необходимо помнить, что средняя арифметическая интервального ряда менее точна, чем средняя арифметическая, исчисленная из конкретных вариантов, потому что при исчислении середин интервалов допускается некоторая условность.

Средняя арифметическая обладает рядом свойств:

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

Если $x_i = y_i + z_i$, то

$$\bar{x} = \frac{\sum x_i}{n} = \frac{\sum (y_i + z_i)}{n} = \frac{\sum y_i}{n} + \frac{\sum z_i}{n} = \bar{y} + \bar{x}$$

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

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

$\sum (x - \bar{x}) = 0$, потому что

$$\sum (x - \bar{x}) = \sum x - \sum \bar{x} = \sum x - n\bar{x} = \sum x - n\frac{\sum x}{n}$$

Это правило показывает, что средняя является равнодействующей.

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

$$\overline{x} = \frac{\sum x_i f_i}{\sum f_i} = \frac{\sum (x_i \mp a)}{\sum f_i} \pm a$$

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

$$\overline{x} = \frac{\sum x_i f_i}{\sum f_i} = \frac{\sum \frac{x_i}{A} f_i}{\sum f_i} \times A = \frac{\sum A x_i f_i}{\sum f_i} \div A$$

5. Если все частоты ряда разделить или умножить на одно и же число с/, то средняя не изменится:

$$\bar{x} = \frac{\sum x_i y_i}{\sum y_i} = \frac{\sum x_i \frac{f_i}{d}}{\sum \frac{f_i}{d}} = \frac{\sum x_i f_i d}{\sum f_i d}$$

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

4.5 Средняя гармоническая

В статистической практике бывают случай, когда при вычислении средней имеются данные об индивидуальных значениях признака (x) и его общем объеме в совокупности ($W = xf$), но неизвестны частоты (f). В таком случае среднее значение признака вычисляется по формуле *средней гармонической*.

Средняя гармоническая – представляет собой величину, обратную средней арифметической из обратных значений вариант.

$$\bar{x}_{гарм.взв.} = \frac{\sum x_i f_i}{\sum f_i} = \frac{\sum W}{\sum \frac{W}{x_i}}$$

Пример. По данным таблицы требуется определить среднюю рентабельность капитала по двум акционерным обществам в целом.

№ АО	**Рентабельность акционерного капитала, % (x)**	**Получено прибыли, тыс.руб. ($W = xf$)**	**Акционерный капитал, тыс.руб. (W/x)**
1	40	6000	15000
2	35	3500	10000
Итого	-	9500	25000

Решение. Основой выбора формы средней является реальное содержание определяемого показателя:

Рентабельность, % = (Прибыль / Акционерный капитал) * 100

Средняя рентабельность равна отношению полученной предприятиями прибыли к сумме их акционерного капитала. В данном примере отсутствуют прямые данные об акционерном капитале, но его сумму можно определить косвенным путем, разделив полученную прибыль (*M*) на рентабельность капитала (*x*).

Средняя рентабельность акционерного капитала будет равна

$$\bar{x} = \frac{\sum M}{\sum \frac{M}{x}} = \frac{6000+3500}{\frac{6000}{0{,}4}+\frac{3500}{0{,}35}} = \frac{9500}{25000} = 0{,}38 \text{ или } 38\ \%.$$

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

$$\bar{x}_{гарм.} = \frac{\sum W}{\sum \frac{W}{x_i}} = \frac{1+1+\ldots+1}{\frac{1}{x_1}+\frac{1}{x_2}+\frac{1}{x_3}+\ldots+\frac{1}{x_n}} = \frac{n}{\sum \frac{1}{x}}$$

где x – отдельные варианты; n – их число

Пример. Фактический объем реализации продукции ОАО «Лада» за месяц составил 14 млн руб. При проверке было выявлено, что объем неучтенной продукции составил 15 %. Фактический объем реализации продукции ОАО «Мир» также составил 14 млн руб., а объем выявленной в результате проверки неучтенной продукции 25 %.

Определить средний процент неучтенной продукции для обоих предприятий.

Решение. Для решения этой задачи необходимо применить формулу простой средней гармонической:

$$\bar{x}_{гарм.} = \frac{2}{\frac{1}{15}+\frac{1}{25}} = 18{,}8\ \%$$

то есть средний процент неучтенной продукции для обоих предприятий составляет 18,8 %.

4.6 Показатели вариации, их значение и виды

Различия индивидуальных значений признака внутри изучаемой совокупности называются вариацией признака. Она возникает в результате того, что индивидуальные значения признака складываются под совокупным влиянием разнообразных факторов, которые по-разному сочетаются в каждом отдельном случае.

При характеристике колеблемости признака применяют систему абсолютных и относительных показателей.

К *абсолютным показателям вариации* относятся:

1. размах вариации R;
2. среднее линейное отклонение d;
3. дисперсия σ^2;
4. среднее квадратическое отклонение $\sigma = \sqrt{\sigma^2}$.

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

К *относительным показателям вариации* относятся:

1. коэффициент осцилляции;
2. линейный коэффициент вариации;
3. простой коэффициент вариации.

Эти показатели выражаются в процентах или относительных величинах.

Размах вариации – наиболее простой измеритель вариации и представляет собой разность между наибольшим и наименьшим значением признака:

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

где x_{max} – наибольшее значение признака; x_{min} – минимальное значение признака.

Величина R показывает, в каких пределах колеблется размер признака, образующего ряд распределения. С его помощью определяют допустимые размеры колебаний, сравнивают их с установленными.

Среднее линейное отклонение представляет собой среднюю из абсолютных значений отклонений отдельных вариант от их средней. Так как алгебраическая сумма отклонений индивидуальных значений признака от средней равна нулю (второе свойство средней арифметической), при исчислении среднего линейного отклонения принимаются во внимание только абсолютные значения отклонений без учета знаков («+» или «-»). Среднее линейное отклонение рассчитывается по формуле средней арифметической простой:

$$\bar{d} = \frac{\sum |x - \bar{x}|}{n}$$

или средней арифметической взвешенной:

$$\bar{d} = \frac{\sum |x - \bar{x}| f}{\sum f}$$

Дисперсией называется средний квадрат отклонений индивидуальных значений признака от их средней величины. Дисперсия равна:

$$\sigma^2 = \frac{\sum (x - \bar{x})^2 f}{\sum f}$$

При равенстве весов или когда они равны 1,

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

Свойства дисперсии:

1) если все значения признака уменьшить или увеличить на одну и ту же постоянную величину *A*, то дисперсия от этого не изменится;

2) если все значения признака уменьшить или увеличить в одно и то же число (i) раз, то дисперсия соответственно уменьшится или увеличится в i^2 раз.

Дисперсия в вариационных рядах с равными интервалами по способу моментов:

$$\sigma^2 = i^2 \left(m_2 - m_1^2 \right) = i^2 \left(\frac{\sum x_1^2 f}{\sum f} - \left(\frac{\sum x_1 f}{\sum f} \right)^2 \right)$$

где i – величина интервала; $x_1 = \frac{x - A}{i}$ – новые (преобразованные) значения вариант (А – условный ноль, в качестве которого удобно использовать середину интервала, обладающего наибольшей частотой);

$m_2 = \frac{\sum x_1^2 f}{\sum f}$ – момент второго порядка;

$m_1 = \left(\frac{\sum x_1 f}{\sum f} \right)^2$ – квадрат момента первого порядка.

Дисперсия имеет большое значение в экономическом анализе. В математической статистике играет важную роль для характеристики качества статистических оценок.

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

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

а при равенстве весов или когда они равны 1,

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

Исчисление дисперсии и среднего квадратического отклонения позволяет устранить недостаток среднего линейного отклонения. *Ведь любое число, положительное или отрицательное, возведенное в квадрат, будет числом положительным.*

Среднее линейное отклонение, дисперсия и среднее квадратическое отклонение дают представление об абсолютной величине колеблемости.

Для сравнения колебаний разнородных явлений, разных по своему характеру и размерам признаков, используется относительный показатель вариации, так называемый коэффициент вариации.

Коэффициент вариации (V) представляет собой процентное отношение среднего квадратического отклонения к средней арифметической и рассчитывается по формуле:

$$V = \frac{\sigma * 100}{\bar{x}}$$

Выражая коэффициент вариации в процентах, различные абсолютные среднеквадратические отклонения приводят к одному основанию и дают возможность сравнивать, оценивать колеблемость величин различных признаков.

При помощи коэффициента вариации возможно, например, сравнение размера колеблемости производительности труда групп рабочих, занятых производством различных видов продукции, размера колеблемости урожаев различных сельскохозяйственных культур и т.д. Чем больше его величина, тем больше разброс значений вокруг средней, тем менее однородна совокупность по своему составу и тем менее представительна средняя. Коэффициент вариации важен и в тех случаях, когда нужно сравнивать средние квадратические отклонения, выраженные в разных единицах измерения.

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

$$V_d = \frac{\bar{d} * 100}{\bar{x}}$$

Отношение размаха вариации к средней арифметической в процентах называется коэффициентом осцилляции:

$$V_R = \frac{R * 100}{\bar{x}}$$

Самый распространенный относительный показатель колеблемости — коэффициент вариации.

ГЛАВА 6 РЯДЫ ДИНАМИКИ

6.1 Ряды динамики и их виды

Ряд динамики – это числовые значения статистических показателей, изменяющихся во времени и расположенных в хронологической последовательности.

Ряд динамики включает два обязательных элемента:

– период времени, за который или по состоянию на который приводятся цифровые значения (показатель времени *t*);
– конкретные числовые значения показателя, характеризующие изучаемый объект или явление (уровни ряда *y*).

Существуют различные виды рядов динамики. Их можно классифицировать по:

– форме представления уровней – ряды абсолютных, относительных или средних величин; (*спросить про относительные и абсолютные показатели*)
– интервалам времени или расстоянию между уровнями – равномерные и неравномерные (полные и неполные);
– по наличию основной тенденции изучаемого процесса – стационарные и нестационарные ряды;
– показателю времени – моментные и интервальные.

Моментные ряды динамики отображают состояние изучаемого явления на определенную дату времени. Например, численность населения, тыс. чел.:

31.12.99.	*31.12.00.*	*31.12.01.*	*31.12.02.*	*31.12.03.*
897	*890,7*	*883,5*	*860,3*	*850*

Сумма уровней моментного ряда не имеет реального содержания, а в основной части представляет собой повторный счет.

Интервальный ряд динамики отображает итоги развития изучаемых явлений за отдельные периоды или интервалы времени. Например, объем валового внутреннего продукта (ВВП) в РФ, млрд руб.:

1999 г.	*2000 г.*	*2001г.*	*2002 г.*	*2003 г.*
4823,2	*7305,6*	*8943,6*	*10834,2*	*13285,2*

Если уровни в интервальном ряду выражены абсолютными показателями, то их можно суммировать или дробить во времени, получая новые числовые значения объема явления, относящиеся к более крупным или мелким промежуткам времени. Сумма уровней интервального ряда дает вполне реальную статистическую величину, так называемые накопленные итоги, например, общий объем налоговых поступлений в госбюджет, общее количество выпускников вузов.

Важнейшим условием правильного построения рядов динамики, получения правильных выводов при анализе и прогнозировании его уровней является сопоставимость уровней, образующих ряд. Статистические данные должны быть сопоставимы: по кругу охватываемых объектов, времени регистрации, территории, идеологии расчета и ценам.

Сопоставимость по кругу охватываемых явлений означает сравнение совокупностей с равным числом элементов, которые должны быть однородны по экономическому содержанию и границам объекта. Несопоставимость может возникнуть в результате перехода ряда объектов из одного подчинения в другое.

Сопоставимость по времени регистрации для интервальных рядов обеспечивается равенством периодов времени, за которые получают данные. Для приведения рядов динамики к сопоставимому виду выделяют среднедневные показатели по декадам, кварталам, месяцам, которые затем сравнивают. Для моментных рядов динамики показатели следует проводить на одну и ту же дату.

Сопоставимость по территории предполагает одни и те же территориальные границы. Данные по странам и регионам, границы которых изменились, должны быть пересчитаны в старых пределах.

Сопоставимость по методологии расчетов характеризуется тем, что при определении уровней динамического ряда необходимо использовать единую методологию их расчета.

Сопоставимость по ценам. При приведении к сопоставимому виду продукции, которая была измерена в стоимостных показателях, трудность заключается в том, что, во-первых, с течением времени происходит непрерывное изменение цен, а во-вторых, существует несколько видов цен.

Прежде чем анализировать ряд динамики, необходимо привести уровни ряда динамики к сопоставимому виду, для чего прибегают к приему «смыкание рядов динамики» путем их приведения к одному ряду. Смыкание может быть произведено *двумя способами*.

Первый (абсолютный способ) – данные за предыдущие периоды умножаются на коэффициент перехода или приведения, равный отношению новых и прежних показателей «переломного» момента времени, когда произошло пересечение показателей в новых и старых границах или изменилось условие формирования уровней ряда.

Второй (относительный способ) – уровень переходного периода принимается для второй части ряда за 100%, и от этого уровня определяются показатели вперед и назад. При этом получается сопоставимый ряд относительных величин.

6.2 Анализ уровней динамического ряда

Показатели анализа ряда динамики могут рассчитываться на постоянной и переменной базах сравнения. При этом принято называть сравниваемый уровень отчетным, а уровень, с которым производится сравнение, – базисным. Для расчета показателей на постоянной базе каждый уровень сравнивается с одним и тем же базисным уровнем. Рассчитанные при этом показатели называются базисными. Для расчета показателей на переменной базе каждый последующий уровень сравнивается с предыдущим, а показатели называются цепными.

1. ***Абсолютный прирост (абсолютное изменение)*** определяется как разность между двумя уровнями динамического ряда и показывает, на сколько единиц данный уровень ряда превышает уровень другого периода. Один и тот же по величине абсолютный прирост может означать разную интенсивность изменения.

а) базисный:

$$\Delta_y^{б} = y_i - y_0$$

б) цепной:

$$\Delta_y^{ц} = y_i - y_{i-1}$$

где

y_i – уровень сравниваемого периода;

y_{i-1} – уровень предшествующего периода;

y_0 – уровень базисного периода.

Цепные и базисные абсолютные приросты связаны между собой определенным правилом: сумма последовательных цепных абсолютных приростов равна последнему базисному:

$$\sum \Delta_y^{ц} = \Delta_y^{б}$$

По знаку абсолютного прироста можно сделать вывод о характере развития явления:

$\Delta_y > 0$ – рост, $\Delta_y < 0$ – спад, $\Delta_y = 0$ – стабильность.

2. ***Темп роста*** определяется как отношение двух сравниваемых уровней и показывает, во сколько раз данный уровень превышает уровень базисного периода.

а) базисный:

$$T_p^{б} = \frac{y_i}{y_0} \times 100\% = K_p^{б} \times 100\%$$

б) цепной:

$$T_p^{ц} = \frac{y_i}{y_{i-1}} \times 100\% = K_p^{ц} \times 100\%$$

Темп роста представляет всегда положительное число.

Коэффициент роста (сила роста):

а) базисный:

$$K_p^{б} = \frac{y_i}{y_0}$$

б) цепной:

$$K_p^{ц} = \frac{y_i}{y_{i-1}}$$

3. ***Темп прироста или темп сокращения*** (темп изменения уровней) показывает, на сколько процентов уровень данного периода больше или меньше определенного уровня, характеризует относительную скорость изменения уровня ряда в единицу времени.

Можно рассчитать двумя способами:

1) как отношение абсолютного прироста к уровню:

а) базисный:

$$T_{ПР}^{б} = \frac{y_i - y_0}{y_0} \times 100\% = \frac{\Delta_y^{б}}{y_0} \times 100\%$$

б) цепной:

$$T_{ПР}^{ц} = \frac{y_i - y_{i-1}}{y_{i-1}} \times 100\% = \frac{\Delta_y^ц}{y_{i-1}} \times 100\%$$

2) как разность между темпом роста и 100%:

$$T_{ПР} = T_p - 100\%$$

Между цепными и базисными показателями изменения уровней ряда существует следующая взаимосвязь:

сумма цепных абсолютных приростов равна базисному приросту;

произведение цепных коэффициентов роста равно базисному;

деление рядом стоящих базисных коэффициентов роста друг на друга равно цепным коэффициентам роста.

4. Темп наращивания (пункт роста) рассчитывается делением цепных абсолютных приростов на уровень, принятый за постоянную базу сравнения:

$$T_н = \frac{y_i - y_{i-1}}{y_0} \times 100\% = \frac{\Delta_y^ц}{y_0} \times 100\%$$

5. ***Абсолютное значение одного процента прироста***. Чтобы знать, что скрывается за каждым процентом прироста, рассчитывается абсолютное значение 1 % прироста как отношение абсолютного прироста уровня за интервал времени к темпу прироста за тот же промежуток времени:

$$A\% = \frac{y_i - y_{i-1}}{\frac{y_i - y_{i-1}}{y_{i-1}}} \times 100\% = 0{,}01 y_{i-1}$$

Иными словами, абсолютное значение 1 % прироста в данном периоде – сотая часть достигнутого уровня в предыдущем периоде. В связи с этим расчет абсолютного значения 1 % прироста базисным методом не имеет смысла, ибо для каждого периода это будет одна и та же величина – сотая часть уровня базисного периода.

Если систематически растут цепные темпы роста, то ряд развивается относительным ускорением. Относительное ускорение можно определить как разность следующих друг за другом темпов роста или прироста; полученная величина выражается в процентных пунктах (п.п.).

6.3 Средние показатели по рядам динамики

Для обобщения характеристики динамики исследуемого явления за ряд периодов определяют различного рода средние показатели, среди которых можно выделить:

средний уровень ряда;

средний абсолютный прирост;

средние темпы роста и прироста.

Способы расчета среднего уровня различаются и зависят от характеристики ряда.

Рассмотрим две категории средних показателей рядов динамики.

1. Средние показатели изменения уровня ряда:

а) средний абсолютный прирост (средняя скорость роста):

цепной $\overline{\Delta} = \frac{\sum \Delta_y^{ц}}{n-1}$

базисный $\overline{\Delta} = \frac{y_n - y_1}{n-1}$

где n – количество уровней ряда; y_n – самое последнее значение уровня ряда; y_1 – самое первое значение;

б) *средняя сила роста (коэффициент роста)*

$$\overline{K}_p = \sqrt[n-1]{K^{ц}_{p_2} \times K^{ц}_{p_3} \times \ldots \times K^{ц}_{p_n}} = \sqrt{K^{б}_{p_n}}$$

Когда приходится вести расчет средних темпов роста по периодам различной продолжительности (неравностоящие ряды динамики), то пользуются средними геометрическими взвешенными по продолжительности периодов:

$$\overline{K}_p = \sqrt[\sum T]{(K^{ц}_{p_2})^{T_1} \times (K^{ц}_{p_3})^{T_2} \times \ldots \times (K^{ц}_{p_n})^{T_{n-1}}}$$

где T_i – продолжительность промежутков времени между моментами i $(i+1)$; $\sum T$ - весь рассматриваемый период времени от t_1 до t_n.

в) *средний темп роста*:

$\overline{T}_p = \overline{K}_p \times 100\%$

Вычитанием 100% из среднего прироста получают соответствующий *средний темп прироста*.

$$\overline{T}_{ПР} = \overline{T}_{p} - 100\%$$

2. Средние уровни ряда зависят от вида временного ряда:

а) по интервальному динамическому ряду из абсолютных величин с равными интервалами средний уровень определяется по средней арифметической простой из уровней ряда:

$$\bar{y} = \frac{\sum y_i}{n}$$

б) для интервального ряда с разными промежутками времени между уровнями используется формула средней арифметической взвешенной, где в качестве весовых коэффициентов используется продолжительность интервалов времени между уровнями:

$$\bar{y} = \frac{\sum y_i t_i}{\sum t_i}$$

где t_i – количество дней между смежными датами;

в) для моментного равно отстающего ряда используется формула средней хронологической:

$$\bar{y}_{хрон} = \frac{\frac{1}{2}y_1 + y_2 + ... + y_{n-1} + \frac{1}{2}y_n}{n-1}$$

Данная формула используется, например, для расчета среднегодовой стоимости основных фондов, товарных запасов и др.;

г) для моментного ряда динамики с неравно отстающими во времени уровнями используется формула средней хронологической взвешенной:

$$\bar{y}_{хрон} = \frac{(y_1 + y_2)t_1 + (y_2 + y_3)t_2 + ... + (y_{n-1} + y_n)t_{n-1}}{\sum t_1}$$

6.4 Методы выявления тенденций в динамических рядах

Одна из важнейших задач статистики – определение в рядах динамики общей тенденции развития.

Основной тенденцией развития называется плавное и устойчивое изменение уровня явления во времени, свободное от случайных колебаний. Задача состоит в выявлении общей тенденции в изменении уровней ряда, освобожденной от действия различных факторов.

Изучение тренда включает два основных этапа:

- ряд динамики проверяется на наличие тренда;
- производится выравнивание временного ряда и непосредственное выделение тренда с экстраполяцией полученных результатов.

С этой целью ряды динамики подвергаются обработке методами укрупнения интервалов, скользящей средней и аналитического выравнивания:

1. Метод укрупнения интервалов.

Одним из наиболее элементарных способов изучения общей тенденции в ряду динамики является укрупнение интервалов. Этот способ основан на укрупнении периодов, к которым относятся уровни ряда динамики. Например, преобразование месячных периодов в квартальные, квартальных в годовые и т. д.

2. Метод скользящей средней.

Выявление общей тенденции ряда динамики можно произвести путем сглаживания ряда динамики с помощью скользящей средней.

Скользящая средняя – подвижная динамическая средняя, которая рассчитывается по ряду при последовательном передвижении на один интервал, то есть сначала вычисляют средний уровень из определенного числа первых по порядку уровней ряда, затем – средний уровень из такого же числа членов, начиная со второго. Таким образом, средняя как бы скользит по ряду динамики от его начала к концу, каждый раз отбрасывая один уровень в начале и добавляя один следующий.

Скользящая средняя обладает достаточной гибкостью, но недостатком метода является укорачивание сглаженного ряда по сравнению с фактическим, что ведет к потере информации. Кроме того, скользящая средняя не дает аналитического выражения тренда.

Период скользящей может быть четным и нечетным. Практически удобнее использовать нечетный период, так как в этом случае скользящая средняя будет отнесена к середине периода скольжения. Скользящие средние с продолжительностью периода, равной 3, следующие:

$$\bar{y}_1 = \frac{y_1 + y_2 + y_3}{3};\ \bar{y}_2 = \frac{y_2 + y_3 + y_4}{3};\ \bar{y}_3 = \frac{y_3 + y_4 + y_5}{3} \text{ и т.д.}$$

Полученные средние записываются к соответствующему срединному интервалу.

Особенность сглаживания по четному числу уровней состоит в том, что каждая из исчисленных (например, четырехчленных) средних относится к соответствующим промежуткам между смежными периодами. Для получения значений сглаженных уровней соответствующих периодов необходимо произвести центрирование расчетных средних.

Недостатком способа сглаживания рядов динамики является то, что полученные средние не дают теоретических рядов, в основе которых лежала бы математически выраженная закономерность.

3. Метод аналитического выравнивания.

При изучении общей тенденции методом аналитического выравнивания исходят из того, что изменения уровней ряда динамики могут быть с той или иной степенью точности приближения выражены определенными математическими функциями. Вид уравнения определяется характером динамики развития конкретного явления. Логический анализ при выборе вида уравнения может быть основан на рассчитанных показателях динамики, а именно:

– если относительно стабильны абсолютные приросты (первые разности уровней приблизительно равны), сглаживание может быть выполнено по прямой;

– если абсолютные приросты равномерно увеличиваются (вторые разности уровней приблизительно равны), можно принять параболу второго порядка;

– при ускоренно возрастающих или замедляющихся абсолютных приростах – параболу третьего порядка;

– при относительно стабильных темпах роста – показательную функцию.

Для аналитического выравнивания наиболее часто используются следующие виды трендовых моделей: прямая (линейная), парабола второго порядка, показательная (логарифмическая) кривая, гиперболическая.

Цель аналитического выравнивания – определение аналитической или графической зависимости. На практике по имеющемуся временному ряду задают вид и находят параметры функции, а затем анализируют поведение отклонений от тенденции. Чаще всего при выравнивании используются следующие зависимости: линейная, параболическая и экспоненциальная.

После выяснения характера кривой развития необходимо определить ее параметры, что можно сделать различными методами:

решением системы уравнений по известным уровням ряда динамики;

методом средних значений (линейных отклонений), который заключается в следующем: ряд расчленяется на две примерно равные части, и вводятся преобразования, чтобы сумма выравненных значений в каждой части совпала с суммой фактических значений, например, в случае выравнивания прямой линии

выравниванием ряда динамики с помощью метода конечных разностей;

методом наименьших квадратов: это некоторый прием получения оценки детерминированной компоненты $f(t)$, характеризующий тренд или ряд изучаемого явления.

Во многих случаях моделирование рядов динамики с помощью полиномов или экспоненциальной функции не дает удовлетворительных результатов, так как в рядах динамики содержатся заметные периодические колебания вокруг общей тенденции. В таких случаях следует использовать гармонический анализ.

6.5 Экстраполяция и интерполяция в рядах динамики

Основа большинства методов прогнозирования – экстраполяция тенденции, связанная с распространением закономерностей, связей и соотношений, действующих в изучаемом периоде, за его пределы или, другими словами, это получение представлений о будущем на основе информации, относящейся к прошлому и настоящему.

Экстраполяция, проводимая в будущее, – это перспектива, а в прошлое, – ретроспектива.

Предпосылки применения экстраполяции:

– развитие исследуемого явления в целом следует описывать плавной кривой;

– общая тенденция развития явления в прошлом и настоящем не должна претерпевать серьезных изменений в будущем.

Экстраполяцию в общем виде можно представить как:

$$\bar{y}_{i+1} = f(y_i, T, a_j)$$

где $\bar{y}_{i+1}$ – прогнозируемый уровень; y_i – текущий уровень прогнозного ряда; T – срок экстраполяции; a_j – параметр уравнения тренда.

Упрощенные приемы целесообразны при недостаточной информации о предыстории развития явления (нет достаточно длинного динамического ряда или информация задана только двумя точками: на начало и конец периода). Упрощенные приемы основываются на средних показателях динамики, и можно выделить:

1. Метод среднего абсолютного прироста.

Для нахождения интересующего нас аналитического выражения тенденции на любую дату необходимо определить средний абсолютный прирост и последовательно прибавить его к последнему уровню ряда столько раз, на сколько периодов экстраполируется ряд.

$$\bar{y}_{i+1} = y_i + \bar{\Delta} t$$

где t – срок прогноза; i – номер последнего уровня.

Применение в экстраполяции среднего абсолютного прироста предполагает, что развитие явления происходит по арифметической профессии и относится в прогнозировании к классу «наивных» моделей, ибо чаще всего развитие явления следует по иному пути, чем арифметическая прогрессия. Вместе с тем в ряде случаев этот метод может найти применение как предварительный прогноз, если у исследователя нет динамического ряда: информация дана лишь на начало и конец периода (например, данные одного баланса).

2. Метод среднего темпа роста.

Осуществляется, когда общая тенденция характеризуется показательной кривой

$$\bar{y}_{i+1} = y_i \bar{k}$$

где y_i – последний уровень ряда динамики; k – средний коэффициент роста.

3. Выравнивание рядов по какой-либо аналитической формуле.

Экстраполяция дает возможность получить точечное значение прогнозов. Точное совпадение фактических данных и прогнозных точечных оценок, полученных путем экстраполяции кривых, имеет малую вероятность.

Любой статистический прогноз носит приближенный характер, поэтому целесообразно определение доверительных интервалов прогноза:

$$\bar{y}_{i+1} \pm t_a S_{\bar{y}}$$

$$\bar{y}_{i+1} \pm t_a \sqrt{\frac{\sum (y_i - y_t)^2}{n-k}}$$

где t_a – коэффициент доверия по распределению Стьюдента при уровне значимости a; $S_{\bar{y}}$ – средняя квадратическая ошибка тренда; k – число параметров в уравнении; y_t – расчетное значение уровня.

Аналитические методы основаны на применении метода наименьших квадратов к динамическому ряду и представлении закономерности развития явления во времени в виде уравнения тренда, то есть математической функции уровней динамического ряда (y) от факторного времени (t): $y = f(t)$

Аналитическое сглаживание позволяет не только определить общую тенденцию изменения явления на рассматриваемом отрезке времени, но и выполнять расчеты для таких периодов, в отношении которых нет исходных данных.

Адаптивные методы используются в условиях сильной колеблемости уровней динамического ряда и позволяют при изучении тенденции учитывать степень влияния предыдущих уровней на последующие значения динамического ряда. К адаптивным методам относятся методы скользящих и экспоненциальных средних, метод гармонических весов, методы авторегрессионных преобразований.

Цель адаптивных методов заключается в построении самонастраивающихся моделей, способных учитывать информационную ценность различных членов временного ряда и давать достаточно точные оценки будущим членам данного ряда.

Прогноз получается как экстраполяция последней тенденции. В разных методиках прогнозирования процесс настройки (адаптации) модели осуществляется по-разному, и можно выделить:

метод скользящей средней (адаптивной фильтрации, метод Бокса-Дженкинса);

методы экспоненциального сглаживания (методы Хольда. Брауна, экспоненциальной средней).

Скользящие средние представляют собой средние уровни за определенные периоды времени путем последовательного передвижения начала периода на единицу времени. При простой скользящей средней все уровни временного ряда считаются равноценными, а при исчислении взвешенной скользящей средней каждому уровню в пределах интервала сглаживания приписывается вес, зависящий от расстояния данного уровня до середины интервала сглаживания.

Особенность метода экспоненциального сглаживания в том, что в процедуре выравнивания каждого наблюдения используется только значение предыдущих уравнений, взятых с определенным весом. Смысл экспоненциальных средних состоит в нахождении таких средних, в которых влияние прошлых наблюдений затухает по мере удаления от момента, для которого определяются средние.

6.6 Изучение сезонных колебаний

При изучении многих социально-экономических явлений и процессов часто обнаруживаются определенные, повторяющиеся колебания. Этим колебаниям свойственны более или менее устойчивые изменения уровней ряда на протяжении изучаемого периода: из года в год в определенные месяцы уровень явления повышается, а в другие – снижается. В статистике данные колебания принято называть «сезонными».

Сезонные колебания (сезонная неравномерность) чаще всего происходят в добывающих и перерабатывающих отраслях сельском хозяйстве, рыбной и лесной промышленности, а также на транспорте, в строительстве, торговле, туризме и т.д.

Погодные изменения влияют на бытовое потребление топлива и электроэнергии, на ассортимент обуви, верхней одежды (зимняя, весенне-осенняя, летняя), фруктов, овощей и многих других товаров. В строительстве наибольшее оживление деятельности проявляется летом; в этот же период года наблюдается максимальный наплыв туристов. Сезонность может проявляться не только к месячным, но и к дневным, недельным данным. Так, кафе, рестораны, театры испытывают подъем спроса к концу недели.

Сезонность проявляется в полном или почти полном прекращении производства на какой-то промежуток времени, обусловленный самой природой продукта и способом его изготовления.

Созревание зерновых, например, требует несколько месяцев, а в садоводстве после посадки саженцев проходит несколько лет до получения готового продукта. В тех же отраслях, которые характеризуются *незначительностью разрыва рабочего периода и времени производства, сезонность проявляется в виде больших внутригодичных подъемов и спадов.*

Итак, вызванные различными причинами, сезонные колебания и в производстве и в обращении обычно отрицательно влияют на результаты производственной деятельности из-за того, что вызывают нарушение ритмичности производства, обусловливают неравномерность использования трудовых ресурсов и оборудования в течение года и т.д. Многие отрасли экономики взаимосвязаны, поэтому проблема сезонности – общая проблема экономики разных стран. Неравномерность производства того или иного продукта ведет к неравномерности его потребления, потребление же, в свою очередь, оказывает воздействие на производство.

Влияние сезонных колебаний полностью устранить невозможно, но некоторые предприятия пытаются его снизить, принимая меры рационального сочетания отраслей, механизации трудоемких процессов и т.д. Вот по этой причине сезонные колебания, отраженные в рядах динамики, необходимо изучать и измерять.

По своему существу все методы анализа сезонности делятся на две группы. К *первой группе* относятся методы, с помощью которых определяется и измеряется сезонность непосредственно из эмпирических данных, без особой предварительной их обработки, – *метод простой средней, метод относительных чисел и метод У. Персонса*.

Суть методов *второй группы* заключается в предварительном определении и исключении общей тенденции развития и в последующем исчислении и количественном измерении сезонных колебаний. К методам анализа сезонности данной группы можно отнести *метод аналитического выравнивания и метод скользящей (подвижной) средней*.

Метод простой средней применяется для анализа сезонности явлений, уровни которых не имеют резко выраженной тенденции увеличения или уменьшения. Сущность этого метода заключается в определении сезонной волны или индекса сезонности. Способы определения индексов сезонности различны, они зависят прежде всего от характера общей тенденции ряда динамики.

Индексы сезонности – процентные отношения фактических (эмпирических) внутригрупповых уровней к теоретическим расчетным уровням, выступающим в качестве базы сравнения. Их вычисляют по данным за несколько лет (не менее трех), распределенным по месяцам или кварталам.

Для каждого месяца рассчитываются средняя величина уровня, а затем – среднемесячный уровень для всего ряда (в %):

$$I_S = \frac{\bar{y}_i}{\bar{y}_0} 100\%$$

где $\bar{y}_i = \frac{\sum y_i}{n}$ осредненные эмпирические уровни ряда по одноименным периодам (месяцам или кварталам); $\bar{y}_0 = \frac{\sum \bar{y}_i}{n}$ или $\bar{y}_0 = \frac{\sum\sum y_i}{\sum n}$ – общий средний уровень ряда.

Метод относительных чисел применяется для анализа сезонности тех рядов динамики, развитие общей тенденции которых происходит равномерно. Основной недостаток – механическое внесение относительно единственной поправки в анализируемые отрезки времени, которая означает признание равномерного развития уровней явления.

Анализ сезонности методом Персонса применяется в рядах динамики, отражающих развитие явлений, общая тенденция которых изменяется по средней геометрической, то есть по сложным процентам. Суть метода заключается в исчислении показателей средней сезонной волны как медианных значений из цепных отношений. Здесь погрешность устраняется с помощью коэффициента подъема или снижения общей тенденции по средней геометрической.

Во многих случаях, когда в рядах динамики наблюдается явно выраженные периодические колебания, для описания тренда следует использовать спектральный анализ, когда динамический ряд аппроксимируется функциями Фурье. Другими словами, он представляет собой операцию по выражению заданной периодической функции в виде ряда Фурье по гармоникам разных порядков. Основы гармонического анализа были разработаны французским математиком Жаном Батистом Жозефом Фурье (1768-1829), заслуги которого на поприще математических наук сосредоточиваются главным образом на теории решения численных уравнений, теории распространения тепла и на разложении функций в ряды тригонометрических функций. Он показал, что дифференцируемая функция может быть представлена в виде некоторого ряда, все члены которого представляют собой гармонические функции. Каждый член ряда представляет собой слагаемое постоянной величины с функциями cos и sin определенного периода. Нахождение конечной суммы уровней с использованием функций косинусов и синусов времени называется гармоническим анализом.

В связи с тем, что уравнение колебательного процесса (гармоники) формируется с помощью основных тригонометрических функций, то оно является предметом подробного рассмотрения в математической статистике.

Обобщающим показателем силы колеблемости динамического ряда из-за сезонного характера производства или обращения служит среднее квадратическое отклонение индексов сезонности, то есть:

$$\sigma_{сез} = \sqrt{\frac{\sum(i_{сез} - 100)^2}{12}}$$

Сравнение показателей $\sigma_{сез}$, вычисленных за разные периоды, показывает сдвиги в сезонности.

ГЛАВА 7 ИНДЕКСЫ

7.1 Общее понятие индексов в статистике. Классификация индексов.

Слово index в переводе с латинского означает «указатель», «показатель», «список».

Индексы – относительные показатели, предназначенные для описания изменения величины какого-либо явления во времени, пространстве или по сравнению с любым эталоном.

Основной элемент индексного соотношения – индексируемая величина, которая представляет собой значение признака статистической совокупности, изменение которой является объектом изучения.

С помощью индексов решаются следующие основные задачи:

– характеризуется общее *изменение сложных социально-экономических явлений и отдельных его элементов* (изучается развитие национальной экономики в целом и ее отдельных отраслей, определяется уровень жизни населения и т.д.);

– выясняется роль *отдельных факторов* в формировании важнейших экономических показателей;

– являются *показателями сравнений* не только с прошлым периодом (сравнение во времени), но и с другой территорией – сравнение в пространстве, а также с планами, нормативами, прогнозами и т. д.;

– используются в международных сопоставлениях макроэкономических показателей, то есть *производится пересчет* значений показателей *из фактических цен в сопоставимые*.

Для удобства применения индексного метода, составления формул индексов и их использования в экономико-статистическом анализе в теории статистики разработана определенная символика, каждая индексируемая величина имеет условное обозначение:

i – индивидуальный индекс,

I – общий (агрегатный) индекс,

p_i – цена за единицу продукции за i-й период,

q_i – количество продукции одного вида в натуральном выражении,

p_iq_i или w_i – товарооборот, стоимость продукции,

z_i – себестоимость единицы продукции,

q_iz_i – полная себестоимость (весь объем),

t_i – производительность, затраты труда (рабочего времени) на единицу продукции.

Знак внизу справа означает период: 0 – базисный, 1 – отчетный.

Свойства индексов:

а) синтетические свойства общих индексов – характеризуются тем, что они выражают и обобщают относительные изменения сложных явлений, отдельные части и элементы которых непосредственно несоизмеримы;

б) аналитические свойства общих индексов – следуют из взаимосвязи индексов и состоят в том, что с помощью индексного метода определяется влияние факторов на изменение изучаемого показателя.

Индексы классифицируются по:

– базе сравнения;
– степени охвата элементов совокупности;
– содержанию изучаемых объектов;
– методам расчета общих индексов.

1. Виды индексов по базе сравнения.

Когда рассматриваются сопоставления уровней изучаемого явления во времени, то говорят о динамических индексах, в пространстве – о территориальных индексах и т. д. Динамические индексы бывают базисные и цепные. Территориальные индексы применяются для межрегиональных сравнений, а также в международной статистике при сопоставлении показателей социально-экономического развития различных стран.

2. Виды индексов по степени охвата элементов совокупности: индивидуальные и сводные.

Индивидуальные индексы служат для характеристики соотношения уровней только одного элемента совокупности, например рост или падение цен на какой-либо продукт. Индивидуальные индексы по существу являются относительными величинами динамики, выполнения плана или сравнения, и индексами их можно назвать только в широком смысле.

Индекс, как относительный показатель, выражается в виде коэффициентов, когда база сравнения принимается за единицу, и в процентах, когда база для сравнения принимается за 100.

Расчеты индивидуальных индексов просты по своей сущности и выполняются путем вычисления отношения двух индексируемых величин. Можно выделить:

а) индивидуальный индекс цен, который показывает относительное изменение уровня цены по каждому виду продукции в отчетном году по сравнению с базисным:

$$i_p = \frac{p_1}{p_0}$$

б) индивидуальный индекс физического объема, характеризующий изменение объемов производства или реализации в натуральном выражении по каждому виду продукции во времени, в пространстве и плана, если фактический выпуск сравнивать с плановым заданием:

$$i_q = \frac{q_1}{q_0}$$

в) индивидуальный индекс выручки, отражающий динамику средств, полученных от реализации отдельных видов продукции:

$$i_{pq} = \frac{p_1 q_1}{p_0 q_0} = \frac{w_1}{w_0}$$

Индивидуальные индексы выручки можно найти и через взаимосвязь индексов:

$$i_{pq} = i_p i_q$$

Пример. Рассмотрим подробнее индивидуальные индексы с учетом условных данных о ценах и реализации продукции за два периода:

Товар	*Ед.изм.*	*III квартал*		*IV квартал*	
		Цена за 1 ед., руб.	*Количество*	*Цена за 1 ед., руб.*	*Количество*
Молоко	*л*	*9,8*	*7500*	*10,1*	*6800*
Яйцо	*шт.*	*26,7*	*1690*	*27,0*	*1830*
Картофель	*кг*	*4,85*	*14750*	*5,2*	*10050*

При определении по данным таблицы статистических индексов Ш квартал принимается за базисный период, в котором цена единицы товара обозначается p_0, а количество – q_0.

IV квартал принимается за текущий или отчетный период, в котором цена единицы товара обозначается p_1, количество – q_1.

Тогда индивидуальные индексы составят:

• цен:

$$i_p^{м} = 10{,}1 / 9{,}8 = 1{,}0306,\ \textit{или}\ 103{,}06\ \%$$

$$i_p^{я} = 27{,}0 / 26{,}7 = 1{,}0112,\ \textit{или}\ 101{,}12\ \%$$

$$i_p^{к} = 5{,}2 / 4{,}85 = 1{,}0722,\ \textit{или}\ 107{,}22\ \%$$

В IV квартале цены увеличились по всем видам представленной продукции: на молоко – на 3,06 %, на яйца – на 1,12 %, на картофель – на 7,22 %;

• физического объема:

$$i_q^{м} = 6800 / 7500 = 0{,}9067,\ \textit{или}\ 90{,}67\ \%$$

$$i_q^{я} = 1830 / 1690 = 1{,}0828,\ \textit{или}\ 108{,}28\ \%$$

$$i_q^{к} = 10050 / 14750 = 0.6814,\ \textit{или}\ 68{,}14\ \%$$

Количество проданного молока в IV квартале снизилось по сравнению с III на 9,33 % (90,67% – 100 %), объем продажи яиц повысился на 8,28 %, а объем реализации картофеля сократился на 31,86 %;

• товарооборота:

$$i_{pq}^{м} = \frac{10,1 \times 6800}{9,8 \times 7500} = \frac{68680}{73500} = 0,9344,\ \textit{или}\ 93,44\ \%$$

$$i_{pq} = \frac{27 \times 1830}{26,7 \times 1690} = \frac{49410}{45123} = 1,095,\ \textit{или}\ 109,5\ \%$$

$$i_{pq}^{к} = \frac{5,2 \times 10050}{4,85 \times 14750} = \frac{52260}{71537,5} = 0,7305,\ \textit{или}\ 73,05\ \%$$

В IV квартале по сравнению с III выручка от реализации молока снизилась на 6,56 %, яиц – выросла на 9,5 %, картофеля – снизилась на 26,95 %.

Если необходимо получить характеристику изменения изучаемого явления во всех последующих периодах по сравнению с первоначальным, то вычисляются базисные индексы. Они имеют постоянную базу сравнения, в качестве которой принимаются данные какого-то одного периода (при анализе динамики), какой-то территории (при территориальных сравнениях) и планового задания (при анализе выполнения плана). Например, сопоставление денежных доходов на душу населения II, III и IV кварталов с I кварталом.

В случае, когда требуется охарактеризовать последовательное изменение изучаемого явления из периода в период, вычисляются цепные индексы. Они представляют собой сравнения текущих уровней с предшествующим уровнем или непрерывно меняющейся базой сравнения. Например, при изучении денежных доходов на душу населения по кварталам года сопоставляют денежные доходы II квартала с I кварталом, III квартала – со II кварталом и IV квартала – с III кварталом.

В зависимости от задачи исследования и характера исходной информации базисные и цепные индексы исчисляются как индивидуальные (однотоварные), так и общие.

Способы расчета индивидуальных базисных и цепных индексов аналогичны расчету относительных величин динамики.

Таблица – Система индивидуальных индексов

Название индивидуального индекса	Система индексов	
	базисных	цепных
Индекс стоимости	$\frac{p_1q_1}{p_0q_0}$; $\frac{p_2q_2}{p_0q_0}$;… $\frac{p_nq_n}{p_0q_0}$	$\frac{p_1q_1}{p_0q_0}$; $\frac{p_2q_2}{p_1q_1}$;… $\frac{p_nq_n}{p_{n-1}q_{n-1}}$
Индекс физического объема	$\frac{q_1}{q_0}$; $\frac{q_2}{q_0}$;… $\frac{q_n}{q_0}$	$\frac{q_1}{q_0}$; $\frac{q_2}{q_1}$; … $\frac{q_n}{q_{n-1}}$
Индекс цен	$\frac{p_1}{p_0}$; $\frac{p_2}{p_0}$; … $\frac{p_n}{p_0}$	$\frac{p_1}{p_0}$; $\frac{p_2}{p_1}$; … $\frac{p_n}{p_{n-1}}$

Между цепными и базисными индивидуальными индексами существует взаимосвязь, которая позволяет переходить от одних индексов к другим.

Отношение базисного индекса отчетного периода к базисному индексу предшествующего периода дает цепной индекс отчетного периода:

$$\frac{p_2}{p_1} = \frac{p_2}{p_0} \div \frac{p_1}{p_0}$$

Произведение последовательных цепных индивидуальных индексов дает базисный индекс последнего периода:

$$\frac{p_2}{p_0} = \frac{p_2}{p_1} \times \frac{p_1}{p_0}$$

Системы базисных и цепных индексов могут быть построены и для агрегатных индексов, только с использованием постоянных и переменных весов.

В связи с тем, что индивидуальные индексы являются разновеликими по направлению и интенсивности изменения, возникает необходимость их обобщения при определении общего для данного ассортимента изменения цен и количества реализованных товаров. Для этого вычисляются соответствующие общие индексы.

Общие (сводные) индексы отражают изменение всех элементов сложного явления и состоят из двух частей: индексируемой величины и соизмерителя, который называется «весом». Одна из особенностей сводных индексов состоит в том, что исследуемый показатель рассматривается не изолированно, а во взаимосвязи с другими показателями. Они имеют более сложную методику построения и расчета.

3. Виды индексов по содержанию изучаемых величин: индексы качественных и индексы количественных показателей.

Примерами индексов *качественных показателей* могут быть: индексы курса валют, цен, себестоимости продукции, производительности труда, заработной платы, урожайности и т. д. Индексируемые показатели этих индексов характеризуют уровень явления в расчете на ту или иную единицу совокупности: цена за единицу продукции, заработная плата одного работника и т. д.

Качественные показатели измеряют уровень явления или иную единицу совокупности, поэтому они являются расчетными, вторичными показателями интенсивности. Обычно это либо средние, либо относительные величины. Рассчитываются данные индексы на базе одинаковых, неизменных объемов продукции.

Количественные показатели характеризуют общий, суммарный размер того или иного явления и выражаются абсолютными величинами. Количества при расчете данных индексов оцениваются в одинаковых, сопоставимых ценах. В качестве примеров можно привести индексы национального дохода, индексы физического объема оптового и розничного товарооборота и т. д.

4. Виды индексов по методам расчета.

По методам расчета общих и групповых индексов различают агрегатные и средние индексы, вычисление которых и составляет особый прием исследования, который называется индексным методом.

Основная форма общих индексов – агрегатные индексы. Свое название они получили от латинского слова «aggrega», что означает «присоединяй». Числитель и знаменатель агрегатного индекса представляют собой набор или «агрегат» – сумму произведения двух величин, одна из которых меняется (индексируется), а другая остается неизменной в числителе и знаменателе (вес индекса).

Коэффициенты соизмерения необходимы для перехода от натуральных измерителей разнородных единиц статистической совокупности к однородным показателям и обеспечивают количественную сравнимость. Поэтому их показатели-сомножители, связанные с индексируемыми величинами, принято называть весами индексов, а умножение на них – взвешиванием. Это необходимо для того, чтобы на величине индекса сказывалось лишь влияние фактора, который определяет изменение индексируемой величины.

Основным условием применения в статистике агрегатных индексов является наличие информации о поступлении или реализации товаров в натуральных измерителях и ценах единицы товара.

Французский экономист Шарль Дюто в 1738 г. предложил первую формулу для вычисления агрегатного индекса:

$$I_p^{д} = \frac{\sum P_1}{\sum P_0}$$

По результатам практических исследований была разработана рекомендация, соблюдение которой обеспечивает увязку сопряженных индексов в системе, что повышает их аналитические возможности. Смысл рекомендации состоит в следующем: в расчетах агрегатных индексов при индексировании вторичных признаков взвешивание обычно производится по отчетным весам, а первичных признаков – по базисным весам.

Средние индексы – средние взвешенные величины из индивидуальных индексов и, являясь производными от агрегатных индексов, рассчитываются по формулам среднего арифметического и среднего гармонического показателей. Первая формула для расчета средних индексов была введена в 1764 г. итальянским экономистом Карли:

$$I_p^{к} = \frac{\sum i_p}{n}$$

Но простые индексы не дают полной картины изменения параметров, так как предполагается, что все составляющие индексируемой величины имеют равное влияние на общий результат.

Пример. Имеются следующие данные о средних ценах на однородные продукты, у.е. за единицу:

Вид продукта	декабрь	январь
А	3,2	3,42
Б	4,1	4,24
В	2,98	3,36
Г	3,84	4,00
Д	4,36	4,58

Вычислите индивидуальные индексы цен, а также субиндексы цен по методикам Дюто и Карли.

Решение

$$\sum i_p = \sum \frac{p_1}{p_0} = \frac{3,42}{3,2} + \frac{4,24}{4,1} + \frac{3,36}{2,98} + \frac{4,00}{3,84} + \frac{4,58}{4,36} = 1,069 + 1,034 + 1,128 + 1,042 + 1,053 = 5,323$$

$$I_p^{д} = \frac{3,42 + 4,24 + 3,36 + 4,00 + 4,58}{3,2 + 4,1 + 2,98 + 3,84 + 4,36} = \frac{19,6}{18,48} = 1,06 \text{ или } 106\ \%$$

$$I_p^{к} = \frac{5,323}{5} = 1,065, \text{ или } 106,5\ \%$$

Следовательно, в январе по сравнению с декабрем цены увеличились по всем видам представленной продукции приблизительно на 6 %.

7.2 Агрегатный индекс как основная форма общего индекса

Агрегатный индекс стоимости продукции.

Индекс физического объема продукции является типичным индексом *количественных показателей*.

Стоимость продукции представляет собой произведение количества продукции в натуральном выражении (q) на цену единицы продукции (p).

Агрегатный индекс стоимости продукции или товарооборота в фактических ценах представляет собой отношение стоимости продукции текущего периода к стоимости базисного периода. Он может быть рассчитан по формуле

$$I_{pq} = \frac{\sum p_1 q_1}{\sum p_0 q_0} = \frac{\sum w_1}{\sum w_0}$$

где

$\sum p_1 q_1$ – фактическая стоимость продукции отчетного периода;

$\sum p_0 q_0$ – фактическая стоимость продукции базисного периода.

Данный индекс показывает, во сколько раз возросла (уменьшилась) стоимость продукции отчетного периода по сравнению с базисным либо сколько процентов составляет рост (снижение) стоимости продукции. Его значение зависит от двух факторов: изменения цены и количества продукции.

Поскольку числитель и знаменатель агрегатных индексов имеют экономический смысл, в статистическом анализе нередко используются их разности. Так, например, разность числителя и знаменателя индекса товарооборота

$$\Delta pq = \sum p_1 q_1 - \sum p_0 q_0$$

характеризует абсолютный прирост (уменьшение) товарооборота в отчетном периоде по сравнению с базисным одновременно за счет:

изменения физического объема продаж;

изменения цен.

Измерить изолированное (элиминированное) влияние каждого из этих двух факторов можно через разность числителя и знаменателя соответствующих аналитических индексов.

Агрегатный индекс физического объема продукции.

Если продукцию товара сравниваемых периодов оценивать по одним и тем же, например базисным, ценам, то такой индекс отразит изменение только одного фактора – индексируемого показателя q и будет называться агрегатным индексом физического объема продукции. Он представляет собой отношение условной стоимости произведенных в текущем периоде товаров в ценах базисного периода к фактической стоимости товаров, произведенных в базисном периоде, – ***индекс Ласпейреса:***

$$I^{Л}{}_{q} = \frac{\sum p_0 q_1}{\sum p_0 q_0}$$

Или – это отношение фактической стоимости товаров, произведенных в отчетном периоде, к условной стоимости товаров, произведенных в базисном периоде, в ценах отчетного периода – ***индекс Пааше***:

$$I^{П}{}_{q} = \frac{\sum p_1 q_1}{\sum p_1 q_0}$$

$\sum p_0 q_1$ – условная стоимость товаров, реализованных в отчетном периоде по базисным ценам;

$\sum p_1 q_0$ – условная стоимость товаров, реализованных в базисном периоде по отчетным ценам.

Абсолютное изменение общей стоимости продукции за счет изменения физического объема продукции можно определить следующим образом:

$$\Delta pq(q)^{Л} = \sum p_0 q_1 - \sum p_0 q_0$$

$$\Delta pq(q)^{П} = \sum p_1 q_1 - \sum p_1 q_0$$

Агрегатные индексы цен Пааше, Ласпейреса и Фишера.

Агрегатная форма индексов цен, в которой изменение цен увязывалось с конкретной массой товаров, была введена в практику расчетов во второй половине XIX в. В 1864 г. немецкий экономист Э. Ласпейрес предложил индекс, который отражает изменение цен и строится по продукции базисного периода. Формула ***агрегатного индекса цен Ласпейреса*** представляет собой следующее отношение:

$$I^{Л}{}_{p} = \frac{\sum p_1 q_0}{\sum p_0 q_0}$$

Индекс цен Ласпейреса показывает, во сколько раз товары базисного периода подорожали или подешевели в результате изменения цен на них в отчетном периоде. Эти особенности индекса Ласпейреса обусловливают его применение при прогнозировании объема товарооборота, в связи с намечаемыми изменениями цен на товары в предстоящем периоде.

В 1874 г. немецкий экономист ***Г. Пааше*** впервые предложил ***агрегатный индекс цен*** с отчетными весами. Формула этого индекса выглядит следующим образом:

$$I^{П}{}_{p} = \frac{\sum p_1 q_1}{\sum p_0 q_1}$$

Индекс цен Пааше характеризует изменение цен отчетного периода по сравнению с базисным по товарам и услугам, реализованным в отчетном периоде, и фактическую экономию или перерасход от изменения цен.

Кроме того, при построении общего индекса цен в качестве соизмерителей индексируемых величин (p_1 и p_0) могут применяться средние величины реализации товаров за два или большее число периодов.

Разность числителя и знаменателя агрегатного индекса цен показывает, как в абсолютном выражении изменилась общая стоимость продукции за счет роста (сокращения) цен на продукцию:

$$\Delta pq(p)^{Л} = \sum p_1 q_0 - \sum p_0 q_0$$

$$\Delta pq(p)^{П} = \sum p_1 q_1 - \sum p_0 q_1$$

Американский экономист И. Фишер предложил «идеальный» индекс, названный его именем, представляющий среднюю геометрическую из произведения двух агрегатных индексов цен Ласпейреса и Пааше:

$$I_p^{Ф} = \sqrt{I_p^{Л} I_p^{П}}$$

Данную формулу можно использовать и для определения индекса физического объема.

Идеальность формулы Фишера заключается в том, что индекс обратим во времени, т.е. не зависит от выбора базы сравнения. Недостаток же формулы в том, что она лишена конкретного экономического содержания (разность между числителем и знаменателем не показывает никакой реальной экономии или потерь вследствие изменения цен).

Индекс Фишера в силу сложности расчета и трудности экономической интерпретации на практике используется довольно едко. Кроме того, многочисленные расчеты показали, что вполне можно применять не среднюю геометрическую, а среднюю арифметическую величину из индексов Ласпейреса и Пааше для получения осредненной величины индекса.

Рассмотрим основные приемы применения индексного анализа на примере

Пример. Имеются следующие данные о ценах реализации товаров, в руб.:

Товар	Ед.изм.	Базисный период		Текущий период	
		Цена за 1 ед.	Количество	Цена за 1 ед.	Количество
А	т	20	7500	25	9500
Б	шт.	30	2000	29	2500

Определить:

1. агрегатный индекс цен на товары, взвешенный по продукции текущего периода (индекс Пааше) и по продукции базисного периода (индекс Ласпейреса), а также «идеальный» индекс Фишера;

2. агрегатный индекс физического объема продажи товаров и услуг в сопоставимых ценах по методикам Пааше и Ласпейреса;

3. агрегатный индекс товарооборота;

4. абсолютный прирост стоимости товаров вследствие изменения цен и объема продаж в целом по двум видам товаров.

Решение.

1) Агрегатные индексы цен:

$$I^{П}{}_{p} = \frac{\sum p_1 q_1}{\sum p_0 q_1} = \frac{25 \times 9500 + 29 \times 2500}{20 \times 9500 + 30 \times 2500} = \frac{237500 + 72500}{190000 + 75000} = \frac{310000}{265000} = 1{,}1698$$

$$I^{л}{}_{p} = \frac{\sum p_1 q_0}{\sum p_0 q_0} = \frac{25 \times 7500 + 29 \times 2000}{20 \times 7500 + 30 \times 2000} = \frac{187500 + 58000}{150000 + 60000} = \frac{245500}{210000} = 1{,}169$$

$$I_p^{ф} = \sqrt{I_p^{Л} I_p^{П}} = \sqrt{1{,}169 \times 1{,}1698} = \sqrt{1{,}3675} = 1{,}1694$$

Таким образом, выполненные расчеты имеют разные показания индексов цен. Это объясняется тем, что индексы Пааше и Ласпейреса характеризуют качественные особенности изменения цен.

В текущем периоде по сравнению с базисным наблюдался рост цен на товары А и Б на 16,9 % по формуле Ласпейреса, на 16,98 % – по формуле Пааше. Индекс Фишера дает осредненную величину повышения цены – 16,94 %.

2) Агрегатные индексы физического объема:

$$I^{Л}_{q} = \frac{\sum p_0 q_1}{\sum p_0 q_0} = \frac{20\times 9500 + 30\times 2500}{20\times 7500 + 30\times 2000} = \frac{265000}{210000} = 1{,}2619$$

$$I^{П}_{q} = \frac{\sum p_1 q_1}{\sum p_1 q_0} = \frac{25\times 9500 + 29\times 2500}{25\times 7500 + 29\times 2000} = \frac{310000}{245500} = 1{,}2627$$

3) Агрегатный индекс выручки от реализации или товарооборота:

$$I_{pq} = \frac{\sum p_1 q_1}{\sum p_0 q_0} = \frac{310000}{210000} = 1{,}4761$$

Для расчета агрегатного индекса выручки от реализации также можно воспользоваться любой из двух моделей:

4) Абсолютный прирост стоимости товаров:

$$\Delta pq = \sum p_1 q_1 - \sum p_0 q_0 = 310000 - 210000 = 100000 \text{ руб.}$$

- вследствие изменения цен:

$$\Delta pq(p)^{Л} = \sum p_1 q_0 - \sum p_0 q_0 = 245500 - 210000 = 35500 \text{ руб.}$$

$$\Delta pq(p)^{П} = \sum p_1 q_1 - \sum p_0 q_1 = 310000 - 265000 = 45000 \text{ руб.}$$

Величина этого показателя характеризует перерасход средств населением вследствие изменения цен при покупке товаров данного ассортимента на 35500 руб. (по формуле Ласпейреса) и на 45000 руб. (по формуле Пааше) по ценам, повышенным на 16,9% или 16,98% соответственно;

- вследствие изменения объема продаж:

$$\Delta pq(q)^{Л} = \sum p_0 q_1 - \sum p_0 q_0 = 265000 - 210000 = 55000 \text{ руб.}$$

$$\Delta pq(q)^{П} = \sum p_1 q_1 - \sum p_1 q_0 = 310000 - 245500 = 64500 \text{ руб.}$$

Таким образом, выручка от реализации продукции вследствие изменения объема продаж увеличилась на 55000 руб. (по формуле Ласпейреса) и на 64500 руб. (по формуле Паше).

7.3 Средние индексы

Средневзвешенный индекс – средняя величина из индивидуальных индексов. При его расчете используются две формы средних величин: арифметическая и гармоническая. Применение той или иной формулы средневзвешенного индекса зависит от имеющейся в распоряжении информации.

Средний гармонический индекс цен применяется в тех случаях, когда неизвестны отдельные значения p_1 и q_1, но дано их произведение p_1q_1 и индивидуальные индексы цен $i_p = p_1/p_0$, а сводный индекс должен быть исчислен с отчетными весами. Индивидуальные индексы должны быть взвешены таким образом, чтобы средний гармонический индекс совпал с агрегатным.

Средний гармонический индекс цен:

$$I^{П}{}_{p} = \frac{\sum p_1 q_1}{\sum p_0 q_1} = \frac{\sum p_1 q_1}{\sum \frac{p_1 q_1}{i_p}}, \text{ т.к. } i_p = \frac{p_1}{p_0}, \text{ то } p_0 = \frac{p_1}{i_p}$$

Весами индивидуальных индексов в данном индексе является стоимость отдельных видов продукции отчетного периода в ценах того же периода p_1q_1..

Средний арифметический индекс цен получается в том случае, если из индивидуального индекса цен $i_p = \frac{p_1}{p_0}$ выразить цену отчетного периода $p_1 = i_p p_0$, а затем подставить ее в числитель агрегатного индекса цен. Данный индекс тождествен агрегатному индексу Ласпейреса и имеет следующий вид:

$$I^{Л}{}_{p} = \frac{\sum p_1 q_0}{\sum p_0 q_0} = \frac{\sum i_p p_0 q_0}{\sum p_0 q_0}$$

Средний арифметического индекса физического объема продукции:

$$I^{Л}{}_{q} = \frac{\sum i_q p_0 q_0}{\sum p_0 q_0}$$

Средний гармонический взвешенный индекс физического объема продукции:

$$I^{П}{}_{q} = \frac{\sum p_1 q_1}{\sum \frac{p_1 q_1}{i_q}}$$

Средние индексы широко применяются для анализа состояния рынка ценных бумаг. Индикаторы рынка ценных бумаг рассчитываются по их видам: акции, облигации, опционы и др. Можно выделить следующие интегральные показатели: индексы Доу-Джонса, индекс Стендарта и Пура.

Покажем особенность применения средних индексов на следующем примере.

Пример. Потребление товаров и услуг населением района характеризуется следующими показателями:

	Стоимость товаров	Стоимость товаров	Изменения во II квартале по

Виды товаров и услуг	и услуг в I квартале в текущих ценах, млн руб.	и услуг в II квартале в текущих ценах, млн руб.	сравнению с I кварталом, %	
			цен	объема продаж
Продовольственные товары	207	216	+20	-10
Непродовольственные товары	332	345	+15	-20
Платные услуги	105	126	+50	Без изменений

Определить:

1) индивидуальные индексы цен и физического объема;

2) общий индекс на товары и услуги;

3) общий индекс физического объема продажи товаров и услуг в сопоставимых ценах;

4) общий индекс товарооборота.

Решение

1) Индивидуальные индексы рассчитываются по формулам:

$$i_p = \frac{p_1}{p_0} \qquad i_q = \frac{q_1}{q_0}$$

следовательно,

$$i^{п}{}_p = \frac{100+20}{100} = 1{,}2 \qquad i^{н}{}_p = \frac{100+15}{100} = 1{,}15$$

$$i^{у}{}_p = \frac{100+50}{100} = 1{,}5$$

$$i^{п}{}_q = \frac{100-10}{100} = 0{,}9 \qquad i^{н}{}_q = \frac{100-20}{100} = 0{,}8 \qquad i^{у}{}_q = \frac{100}{100} = 1{,}0$$

2) Тогда сводные индексы цен на товары и услуги, рассчитанные по формулам Паше и Ласпейреса с учетом индивидуальных индексов, составят:

$$I_p^{П} = \frac{\sum p_1 q_1}{\sum p_0 q_1} = \frac{\sum p_1 q_1}{\sum \frac{p_1 q_1}{i_p}} = \frac{216+345+126}{\frac{216}{1{,}2}+\frac{345}{1{,}15}+\frac{126}{1{,}5}} = \frac{687}{564} = 1{,}218$$

или 121,8 %;

$$I_p^л = \frac{\sum p_1 q_0}{\sum p_0 q_0} = \frac{\sum i_p p_0 q_0}{\sum p_0 q_0} = \frac{1{,}2 \times 207 + 1{,}15 \times 332 + 1{,}5 \times 105}{207 + 332 + 105} = \frac{787{,}7}{644} = 1{,}223$$

или 122,3 %.

Во II квартале по сравнению с I кварталом наблюдался рост цен на все виды товаров и услуг на 21,8 % по формуле Пааше, на 22,3 % - по формуле Ласпейреса.

3) Сводные индексы физического объема на товары и услуги, рассчитанные по формулам Паше и Ласпейреса с учетом индивидуальных индексов, составят:

$$I_q^л = \frac{\sum i_q p_0 q_0}{\sum p_0 q_0} = \frac{0{,}9 \times 207 + 0{,}8 \times 332 + 1 \times 105}{207 + 332 + 105} = \frac{556{,}9}{644} = 0{,}865$$

или 86,5 %

$$I_q^П = \frac{\sum p_1 q_1}{\sum \frac{p_1 q_1}{i_q}} = \frac{216 + 345 + 126}{\frac{216}{0{,}9} + \frac{345}{0{,}8} + \frac{126}{1}} = \frac{687}{776{,}25} = 0{,}885$$

или 88,5 %.

Таким образом, во II квартале по сравнению с I кварталом в целом имеет место снижение объема реализации продукции и услуг в натуральном выражении на 13,5 % по формуле Ласпейреса, на 11,5 % – по формуле Пааше.

4) Общий индекс товарооборота составит:

$$I_{pq} = \frac{\sum p_1 q_1}{\sum p_0 q_0} = \frac{687}{644} = 1{,}0668, \text{ или } 106{,}68\ \%.$$

Индекс показывает, что стоимость товаров и услуг в среднем увеличилась на 6,68 %.

7.4 Индексный анализ средних величин: индексы постоянного, переменного составов и структурных сдвигов

Для однородной совокупности возможно использование средних значений, общее изменение которых обусловлено взаимодействием двух факторов: изменением отдельных уровней показателя и изменением в структуре весов.

Под изменением структуры здесь понимают изменение доли отдельных единиц совокупности, из которых формируются средние в общей их численности. *Например, динамика среднедушевого дохода населения зависит от изменения среднего дохода каждого человека и от изменения количества людей с более высокими и низкими доходами в общей численности; средняя цена на хлеб может изменяться не только под влиянием изменения* *цены хлеба, но и в результате изменения состава товарной массы.* Эта задача решается с помощью индексного метода, то есть путем построения системы взаимосвязанных индексов. Система включает три индекса: переменного состава, постоянного состава и структурных сдвигов.

Индекс переменного состава представляет собой соотношение средних величин какого-либо признака в отчетном и базисном периодах:

$$I_{перем} = \frac{\bar{x}_1}{\bar{x}_0} = \frac{\sum x_1 f_1}{\sum f_1} \div \frac{\sum x_0 f_0}{\sum f_0}$$

Данный индекс характеризует не только изменение индивидуальных цен в местах продажи, но и изменение структуры реализации по предприятиям розничной и оптовой торговли, рынкам, городам и регионам.

Индекс постоянного (фиксированного) состава характеризует динамику средней величины при одной и той же фиксированной структуре какой-либо однородной совокупности, то есть когда влияние структурного фактора устранено:

$$I_{пост} = \frac{\sum x_1 f_1}{\sum f_1} \div \frac{\sum x_0 f_1}{\sum f_1}$$

Индекс постоянного состава может быть рассчитан и в агрегатной форме:

$$I_{пост} = \frac{\sum x_1 f_1}{\sum x_0 f_1}$$

Индекс структурных сдвигов необходим для измерения влияния только структурных изменений на исследуемый средний показатель и может быть рассчитан по формуле:

$$I_{стр} = \frac{\sum x_0 f_1}{\sum f_1} \div \frac{\sum x_0 f_0}{\sum f_0}$$

Индексы переменного, постоянного состава и структурных сдвигов увязываются в следующую систему:

$$I_{перем} = I_{пост} I_{стр}$$

Если в индексах средних уровней в качестве весов используются удельные веса единиц совокупности в общей численности совокупности, то есть показатели доли ($d = f/\Sigma f$), то система индексов может быть записана в следующем виде:

$$I_{перем} = \frac{\sum x_1 d_1}{\sum x_0 d_0} \qquad I_{пост} = \frac{\sum x_1 d_1}{\sum x_0 d_1} \qquad I_{стр} = \frac{\sum x_0 d_1}{\sum x_0 d_0}$$

Система индексов переменного, постоянного состава и структурных сдвигов строится для изучения динамики среднего уровня цен, себестоимости, фондоотдачи, рентабельности, производительности труда, заработной платы и других вторичных признаков.

Помимо мультипликативной модели, на основе индексов переменного, постоянного состава и структурных сдвигов может быть построено и *аддитивное разложение.* Так, общий абсолютный прирост (уменьшение) среднего уровня признака в целом по совокупности находится по формуле:

$$\Delta\bar{x} = \bar{x}_1 - \bar{x}_0 = \frac{\sum x_1 f_1}{\sum f_1} - \frac{\sum x_0 f_0}{\sum f_0} \text{ или } \Delta\bar{x} = \sum x_1 d_1 - \sum x_0 d_0$$

Абсолютный прирост (уменьшение) среднего уровня признака в целом по совокупности за счет изменения значений изучаемого признака у отдельных единиц совокупности и за счет структурных изменений рассчитывается соответственно как разность числителей и знаменателей индексов постоянного состава и структурных сдвигов:

$$\Delta\bar{x}_{(x)} = \frac{\sum x_1 f_1}{\sum f_1} - \frac{\sum x_0 f_1}{\sum f_1} \text{ или } \Delta\bar{x} = \sum x_1 d_1 - \sum x_0 d_1$$

$$\Delta\bar{x}_{(f)} = \frac{\sum x_1 f_1}{\sum f_1} - \frac{\sum x_1 f_0}{\sum f_0} \text{ или } \Delta\bar{x} = \sum x_1 d_1 - \sum x_1 d_0$$

В общем виде аддитивное разложение имеет вид:

$$\Delta\bar{x} = \Delta_{\bar{x}_{(x)}} + \Delta_{\bar{x}_{(f)}}$$

ГЛАВА 8 ВЫБОРОЧНОЕ НАБЛЮДЕНИЕ

8.1 Выборочное наблюдение, его сущность и значение

Выборочное наблюдение – это вид несплошного наблюдения, 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.

Этапы подготовки и проведения выборочного наблюдения:

1 этап – определение цели обследования;

2 этап – установление границ генеральной совокупности;

3 этап – составление программы наблюдения и программы разработки данных;

4 этап – определение вида выборки, процента отбора и метода отбора;

5 этап – отбор и регистрация наблюдаемых признаков у отобранных единиц;

6 этап – расчет выборочных характеристик и их ошибок;

7 этап – распространение полученных результатов на генеральную совокупность.

Генеральная совокупность – это совокупность, из которой производится отбор единиц совокупности.

Выборочная совокупность – это совокупность отобранных в определенном порядке единиц, по которым собирается информация.

Основные характеристики генеральной и выборочной совокупностей приведены в таблице 73.

Таблица 73 − Основные характеристики генеральной и выборочной совокупностей

Характеристика	Генеральная совокупность	Выборочная совокупность
Объем совокупности (число единиц)	N	n
Численность единиц, обладающих обследуемых признаком	M	m

Средний размер признака	$\bar{x} = \frac{\sum x_i}{N}$	$\tilde{x} = \frac{\sum x_i}{n}$
Доля единиц, обладающих обследуемым признаком	$p = \frac{M}{N}$	$\omega = \frac{m}{n}$
Дисперсия количественного признака	$\sigma_{\bar{x}}^2 = \frac{\left[\sum\left(x_i - \bar{x}\right)^2\right]}{N}$	$\sigma_{\tilde{x}}^2 = \frac{\left[\sum\left(x_i - \tilde{x}\right)^2\right]}{n}$
Дисперсия доли	$\sigma_p^2 = p(1-p)$	$\sigma_\omega^2 = \omega(1-\omega)$

8.2 Способы формирования выборочной совокупности

Отбор единиц в выборочную совокупность может быть повторным или бесповторным.

При *повторном отборе* попавшая в выборку единица подвергается обследованию, возвращается в генеральную совокупность и наравне с другими единицами участвует в дальнейшей процедуре отбора. 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. Число единиц, 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.

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 являются более точными по сравнению с результатами, основанными на повторной выборке.

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;
– комбинированная.

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.

После проведения отбора с использованием какого-либо алгоритма, реализующего принцип случайности, или на основе таблицы случайных чисел, необходимо определить границы генеральных характеристик.

Механическая (систематическая) выборка применяется в тех случаях, когда генеральная совокупность каким-либо образом упорядочена, т.е. имеется определенная последовательность в расположении единиц (например, номера домов, списки избирателей).

При проведении механического отбора устанавливается *шаг отсчета* (h), т.е. расстояние между отбираемыми единицами ($h = \frac{N}{n}$ – величина, обратная доле выборки), и *начало отсчета* – номер единицы, которая должна быть обследована первой.

Механический отбор всегда бывает *бесповторным*. При этом отборе используются формулы, применяемые при собственно случайном бесповторном отборе. Данный вид отбора имеет преимущество перед случайным отбором, его не только легче организовать, но при нем единицы выборочной совокупности равномернее распределяются в генеральной совокупности.

Типическая выборка. Используется в случаях, когда все единицы генеральной совокупности объединены в несколько крупных типических групп (страт). Отбор единиц в выборочную совокупность из каждой типической группы осуществляется собственно случайным или механическим способом.

Отбор единиц в типическую выборку может быть организован либо пропорционально объему типических групп, либо пропорционально внутригрупповой вариации (дифференциации) признака.

Объем выборки из типической группы при отборе, пропорциональном численности единиц типических групп, определяется по формуле

$$n_i = n \times \frac{N_i}{N},$$

где n_i – объем выборки из i-й типической группы;

N_i – объем i-й типической группы в генеральной совокупности.

Так как в типическую выборку должны попасть представители всех групп, средняя ошибка типической выборки зависит только от средней из внутригрупповых дисперсий $\overline{\sigma}_i^2$ или $\overline{\omega(1-\omega)}$, а не от общей дисперсии $\sigma_{\tilde{x}}^2$ или $\omega \times (1-\omega)$.

Серийная выборка. 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. При серийном отборе внутри отобранных групп обследуются все без исключения единицы.

Поскольку внутри серий обследуются все без исключения единицы, средняя ошибка выборки при отборе равновеликих серий зависит от величины только *межгрупповой дисперсии* ($\delta_{\tilde{x}}^2$ или δ_{ω}^2):

межгрупповая дисперсия средних $\delta_{\tilde{x}}^2 = \frac{\sum(\tilde{x}_i - \tilde{x})^2}{r}$,

где $\tilde{x}_i$ – средняя i-й серии;

$\tilde{x}$ – средняя по всей выборочной совокупности;

r – число отобранных серий;

межгрупповая дисперсия доли $\delta_{\omega}^2 = \frac{\sum(\omega_i - \overline{\omega})^2}{r}$,

где ω_i – доля признака i-й серии;

$\overline{\omega}$ – общая доля признака во всей выборочной совокупности.

Комбинированная выборка. Предполагает использование нескольких способов выборки.

При проведении выборочного наблюдения неизбежна некоторая свойственная ему погрешность (ошибки). Классификация ошибок выборочного наблюдения представлена на рисунке 11.

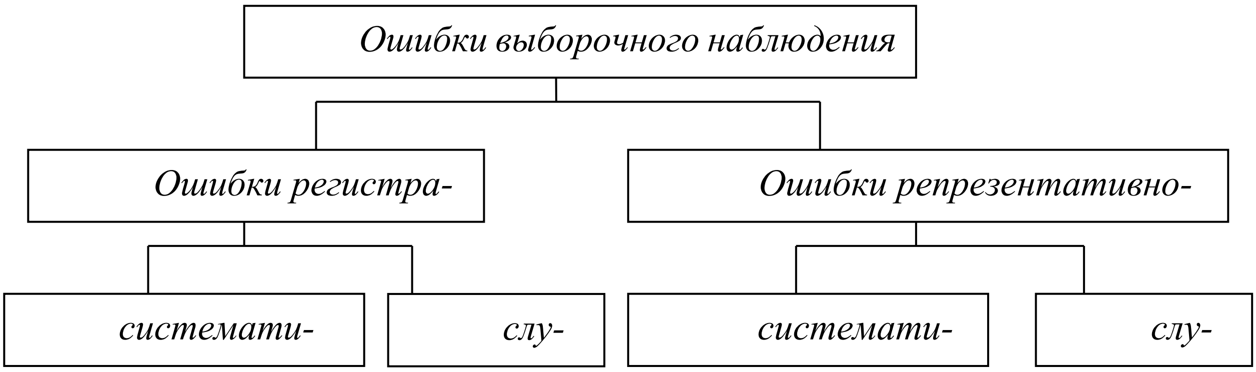

Рисунок 11 – Классификация ошибок выборочного наблюдения

Ошибки регистрации являются следствием неправильного установления значения наблюдаемого признака или неправильной записи. Случайные ошибки регистрации – это результат действия различных случайных факторов (например, перепутаны строки или графы при заполнении статистического формуляра). Такие ошибки имеют разную направленность: они могут и повышать, и понижать значения показателей.

Систематические ошибки регистрации всегда имеют одинаковую тенденцию либо к увеличению, либо к уменьшению значения показателей по каждой единице наблюдения, и поэтому величина показателя по совокупности в целом будет включать в себя накопленную ошибку.

Ошибки репрезентативности характерны только для несплошного наблюдения и обусловлены тем, что выборочная совокупность не может по всем параметрам в точности воспроизвести генеральную совокупность. Получаемые расхождения отражают, в какой степени попавшие в выборку единицы могут представлять всю генеральную совокупность.

Систематические ошибки репрезентативности связаны с нарушением принципов формирования выборочной совокупности.

Случайные ошибки репрезентативности обусловлены действием случайных факторов, не содержащих каких-либо элементов системности в направлении воздействия на рассчитываемые выборочные характеристики.

Так как случайная ошибка выборки возникает вследствие случайных различий между границами выборочной и генеральной совокупностей, то при достаточно большом объеме выборки эта ошибка будет сколь угодно мала. Случайные ошибки могут быть доведены до незначительных размеров, и их размеры и пределы можно определить с достаточной точностью на основании закона больших чисел.

Средняя (стандартная) ошибка выборки представляет собой такое расхождение между средним выборочной и генеральной совокупностей $(\tilde{x} - \bar{x})$, которое не превышает $\pm\sigma$.

Средняя ошибка выборки *зависит от объема выборки* (чем больше численность при прочих равных условиях, тем меньше величина средней ошибки выборки) и *степени варьирования признака* (чем меньше вариация признака, следовательно, и дисперсия, тем меньше ошибка выборки, и наоборот). Формулы расчета средней ошибки выборки представлены в таблице 74.

Таблица 74 – Средняя ошибка в зависимости от способа формирования выборочной совокупности

Вид выборки	Повторный отбор		Бесповторный отбор	
	для средней ($\mu_{\tilde{x}}$)	для доли (μ_{ω})	для средней ($\mu_{\tilde{x}}$)	для доли (μ_{ω})
Собственно-случайная и механическая	$\frac{\sigma}{\sqrt{n}}$	$\sqrt{\frac{w(1-w)}{n}}$	$\sqrt{\frac{\sigma^2}{n}\left(1-\frac{n}{N}\right)}$	$\sqrt{\frac{w(1-w)}{n}\left(1-\frac{n}{N}\right)}$
Типическая (при отборе пропорциональном объему групп)	$\sqrt{\frac{\bar{\sigma}^2}{n}}$	$\sqrt{\frac{\overline{w(1-w)}}{n}}$	$\sqrt{\frac{\bar{\sigma}^2}{n}\left(1-\frac{n}{N}\right)}$	$\sqrt{\frac{\overline{w(1-w)}}{n}\left(1-\frac{n}{N}\right)}$

Серийная (гнездовая)	$\sqrt{\frac{\delta^2_{\tilde{x}}}{r}}$	$\sqrt{\frac{\delta^2_{\omega}}{r}}$	$\sqrt{\frac{\delta^2_{\tilde{x}}}{r}\left(\frac{R-r}{R-1}\right)}$	$\sqrt{\frac{\delta^2_{\omega}}{r}\left(\frac{R-r}{R-1}\right)}$

Предельная ошибка выборки – максимально возможное расхождение выборочной и генеральной средних $(\tilde{x}-\bar{x})$, т.е. максимум ошибки при заданной вероятности ее появления.

О величине предельной ошибки можно судить с определенной вероятностью, на величину которой указывает коэффициент доверия (t).

Предельная ошибка выборки (Δ) определяется по формуле

$$\Delta_{\tilde{x}} = t\mu_{\tilde{x}} \text{ или } \Delta_{\omega} = t\mu_{\omega},$$

где Δ – предельная ошибка выборки;

t – коэффициент доверия, зависящий от вероятности, с которой гарантируется предельная ошибка выборки (см. приложение 5).

Предельная ошибка выборки позволяет определять предельные значения характеристик генеральной совокупности при заданной вероятности и их доверительные интервалы.

Интервальная оценка генеральной совокупности

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

т.е., с заданной вероятностью можно утверждать, что значение генеральной средней можно ожидать в пределах от $\tilde{x} - \Delta_{\tilde{x}}$ до $\tilde{x} + \Delta_{\tilde{x}}$.

Интервальная оценка генеральной доли

$$\omega - \Delta_{\omega} \le p \le \omega + \Delta_{\omega}.$$

Пример. Для определения среднего возраста рабочих предприятия была проведена 5%-я механическая выборка. В результате обследования получены данные, приведенные в таблице 75.

Таблица 75 – Данные обследования для определения среднего *возраста рабочих предприятия*

Возраст рабочего, лет	18-28	28-38	38-48	Свыше 48
Число рабочих, человек	10	50	35	5

С вероятностью 0,954 нужно определить:

1) пределы, в которых находится средний возраст рабочих предприятия;

2) пределы, в которых находится доля рабочих старше 48 лет.

Решение.

1. Рассчитаем среднюю арифметическую и выборочную дисперсию по данным таблицы 76.

Таблица 76 – Данные для расчета

Границы интервалов, лет	Число рабочих, чел. (n_i)	Середина интервала, лет (x_i)	$x_i n_i$	$x_i^2 n_i$
18−28	10	23	230	5290
28−38	50	33	1650	54450
38−48	35	43	1505	64715
Свыше 48	5	53	265	14045
Итого	100	-	3650	138500

Так как численность выборочной совокупности $n = \sum n_i$, значение выборочной средней равно:

$$\tilde{x} = \frac{\sum x_i n_i}{n} = \frac{3650}{100} = 36{,}5 \text{ лет}$$

Выборочная дисперсия равна:

$$\sigma_{\tilde{x}}^2 = \frac{\sum x_i^2 n_i}{\sum n_i} - \left(\tilde{x}\right)^2 = \frac{138500}{100} - (36{,}5)^2 = 52{,}75$$

Стандартная ошибка средней равна:

$$\mu_{\tilde{x}} = \sqrt{\sigma_{\tilde{x}}^2 \div n \times \left(1 - n \div N\right)} = \sqrt{52{,}75 \div 100 \times (1 - 0{,}05)} = 0{,}7 \text{ года}$$

Значение предельной ошибки выборки для средней при $F(t) = 0{,}954$ и коэффициенте доверия $t = 2$ составит (см. приложение 5):

$$\Delta_{\tilde{x}} = \mu_{\tilde{x}} \times t = 0{,}7 \times 2 = 1{,}4 \text{ года}$$

Пределы, в которых находится средний возраст рабочих предприятия, следующие:

$$\tilde{x} - \Delta_{\tilde{x}} \le \bar{x} \le \tilde{x} + \Delta_{\tilde{x}}$$
$$36{,}5 - 1{,}4 \le \bar{x} \le 36{,}5 + 1{,}4$$
$$35{,}1 \le \bar{x} \le 37{,}9$$

Значит, с вероятностью 0,954 можно утверждать, что средний возраст рабочих предприятия, находится в интервале от 35,1 лет до 37,9 лет.

2. Выборочная доля составляет:

$$\omega = m \div n = 5 \div 100 = 0{,}05$$

Стандартная ошибка доли равна:

$$\mu_{\omega} = \sqrt{\frac{\omega(1-\omega)}{n}\left(1 - \frac{n}{N}\right)} = \sqrt{\frac{0{,}05(1-0{,}05)}{100}\left(1 - 0{,}05\right)} = 0{,}021$$

Значение предельной ошибки выборки для средней при $F(t) = 0{,}954$ и коэффициенте доверия $t = 2$ (см. приложение 5) составит:

$$\Delta_{\omega} = \mu_{\omega} \times t = 0{,}021 \times 2 = 0{,}042$$

Границы, в которых находится доля рабочих предприятия следующие:

$$\omega - \Delta_{\omega} \le p \le \omega + \Delta_{\omega}$$

$$0{,}05-0{,}042 \le p \le 0{,}05+0{,}042$$
$$0{,}008 \le p \le 0{,}092$$

С вероятностью 0,954 можно утверждать, что доля рабочих старше 48 лет будет находиться в пределах от 0,8 до 9,2 %.

8.3 Определение необходимого объема выборочной совокупности

При подготовке выборочного наблюдения с заранее заданным значением допустимой ошибки выборки важно правильно определить объем (численность) выборочной совокупности. Объем выборки (n) должен быть достаточным для того, чтобы обеспечить репрезентативность выборки.

Формула необходимой численности выборки для различных способов отбора выводится из формулы предельной ошибки выборки и представлена в таблице 77.

Таблица 77 – Необходимый объем (численность) выборки

Вид выборки	Повторный отбор		Бесповторный отбор	
	при определении среднего размера признака	при определении доли признака	при определении среднего размера признака	при определении доли признака
Собственно случайная и механическая	$\frac{t^2\sigma_{\tilde{x}}^{\ 2}}{\Delta_{\tilde{x}}^2}$	$\frac{t^2\omega(1-\omega)}{\Delta_{\omega}^2}$	$\frac{t^2\sigma_{\tilde{x}}^2 N}{\Delta_{\tilde{x}}^2 N + t^2\sigma_{\tilde{x}}^2}$	$\frac{t^2\omega(1-\omega)N}{\Delta_{\omega}^2 N + t^2\omega(1-\omega)}$
Типическая	$\frac{t^2\overline{\sigma}_i^{\ 2}}{\Delta_{\tilde{x}}^2}$	$\frac{t^2\overline{\omega(1-\omega)}}{\Delta_{\omega}^2}$	$\frac{t^2\overline{\sigma}_i^2 N}{\Delta_{\tilde{x}}^2 N + t^2\overline{\sigma}_i^2}$	$\frac{t^2\overline{\omega(1-\omega)}N}{\Delta_{\omega}^2 N + t^2\overline{\omega(1-\omega)}}$
Серийная	$\frac{t^2\delta_{\tilde{x}}^{\ 2}}{\Delta_{\tilde{x}}^2}$	$\frac{t^2\delta_{\omega}^{\ 2}}{\Delta_{\omega}^2}$	$\frac{t^2\delta_{\tilde{x}}^{\ 2} R}{\Delta_{\tilde{x}}^2 R + t^2\delta_{\tilde{x}}^2}$	$\frac{t^2\delta_{\omega}^{\ 2} R}{\Delta_{\omega}^2 R + t^2\delta_{\omega}^2}$

Пример. Для определения среднего дохода на душу населения необходимо провести выборочное наблюдение способом случайного отбора. В городе проживает 15 тыс. семей, среднее квадратическое отклонение составляет 60 руб. Определим необходимую численность выборки при условии, что с вероятностью 0,954 ошибка выборки не превысит 10 руб.

Численность выборки при случайном отборе определяется по формуле

$$n = \frac{t^2 \sigma_{\tilde{x}}^2 N}{\Delta_{\tilde{x}}^2 N + t^2 \sigma_{\tilde{x}}^2}.$$

Коэффициент доверия, зависящий от вероятности, с которой гарантируется предельная ошибка выборки, устанавливается по приложению 5: при $F(t) = 0{,}954$, $t = 2{,}0$.

Подставим данные в формулу:

$$n = \frac{t^2 \sigma_{\tilde{x}}^2 N}{\Delta_{\tilde{x}}^2 N + t^2 \sigma_{\tilde{x}}^2} = \frac{2^2 \times 60^2 \times 15000}{10^2 \times 15000 + 2^2 \times 60^2} = 142{,}6 \approx 143 \text{ семьи}$$

Таким образом, для наблюдения необходимо отобрать 143 семьи. Это позволит определить средний доход на душу населения с ошибкой не более 10 руб. Вероятность, что ошибка не превысит 10 руб., равна 0,954.

8.4 Распределение результатов выборочного наблюдения на генеральную совокупность

Конечная цель любого выборочного наблюдения – распространение его характеристик на генеральную совокупность. Существуют два способа распространения данных выборочного наблюдения на генеральную совокупность: прямого пересчета и поправочных коэффициентов.

Способ прямого пересчета применяется в том случае, если цель выборочного наблюдения заключается в определении объема признака генеральной совокупности, когда известна лишь численность ее единиц.

Пример. При выборочном обследовании партии нарезных батонов в 2000 ед. доля нестандартных изделий в выборке составляет: $\omega = 0{,}1$ (10:100) при установленной с вероятностью $F(t) = 0{,}954$ предельной ошибке выборки $\Delta_\omega = \pm 0{,}06$.

На основе этих данных доля нестандартных изделий во всей партии составит: $p = 0{,}1 \pm 0{,}06$, или от 0,04 до 0,16.

Способом прямого пересчета можно определить пределы абсолютной численности нестандартных изделий во всей партии:

минимальная численность, шт.: $2000 \times 0{,}04 = 80$;

максимальная численность, шт.: $2000 \times 0{,}16 = 320$.

Способ поправочных коэффициентов применяется в тех случаях, когда цель выборочного метода состоит в уточнении результатов сплошного наблюдения.

Пример. По данным годовых отчетов в сельскохозяйственных организациях имеется 127500 голов коров. В целях проверки данных сплошного учета провели контрольные обходы части обследованных предприятий и выявили, что если данные сплошного учета в организациях, попавших в выборку, показали 550 голов коров, то данные выборки в этих же организациях – 560 голов коров. Необходимо определить фактическую численность коров в сельскохозяйственных организациях.

Определим «процент недоучета» (коэффициент) при сплошном наблюдении:

$$\frac{560 - 550}{550} \times 100 = 1{,}82\,\%$$

Количество голов коров необходимо умножить на этот коэффициент:

$$127500 \times 1{,}0182 = 129821 \text{ голов}$$

Это значит, что при сплошном учете недоучтена 2321 голова коров.

ГЛАВА 9 СТАТИСТИЧЕСКОЕ ИЗУЧЕНИЕ ВЗАИМОСВЯЗЕЙ СОЦИАЛЬНО-ЭКОНОМИЧЕСКИХ ЯВЛЕНИЙ

9.1 Основные виды взаимосвязей

Социально-экономические явления представляют собой результат одновременного воздействия большого числа разнообразных и взаимосвязанных факторов. Понять и изучить какое-либо явление можно, исследуя его во взаимосвязи с окружающими признаками. В статистике различают факторные и результативные признаки. *Факторные признаки* обусловливают изменения других, связанных с ними признаков. *Результативные признаки* изменяются под действием факторных признаков.

В статистике различают:

– *функциональную связь*, при которой определенному значению факторного признака соответствует одно и только одно значение результативного признака. Функциональная связь проявляется во всех случаях наблюдения и для каждой конкретной единицы исследуемой совокупности;

– *стохастическую зависимость*, при которой причинная зависимость проявляется не в каждом отдельном случае, а в общем, среднем при большом числе наблюдений. Частным случаем стохастической зависимости является корреляционная связь, при которой изменение среднего значения результативного признака обусловлено изменением факторных признаков.

Для *функциональной связи* характерны следующие особенности:

1) каждому значению величины факторного признака соответствует только *одно* или *несколько точно определенных значений* результативного признака;

2) эта связь обычно *выражается формулами*, что в большей степени присуще точным наукам;

3) функциональная зависимость *с одинаковой силой* проявляется у всех единиц совокупности;

4) она является *полной и точной*, так как обычно известны перечень всех факторов и механизм их воздействия на результативный признак (в виде уравнения).

Корреляционные связи имеют следующие особенности:

1) средняя величина результативного признака меняется под влиянием изменения *многих факторных признаков*, ряд из которых может быть неизвестен;

2) разнообразие факторов, их взаимосвязи и противоречивое действие вызывают *широкое варьирование результативного признака*;

3) корреляционные связи обнаруживаются не в единичных случаях, а в массе, для их исследования требуются *массовые наблюдения*;

4) связь между признаками-факторами и результативным признаком *неполная*, а проявляется в общем, среднем.

В *зависимости от направления действия* функциональные и корреляционные связи делят на прямые и обратные; *по аналитическому выражению* – на прямолинейные и криволинейные.

При наличии *прямых связей* направление изменения результативного признака совпадает с направлением изменения признака-фактора. С увеличением (уменьшением) значений факторного признака происходит увеличение (уменьшение) результативного признака.

Обратные связи характеризуются тем, что направление изменения результативного признака не совпадает с направлением изменения признака-фактора. С увеличением (уменьшением) значений факторного признака происходит уменьшение (увеличение) результативного признака.

Прямолинейные связи выражаются уравнением прямой линии. При наличии этих связей с возрастанием величины факторного признака происходит непрерывное возрастание (или убывание) величин результативного признака.

Криволинейные связи выражаются уравнениями кривых линий – гиперболой, параболой и другими, поскольку с возрастанием величины факторного признака возрастание (или убывание) результативного признака происходит неравномерно, или направление его изменения меняется на обратное.

Корреляционные связи *в зависимости от количества признаков, включенных в модель*, делятся на однофакторные (парные) и многофакторные (множественные).

Однофакторные (парные) связи отражают зависимость между одним признаком-фактором и результативным признаком (при абстрагировании от влияния других признаков).
Многофакторные (множественные) связи характеризуются зависимостью между несколькими факторными признаками и результативным признаком (факторы действуют комплексно, т.е. одновременно и во взаимосвязи).
Для выражения функциональных связей применяют балансовый и индексный методы.
Метод балансовых построений широко используют для анализа связей и пропорций в экономике. Статистический баланс представляет собой систему показателей, которая состоит из двух сумм абсолютных величин, связанных знаком равенства:

$$A+C=D+E$$

Посредством балансов связывают в единую систему абсолютные величины, показывающие движение ресурсов. Суммы показателей образуют систему величин, характеризующих размер ресурсов на начало периода, поступление и выбытие, размер ресурсов на конец периода.
При рассмотрении корреляционных связей используют:

– методы взаимной сопряженности – для изучения связи между атрибутивными (качественными) признаками;
– метод параллельных рядов;
– графический метод (корреляционного поля);
– табличный метод (корреляционной таблицы);
– метод аналитических группировок;
– корреляционно-регрессионный анализ и другие методы – для выявления связей между количественными (варьирующими) признаками.

Изучение и оценка связей между атрибутивными (качественными) признаками в статистике осуществляется с использованием методов взаимной сопряженности (непараметрических методов оценки связи).
Методы взаимной сопряженности строятся на применении следующих показателей:

– коэффициента ассоциации;
– коэффициента контингенции;
– биссериального коэффициента корреляции;
– коэффициента взаимной сопряженности А.А. Чупрова;
– коэффициента взаимной сопряженности Пирсона.

Коэффициент ассоциации ($K_{\text{а}}$) применяется для изучения связи между качественными признаками в том случае, когда каждый признак принимает только два значения (состоит только из двух групп), т.е. позволяет изучить связь между альтернативными признаками. Коэффициент вычисляется по формуле

$$K_{\text{а}} = \frac{ad - bc}{ad + bc},$$

где a, b, c, d − частоты «таблицы четырех полей» (таблица 83).

Коэффициент ассоциации изменяется в пределах от (−1) до (+1). Чем ближе этот показатель к 1 или (−1), тем сильнее связаны между собой изучаемые признаки. Если коэффициент ассоциации не ниже 0,3, можно говорить о наличии существенной связи между признаками.

Коэффициент контингенции ($K_{\text{к}}$) определяется по формуле

$$K_{\text{к}} = \frac{ad - bc}{\sqrt{(a+b)\cdot(b+d)\cdot(a+c)\cdot(c+d)}}$$

Коэффициент контингенции применяются в том случае, когда хотя бы одно значение из четырех показателей в «таблице четырех полей» отсутствует.

Таблица 83 − Таблица для вычисления коэффициентов ассоциации и контингенции

a	b	$a+b$
c	d	$c+d$
$a+c$	$b+d$	$a+b+c+d$

По абсолютной величине коэффициент контингенции всегда меньше коэффициента ассоциации и изменяется от (−1) до (+1). Чем ближе к 1 или (−1), тем сильнее связаны между собой изучаемые признаки.

Пример. В результате исследования крупнейших страховых компаний получены нижеследующие данные (таблица 84).

Таблица 84 – Группировка страховых компаний по классу устойчивости

Группа страховых компаний по классу устойчивости	По величине выплат, %		
	< 30	> 30	Итого
1−3	6	22	28
4−5	11	14	25
Итого	17	36	53

Определите степень тесноты связи меду признаками с использованием коэффициентов ассоциации и контингенции.

1. Коэффициент ассоциации определяется по формуле

$$K_{\text{а}} = \frac{ad - bc}{ad + bc} = \frac{6 \times 14 - 22 \times 11}{6 \times 14 + 22 \times 11} = \frac{-158}{326} = -0{,}48$$

Связь между признаками разнонаправленная, и степень тесноты существенная.

2. Коэффициент контингенции определяется по формуле

$$K_{\text{к}} = \frac{ad - bc}{\sqrt{(a+b)\cdot(b+d)\cdot(a+c)\cdot(c+d)}} =$$

$$= \frac{6 \times 14 - 22 \times 11}{\sqrt{(6+22) \times (6+11) \times (14+22) \times (14+11)}} = -0{,}24$$

Связь между признаками присутствует, и она разнонаправленная.

Биссериальный коэффициент корреляции (r) позволяет изучить связь между качественным альтернативным и количественным варьирующим признаками и определяется по формуле

$$r = \frac{\left|\overline{Y}_2 - \overline{Y}_1\right|}{\sigma_Y} \times \frac{pq}{Z},$$

где $\overline{Y}_2, \overline{Y}_1$ – средние значения признака в группах;

σ_Y – среднее квадратическое отклонение фактических значений признака от среднего уровня;

p – доля первой группы в совокупности;

q – доля второй группы;

Z – табличные значения Z-распределения в зависимости от p.

Величина коэффициента варьирует в пределах от 0 до 1.

Коэффициент взаимной сопряженности А.А.Чупрова применяется для измерения тесноты связи между варьированием двух атрибутивных признаков, когда это варьирование образует несколько (три и более) групп и определяется по формуле

$$K_{\text{ч}} = \sqrt{\frac{\chi^2}{\sqrt{(m_1 - 1) \times (m_2 - 1)}}} = \sqrt{\frac{\chi^2}{n \times \sqrt{(m_1 - 1) \times (m_2 - 1)}}},$$

где χ^2 – хи-квадрат;

$$\chi^2 = \left(\sum_1^{m_1} \sum_1^{m_2} \frac{f_{ij}^2}{f_i \times f_j} \right) - 1,$$

где f_i f_j – эмпирические частоты в i-й строке j-м столбце;

m – число групп по каждому признаку;

n – количество наблюдений.

Числовое значение коэффициента изменяется от 0 до 1, но уже при значении 0,3 можно говорить о тесной связи между вариацией изучаемых качественных признаков.

Пример. В результате исследования крупнейших страховых компаний получены нижеследующие данные (таблица 85).

Определите связь между признаками с использованием коэффициента взаимной сопряженности А.А. Чупрова.

Таблица 85 – Группировка страховых компаний по классу устойчивости

Группа страховых компаний по классу устойчивости	По величине выплат, %			
	< 30	30−40	> 40	Итого
1–2	2	6	10	18
3–4	14	9	7	30
5	1	2	2	5
Итого	17	17	19	53

Коэффициент взаимной сопряженности А.А. Чупрова определяется по формуле

$$K_ч = \sqrt{\frac{\chi^2}{\sqrt{(m_1 - 1)\times(m_2 - 1)}}}.$$

Для его расчета необходимо определить хи-квадрат по формуле:

$$\chi^2 = \left(\sum_1^{m_1}\sum_1^{m_2}\frac{f_{ij}^2}{f_i \times f_j}\right) - 1 =$$

$$= \left(\begin{array}{l}\dfrac{10^2}{19\times18} + \dfrac{14^2}{17\times30} + \dfrac{9^2}{17\times30} + \dfrac{2^2}{17\times18} + \dfrac{6^2}{17\times18} + \\ + \dfrac{7^2}{19\times30} + \dfrac{1^2}{17\times5} + \dfrac{2^2}{17\times5} + \dfrac{2^2}{19\times5} - 1\end{array}\right) = 0{,}153$$

Следовательно, коэффициент А.А. Чупрова равен:

$$K_{\text{ч}} = \sqrt{\frac{0{,}153}{\sqrt{(3-1)\times(3-1)}}} = 0{,}277.$$

Связь между признаками средней степени тесноты.

Коэффициент взаимной сопряженности Пирсона ($K_{\text{п}}$) определяется по формуле

$$K_n = \sqrt{\frac{\chi^2}{\chi^2 + n}},$$

где n – количество наблюдений.

Коэффициент изменяется от 0 до 1. Чем ближе к единице, тем теснее связь между атрибутивными признаками.

Метод приведения параллельных данных. Метод основан на сопоставлении двух или нескольких рядов статистических величин. Такое сопоставление позволяет установить наличие связи и получить представление о ее характере.

Взаимосвязь двух признаков изображается графически 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, обозначается точкой. При отсутствии тесных связей наблюдается беспорядочное расположение точек на графике. Чем сильнее связь между признаками, тем теснее будут группироваться точки вокруг определенной линии, выражающей форму связи.

На основе сравнения параллельных рядов можно применить элементарные показатели, характеризующие направление и тесноту связи:

– коэффициент Фехнера (коэффициент корреляции знаков);
– коэффициент Спирмена (коэффициент корреляции рангов);
– множественный коэффициент ранговой корреляции (коэффициент конкордации).

Коэффициент Фехнера ($K_ф$) основан на степени согласованности направлений отклонений индивидуальных значений факторного и результативного признаков от соответствующих средних величин. Для расчета этого показателя исчисляют средние значения факторного и результативного признаков (по арифметической простой), а затем проставляют знаки отклонений для значений взаимосвязанных пар признаков (если фактическое значение признака больше средней величины, то ставится знак «+», если меньше, то знак «−»). Коэффициент определяется по формуле следующего вида:

$$K_{\phi} = \frac{C - H}{C + H},$$

где C – число совпадений знаков отклонений;
H – число несовпадений знаков отклонений.

Значение коэффициента Фехнера находится в пределах от −1 до +1, т.е. $-1 \le К_{ф} \le +1$. Если: $К_{ф} = \pm 1$, то связь между между признаками функциональная; $К_{ф} = 0$, то связь отсутствует.
Данный показатель позволяет уловить направление вариации, но не учесть точно ее величину.
Пример. Имеются данные по совокупности стран о размере ВВП на душу населения по полной первоначальной стоимости основных фондов и фактическом конечном потреблении домашних хозяйств по полной первоначальной стоимости основных фондов (таблица 86).
Определите степень тесноты связи между признаками с использованием коэффициента Фехнера.
Решение.

1. Определим средние значения по факторному и результативному признакам (по арифметической простой):

$$\bar{x} = \frac{\sum x}{n} = \frac{58800}{8} = 7350$$

$$\bar{y} = \frac{\sum y}{n} = \frac{454}{8} = 56{,}75$$

2. Сопоставляя фактические значения с их средними, расставим знаки, например, 14400 больше 7350, значит, знак «+», 2390 меньше 7350, значит, знак «−». В результате получаем семь совпадений знаков и одно несовпадение.
3. Коэффициент Фехнера определяется по формуле

$$K_{ф} = \frac{C - H}{C + H} = \frac{7 - 1}{7 + 1} = 0{,}75$$

Можно предположить наличие тесной и однонаправленной связи между признаками.

Таблица 86 – Размер ВВП на душу населения по полной первоначальной стоимости основных фондов и фактическом конечном потреблении домашних хозяйств по полной первоначальной стоимости основных фондов

Страна	Исходные данные		Расчетные данные	
	ВВП на душу населения по полной первоначальной стоимости основных фондов, долл. (x)	фактическом конечном потреблении домашних хозяйств по полной первоначальной стоимости основных фондов к РФ, % (y)	(x)	(y)
РФ	14400	100	+	+
Белоруссия	10740	85	+-	+
Молдавия	2930	37	-	-
Украина	6810	59	-	+
Армения	5900	34	-	-
Азербайджан	6370	46	-	-
Казахстан	9700	69	+	+
Киргизия	1950	24	-	-

Коэффициент корреляции рангов (коэффициент Спирмэна) (ρ) применяют для анализа связи двух признаков (x, y). Он учитывает согласованность рангов, т.е. номеров, которые занимают единицы совокупности по каждому из этих признаков, и определяется по формуле

$$\rho = 1 - \frac{6\sum d^2}{n^3 - n} \text{ или } \rho = 1 - \frac{6\sum d^2}{n(n^2 - 1)},$$

где d – разность рангов x и y;
n – число наблюдаемых пар значений x и y.

Коэффициент Спирмэна принимает любые значения в интервале от (+1) (полная корреляция рангов, в этом случае $\sum d = 0$) до (−1) (полная обратная корреляция рангов, в этом случае $\frac{6\sum d^2}{n^3 - n} = 2$.

При ρ = 0, когда $\frac{6\sum d^2}{n^3 - n} = 1$, корреляция рангов отсутствует.

Значимость коэффициента корреляции рангов Спирмэна проверяется на основе t-критерия Стьюдента. Расчетное значение критерия определяется по формуле

$$t_{\text{р}} = \rho_{x/y}\sqrt{\frac{n-2}{1-\rho^2_{X/Y}}}$$

Значение коэффициента корреляции считается статистически существенным, если $t_{\text{р}} > t_{\text{кр}}$ *(α; k = n−2)*.

Пример. Определите степень тесноты связи между признаками с использованием коэффициента Спирмена на основании данных по совокупности стран ближнего зарубежья о размере ВВП на душу населения по полной первоначальной стоимости основных фондов и уровню безработицы, представленных в таблице 87.

Решение.

1. Расставим ранги в порядке возрастания величин факторного и результативного признаков и рассчитаем $\sum d^2$:

$$\sum d^2 = \sum (x - y)^2 = 148.$$

2. Определим коэффициент Спирмена:

$$\rho = 1 - \frac{6\sum d^2}{n(n^2 - 1)} = 1 - \frac{6 \times 148}{8 \times (8^2 - 1)} = -0{,}76.$$

Связь между признаками довольно тесная и разнонаправленная.

Таблица 87 − Размер ВВП на душу населения по полной первоначальной стоимости основных фондов и уровень безработицы в странах ближнего зарубежья

Страна	Исходные данные		Расчетные данные		$\sum(x-y)^2$
	ВВП на душу населения по полной первоначальной стоимости основных фондов, долл.	уровень безработицы, %	ранг по ВВП на душу населения по полной первоначальной стоимости основных фондов, долл. (x)	ранг по уровню безработицы, % (y)	
РФ	14400	13,4	8	3	25
Белоруссия	10740	7,7	7	1	36
Молдавия	2930	25,8	2	5	9
Украина	6810	29,3	5	7	4
Армения	5900	26,5	3	6	9
Азербайджан	6370	15,8	4	4	0
Казахстан	9700	12,7	6	2	16
Киргизия	1950	35,0	1	8	49
Итого	-	-	-	-	148

Множественный коэффициент ранговой корреляции (коэффициент конкордации) (ω) применяется для оценки тесноты связи между несколькими (тремя и более) признаками при использовании ранговой корреляции. Коэффициент изменяется в пределах от 0 до 1 и характеризует степень тесноты связи, но уже при значении 0,5 можно говорить о тесной связи между вариацией изучаемых признаков.

Множественный коэффициент ранговой корреляции (коэффициент конкордации) определяется по формуле

$$\omega = \frac{12S}{m^2 \cdot (n^3 - n)},$$

где m – количество факторов;

n – число наблюдений;

S – отклонение суммы квадратов рангов от средней квадратов рангов.

$$S=\sum_{1}^{n}\left(\sum_{1}^{m}r_{ij}\right)^{2}-\frac{\left(\sum_{1}^{n}\sum_{1}^{m}r_{ij}\right)^{2}}{n},$$

где r_{ij} – ранг i-го фактора у j-й единицы.

Значимость коэффициента конкордации проверяется на основе χ^2-критерия Пирсона:

$$\chi^{2}=\frac{12S}{m\cdot n(n-1)}.$$

Если фактическое значение хи-квадрата больше табличного значения (приложение 4), при вероятности $L=0{,}05(0{,}01;\,0{,}10)$ и числе степеней свободы $\nu=\eta-1$, то это подтверждает значимость коэффициента конкордации.
В случае наличия связанных рангов коэффициент конкордации определяется по формуле

$$\omega=\frac{12S}{m^{2}\cdot\left(n^{3}-n\right)-m\cdot\sum_{j=1}^{m}Tj},$$

где $Tj=\dfrac{1}{12}\sum_{j=1}^{m}\left(t_{j}^{3}-t_{i}\right);$

t – количество связанных рангов по отдельным показателям.

Проверка значимости осуществляется по формуле

$$\chi^{2}=\frac{12S}{m\cdot n(n+1)-\frac{1}{n-1}\cdot\sum_{j=1}^{m}Tj}.$$

Ранговые коэффициенты корреляции Спирмэна и конкордации имеют преимущество, что с помощью их можно измерять и оценивать связи как между количественными, так и между атрибутивными признаками, которые поддаются ранжированию.

Связь между признаками можно наглядно увидеть, если построить *график*, отложив на оси абсцисс значения факторного признака (x), а на оси ординат – значения результативного признака (y). Нанеся на графике точки, соответствующие значениям x и y, можно получить корреляционное поле, по характеру расположения точек в котором можно судить о направлении и силе связи. Если точки беспорядочно разбросаны по всему полю, это говорит об отсутствии зависимости между двумя признаками. Если точки концентрируются вокруг оси, идущей от нижнего левого угла в верхний правый, то имеется прямая зависимость между варьирующими признаками. Если точки будут концентрироваться вокруг оси, идущей от верхнего левого угла в нижний правый, то существует обратная зависимость.

Табличный метод (корреляционной таблицы) применяется при наличии большого числа значений результативного признака, соответствующих одному значению факторного признака. Для построения корреляционной таблицы сначала осуществляют группировку совокупности по факторному и результативному признакам (комбинированную группировку), а затем в графах таблицы располагают группы по факторному признаку. Числа, расположенные на пересечении граф и строк, показывают частоту повторения данного сочетания значений x и y.

Направление расположения частот корреляционной таблицы дает возможность предполагать наличие или отсутствие связи, а также ее направление.

Если частоты располагаются по диагонали из левого верхнего угла в правый нижний угол таблицы (т.е. с ростом значений факторного признака растут значения результативного), то можно предполагать наличие прямой однонаправленной связи. Если частоты располагаются по другой диагонали, то предполагают наличие прямой обратной связи.

Пример. Постройте корреляционную таблицу на основании данных о численности врачей (на 10000 человек) и заболеваемости населения (на 1000 человек) по совокупности регионов РФ в 2007 г. представленных в таблице 88.

Таблица 88 − Численность врачей и заболеваемость по совокупности регионов РФ в 2007 г.

Регион	Численность врачей на 10000 человек населения, чел. (*x*)	Заболеваемость на 1000 человек населения, чел. (*y*)
1	8,2	506,9
2	8,4	820,3
3	11,8	558,0
4	13,8	684,0
5	14,2	824,5
6	14,7	990,3
7	15,7	886,0
8	17,0	1010,8
9	17,5	1056,9
10	17,8	904,8
11	18,2	793,2
12	18,9	914,0
13	19,0	910,7
14	20,9	1214,9

Решение.

1. Проведем группировку по факторному признаку:

$$n = 4,$$

$$h = \frac{20{,}9 - 8{,}2}{4} = 3{,}175.$$

Границы групп:
1-я группа: 8,200 − 11,375
2-я группа: 11,375 – 14,550
3-я группа: 14,550 – 17,725
4-я группа: 17,725 – 20,900
2. Проведем группировку по результативному признаку:

$$n = 4$$

$$h = \frac{1214{,}9 - 506{,}9}{4} = 177.$$

Границы групп:
1-я группа: 506,9 – 683,9
2-я группа: 683,9 – 860,9
3-я группа:860,9 – 1037,9
4-я группа:1037,9 – 1214,9
3. Сформируем корреляционную таблицу.

Таблица 89 – Корреляционная таблица

	8,2 – 11,375	11,375 – 14,550	14,550 – 17,725	17,725 – 20,900	$\sum f_y$
506,9 – 683,9	1	1			2
683,9 – 860,9	1	2		1	4
860,9 – 1037,9			3	3	6
1037,9 – 1214,9			1	1	2
$\sum f_x$	2	3	4	5	14

Частоты в таблице 89 расположились по диагонали из левого верхнего угла в правый нижний угол таблицы (с ростом значений факторного признака растут значения результативного), т.е. имеет место быть прямая однонаправленная связь.

Существуют ситуации, когда все клетки корреляционной таблицы заполнены. Однако это не говорит об отсутствии связи. В этих ситуациях необходимо установить, как расположена в таблице основная масса случаев. Для этого в каждой графе рассчитывают средние значения результативного признака, соответствующие определенному значению фактора (по формуле арифметической средней взвешенной). По изменению размера значений результативного признака судят о форме и направлении связи.

9.2 Сущность КРА и этапы его проведения

Корреляционно-регрессионный анализ *предполагает установление аналитической формы связи (регрессионный анализ) и измерение тесноты, направление связи (корреляционный анализ)*.

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 (факторов), а множество всех прочих факторов, также оказывающих влияние на зависимую величину, принимается за постоянные и средние значения. Регрессия может быть однофакторной (парной) и многофакторной (множественной).

По форме зависимости различают: линейную регрессию, нелинейную регрессию.

По направлениям связи различают:

– прямую регрессию (положительную), возникающую при условии, что если с увеличением или уменьшением независимой величины значения зависимой также соответственно увеличиваются или уменьшаются;

– обратную (отрицательную) регрессию, проявляющуюся при условии, что с увеличением или уменьшением независимой величины зависимая соответственно уменьшается или увеличивается.

Парная регрессия характеризует связь между двумя признаками – результативным и факторным. Аналитическая связь между ними описывается следующими уравнениями:

прямой $\bar{y}_x = a_0 + a_1 x$;

параболой $\bar{y}_x = a_0 + a_1 x + a_2 x^2$;

гиперболой $\bar{y}_x = a_0 + \dfrac{a_1}{x}$ и т.д.

Оценка параметров уравнений регрессии (a_0, a_1 и a_2 в уравнении параболы второго порядка) осуществляется методом наименьших квадратов, в основе которого лежит предположение о независимости наблюдений исследуемой совокупности.

Если результативный и факторный признаки возрастают одинаково, примерно в арифметической прогрессии, то это свидетельствует о том, что связь между ними линейная, а при обратной связи – гиперболическая. Если факторный признак увеличивается в арифметической прогрессии, а результативный – значительно быстрее, то используется связь параболическая или степенная.
В уравнениях регрессии параметр:
a_0 – показывает усредненное влияние на результативный признак неучтенных (не выделенных для исследования) факторов;
a_1 (а в уравнениях параболы a_2) – коэффициент регрессии который показывает, насколько изменяется в среднем значение результативного признака при увеличении факторного на единицу собственного измерения.
Система нормальных уравнений для нахождения параметров линейной парной регрессии методом наименьших квадратов имеет следующий вид:

$$\left.\begin{aligned} &na_0 + a_1\sum x = \sum y \\ &a_0\sum x + a_1\sum x^2 = \sum xy \end{aligned}\right\},$$

где n – объем исследуемой совокупности (число единиц наблюдения).

Если связь между признаками X и Y криволинейная и описывается уравнением параболы второго порядка, то система нормальных уравнений имеет вид

$$\begin{cases} na_0 + a_1\sum x + a_2\sum x^2 = \sum y \\ a_0\sum x + a_1\sum x^2 + a_2\sum x^3 = \sum xy \\ a_0\sum x^2 + a_1\sum x^3 + a_2\sum x^4 = \sum x^2 y \end{cases}$$

Система нормальных уравнений для нахождения параметров гиперболы

$$na_0 + a_1\sum \frac{1}{x} = \sum y,$$

$$a_0\sum\frac{1}{x}+a_1\sum\frac{1}{x^2}=\sum\frac{y}{x}$$

Интерпретация моделей регрессии осуществляется методами той отрасли знаний, к которой относятся исследуемые явления. Но всякая интерпретация начинается со статистической оценки уравнения регрессии в целом и оценки значимости входящих в модель факторных признаков, т.е. выяснения, как они влияют на величину результативного признака. Чем больше величина коэффициента регрессии, тем значительнее влияние данного признака на моделируемый. Особое значение при этом имеет знак перед коэффициентом регрессии. 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.

Множественная (многофакторная) регрессия – это изучение связи между тремя и более связанными между собой признаками, она описывается функцией вида

$$\bar{y}_{1,2,\ldots,k}=f\left(x_1,x_2,\ldots,x_k\right).$$

Построение моделей множественной регрессии включает несколько этапов:

– выбор формы связи (уравнения регрессии);
– отбор факторных признаков;
– обеспечение достаточного объема совокупности для получения несмещенных оценок.

Зависимости между социально-экономическими явлениями описываются следующими типами моделей:

линейной: $\bar{y}_{1,2,\ldots,k}=a_0+a_1x_1+a_2x_2+\ldots+a_kx_k$;

степенной: $\bar{y}_{1,2,\ldots,k}=a_0x_1^{a_1}\times x_2^{a_2},\ldots,x_k^{a_k}$

показательной: $\bar{y}_{1,2,\ldots,k}=e^{a_0+a_1x\ +a_2x_2+\ldots+a_kx_k}$

параболической: $\bar{y}_{1,2,\ldots,k}=a_0+a_1x_1^2+a_2x_2^2+\ldots+a_kx_k^2$

гиперболической: $\bar{y}_{1,2,...k} = a_0 + \frac{a_1}{x_1} + \frac{a_2}{x_2} + ... + \frac{a_k}{x_k}$

Наиболее приемлемым способом отбора факторных признаков является метод шаговой регрессии. Сущность метода заключается в последовательном включении факторов в уравнение регрессии и последующей проверке их значимости.

Система нормальных уравнений для нахождения параметров линейной множественной регрессии методом наименьших квадратов имеет следующий вид:

$$\begin{cases} na_0 + a_1\sum x_1 + a_2\sum x_2 + ... + a_i\sum x_i + ... + a_k\sum x_k = \sum Y, \\ a_0\sum x_1 + a_1\sum x_1^2 + a_2\sum x_2 x_1 + ... + a_i\sum x_i x_1 + ... + a_k\sum x_k x_1 = \sum Y x_1 \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ a_0\sum x_k + a_1\sum x_1 x_k + a_2\sum x_2 x_k + ... + a_i\sum x_i x_k + ... + a_k\sum x_k^2 = \sum Y x_k \end{cases}$$

Фактор является незначимым, если его включение в уравнение регрессии только изменяет значение 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 коэффициент регрессии не изменяется (или меняется несущественно), то данный признак существенен, и его включение в уравнение регрессии необходимо.

Если же при включении в модель факторного признака коэффициенты регрессии меняют не только величину, но и знаки, а множественный коэффициент корреляции не возрастает, то данный факторный признак признается нецелесообразным для включения в модель связи.

Сложность и взаимное переплетение отдельных факторов, обуславливающих исследуемое экономическое явление, могут проявляться в *мультиколлинеарности* – тесной связи между факторными признаками, включенными в модель. Одним из индикаторов определения наличия мультиколлинеарности между признаками является превышение парным коэффициентом корреляции величины 0,8 ($r_{x_i x_j}$) и др.

Устранение мультиколлениарности осуществляется путем исключения из корреляционной модели одного или нескольких линейно-связанных факторных признаков или преобразование исходных факторных признаков в новые, укрупненные факторы.
Проверка адекватности моделей, построенных на основе уравнений регрессии, начинается с проверки значимости каждого коэффициента регрессии.
Значимость коэффициентов регрессии осуществляется с помощью t-критерия Стьюдента:

$$t_{\text{р}} = \frac{|a_i|}{\sqrt{\sigma^2_{a_i}}},$$

где $\sigma^2_{a_i}$ – дисперсия коэффициента регрессии, определяемая по формуле

$$\sigma^2_{a_i} = \frac{\sigma^2_Y}{k},$$

где σ^2_Y – дисперсия результативного признака;
k – число факторных признаков в уравнении.

Параметр модели признается статистически значимым, если

$$t_p > t_{kp}\ (\alpha;\ \nu = n - k - 1),$$

где α – уровень значимости;
$\nu = n - k - 1$ – число степеней свободы.
Табличное значение (t_{kp}) находится по приложению 6.

Проверка *адекватности всей модели* осуществляется с помощью расчета F-критерия Фишера:

$$F = \frac{\delta^2_{yx}}{\sigma^2_{ост}} \div \frac{k}{n-k-1} \text{ или } F = \frac{r^2_{yx}}{1-r^2_{yx}} \div \frac{k}{n-k-1}$$

входные параметры $\alpha, \nu_1 = k+1, \nu_2 = n-k-1,$

где k – число факторных признаков в уравнении.
n – количество единиц совокупности;
r^2_{yx} – коэффициенты детерминации.

Если $F_{факт} > F_{табл}$, то можно говорить о надежности построенного уравнения. $F_{табл}$ определяется по приложению 7.

Корреляционный анализ имеет задачей количественное определение тесноты связи между двумя признаками (при парной связи) и между результативным и множеством факторных признаков (при многофакторной связи).

В статистике различают следующие варианты зависимостей:

– парная корреляция – связь между двумя признаками (результативным и факторным или двумя факторными);

– частная корреляция – зависимость между результативным и одним факторным признаками при фиксированном значении других факторных признаков;

– множественная корреляция – зависимость результативного и двух или более факторных признаков, включенных в исследование.

На практике различают следующие показатели корреляции:

– линейный коэффициент корреляции;
– эмпирическое корреляционное отношение;
– индекс корреляции (теоретическое корреляционное отношение);
– частные коэффициенты корреляции;
– коэффициент множественной корреляции;
– коэффициент детерминации.

Линейный коэффициент корреляции используется для изучения связи между двумя признаками при наличии между ними линейной зависимости и определяется по формуле

$$r_{yx} = \frac{\overline{yx} - \bar{y} \times \bar{x}}{\sigma_x \sigma_y},$$

где $\bar{x}, \bar{y}, \overline{yx}$ – средние значения факторного, результативного признаков и произведения факторного и результативного признаков;

σ_x, σ_y − средние квадратические отклонения факторного и результативного признаков.

Линейный коэффициент корреляции изменяется в пределах от (−1) до (+1): $-1 \le r \le 1$. Интерпретация выходных значений представлена в таблице 90.

Таблица 90 − Оценка линейного коэффициента корреляции

Значение коэффициента связи	Характер связи	Интерпретация связи
$r = 0$	Отсутствует	-
$0 < r < 1$	Прямая, однонаправленная	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

Производя расчет по итоговым значениям исходных переменных, линейный коэффициент корреляции можно вычислить по формуле

$$r_{xy} = \frac{n \cdot \sum xy - x \cdot \sum y}{\sqrt{\left[n\sum x^2 - \left(\sum x\right)^2\right] \cdot \left[n\sum y^2 - \left(\sum y\right)^2\right]}}.$$

В том случае, когда исходная информация представлена в виде корреляционной таблицы, необходимо учесть частоты повторений индивидуальных значений факторного и результативного признаков, а также число повторений данного сочетания их значений. В этом случае формула коэффициента корреляции будет иметь следующий вид:

$$r_{xy} = \frac{n \cdot \sum xyf_{xy} - \sum xf_x \cdot \sum yf_y}{\sqrt{\left[n\sum x^2 f_x - \left(\sum xf_x\right)^2\right] \cdot \left[n\sum y^2 f_y - \left(\sum yf_y\right)^2\right]}}.$$

Линейный коэффициент корреляции может быть также выражен через дисперсии слагаемых:

$$r_{xy} = \frac{\sigma_x^2 + \sigma_y^2 - \sigma_{x-y}^2}{2\sigma_x\sigma_y}.$$

Эмпирическое корреляционное отношение (η) рассчитывается по данным группировки в случае нелинейной зависимости между признаками и определяется по формуле

$$\eta = \sqrt{\frac{\delta^2}{\sigma^2}},$$

где δ^2 – межгрупповая дисперсия, характеризующая вариацию результативного признака, обусловленную группировочным признаком;
σ^2 – общая дисперсия результативного признака.

Показатель изменяется в пределах от 0 до 1. Интерпретация значений коэффициента представлена в таблице 91.

Таблица 91 − Оценка эмпирического корреляционного отклонения

Значение коэффициента	Интерпретация связи
0,1–0,3	Слабая
0,3–0,5	Умеренная
0,5–0,7	Заметная
0,7–0,9	Тесная
0,9–0,99	Весьма тесная

Индекс корреляции (теоретическое корреляционное отношение) (R) используют для измерения связи при любой ее форме и определяется по формуле

$$R = \sqrt{\frac{\sigma_y^2 - \sigma_{y-y_x}^2}{\sigma_y^2}},$$

где $\sigma_{y-y_x}^2 = \dfrac{\sum (y - y_x)^2}{n}$ – дисперсия отклонений;

σ_y^2 – дисперсия фактических значений результативного признака;

$\sigma_{y_x}^2 = \sigma_y^2 - \sigma_{y-y_x}^2$ – дисперсия теоретических значений результативного признака.

Индекс корреляции изменяется в пределах от 0 до 1. Если он равен нулю, то связи между вариацией признаков *x* и *y* нет. Чем он ближе к 1, тем связь между признаками теснее.

Частные коэффициенты корреляции (r) применяются для характеристики тесноты связи между двумя признаками при фиксированном значении других признаков и определяется по формуле

$$r_{1,2,3,\ldots,n} = \frac{r_{1,2,3,\ldots,n-1} - r_{1,n,3,\ldots,n-1} \times r_{2,n,3,\ldots,n-1}}{\sqrt{\left(1 - r_{1,n,3,\ldots,n-1}^2\right) \times \left(1 - r_{2,n,3,\ldots,n-1}^2\right)}}.$$

Коэффициент множественной корреляции (R) применяется в случае оценки связи между результативным (*Y*) и двумя факторными (X_1, X_2) признаками. Множественный коэффициент корреляции имеет вид

$$R = \sqrt{\frac{r_{yx_1}^2 + r_{yx_2}^2 - 2 \times r_{yx_1} \times r_{yx_2} \times r_{x_1x_2}}{1 - r_{x_1x_2}^2}},$$

где *r* – парные коэффициенты корреляции между признаками.

Значения коэффициента находятся в пределах от 0 до 1. Чем ближе значение коэффициента к единице, тем теснее связь между признаками. Множественный коэффициент корреляции можно рассчитать, используя парные коэффициенты корреляции *(r_{yx})* и β-коэффициенты:

$$R_{x_1,x_2,\dots,x_n\dots} = \sqrt{\beta_1 \times r_{yx_1} + \beta_2 \times r_{yx_2} + \dots + \beta_n \times r_{yx_n}} \;.$$

Коэффициент детерминации (d) показывает, какая доля вариации изучаемого результативного признака объясняется влиянием факторов, включенных в уравнение множественной регрессии, и представляет собой квадрат коэффициента корреляции:

$$d = r^2 \times 100\,\% \,,$$

$$d = R^2 \times 100\,\%$$

Коэффициент изменяется в пределах от 0 до 100 и показывает, на сколько процентов изменение результативного признака зависит от выбранных в модель факторных признаков. Остальные проценты (до 100) отражают влияние других, не учтенных в модели признаков.

Значимость линейного коэффициента корреляции проверяется на основе t-критерия Стьюдента (приложение 6).

Если объем совокупности (n) <50 единиц, то формула критерия имеет вид

$$t_{\text{расч}} = \sqrt{\frac{r^2}{1-r^2} \times (n-2)} = \frac{|r|}{\sqrt{1-r^2}} \times \sqrt{n-2}$$

входные параметры $\alpha,\, k = n-2$.

Если расчетное значение $t_{\text{расч}} > t_{\text{табл}}$, то коэффициент корреляции принято считать значимым.

Если объем совокупности (n) более 100 единиц, то используется формула

$$t_{\text{расч}} = \frac{|r|}{\sqrt{1-r^2}} \times \sqrt{n}\ .$$

Проверка значимости коэффициента множественной корреляции осуществляется на основе *F-критерия Фишера*:

$$F_p = \frac{\frac{1}{2} R^2_{yx_1x_2}}{\frac{1}{n-3} \cdot \left(1 - R^2_{yx_1x_2}\right)}$$

входные параметры $\alpha, \nu_1 = 2, \nu_2 = n-3$.

Если $F_{\text{расч}} > F_{\text{табл}}$, то коэффициент множественной корреляции считается значимым. $F_{\text{табл}}$ определяется по приложению 7.

Для выявления влияния каждого отдельного фактора на результативный признак вычисляют *стандартизированные коэффициенты:*

– коэффициенты эластичности (α-коэффициент);
– β-коэффициенты.

Коэффициенты эластичности (Э) (α-коэффициент), определяется по формуле

$$Э_{x_i} = a_i \times \frac{\bar{x}_i}{\bar{y}},$$

где $\bar{x}_i$ – среднее значение соответствующего факторного признака;

$\bar{y}$ – среднее значение результативного признака;

a_i – коэффициент регрессии при соответствующем факторном признаке.

Коэффициент эластичности показывает, на сколько процентов в среднем изменяется значение результативного признака при изменении факторного признака на 1 %.

β-коэффициент определяется по формуле

$$\beta_i = a_i \times \frac{\sigma_{x_i}}{\sigma_y}$$

где σ_{x_i} – среднее квадратическое отклонение *i*-го фактора;

σ_y − среднее квадратическое отклонение результативного признака.

β-коэффициент показывает, на какую часть среднего квадратического отклонения изменится результативный признак при изменении соответствующего факторного признака на свое среднее квадратическое отклонение.

Пример. По данным о сумме активов и заемных средств по семи сельскохозяйственным организациям Пензенской области определим:

1) зависимость между признаками с использованием уравнения парной линейной регрессии;
2) линейный коэффициент корреляции и детерминации;
3) коэффициент эластичности и *β*-коэффициент.

Исходные данные и расчетные показатели представим в таблице 92.

Таблица 92 − Уровень активов и заемных средств сельскохозяйственных организаций в Пензенской области на 01.01.2010 г., (млн. руб.)

№ предприятия	Заемные средства, *(x)*	Сумма активов, *(y)*	x^2	xy	y^2	$\bar{y}_x$
1	12,8	81,1	163,84	1038,08	6577,21	64,63
2	43,2	85,0	1866,24	3672	7225,00	81,17
3	7,3	48,3	53,29	352,59	2332,89	61,64
4	61,9	76,6	3831,61	4766,3	5867,56	91,34
5	13,8	58,1	190,44	801,78	3375,61	65,18
6	25,3	64,3	640,09	1626,79	4134,49	71,43
7	40,3	101,6	1624,09	4094,48	10322,56	79,59
Итого	204,6	515	8369,6	16352,02	39835,32	515

1. Определим параметры уравнения с помощью системы нормальных уравнений:

$$\begin{cases} 7a_0 + a_1 204{,}6 = 515 \\ a_0 204{,}6 + a_1 8369{,}6 = 16352{,}02 \end{cases}$$

$$a_0 = 57{,}67;\ a_1 = 0{,}544$$

Подставим полученные значения в уравнение линейной регрессии $\bar{y}_x = a_0 + a_1 x$:

$$\bar{y}_x = 57{,}67 + 0{,}544x$$

С увеличением заемных средств на 1 млн. руб. сумма активов возрастет в среднем на 0,544 млн руб.

2. Для расчета линейного коэффициента корреляции исчислим необходимые составляющие:

$$\bar{x} = \frac{204{,}6}{7} = 29{,}23$$

$$\bar{y} = \frac{515}{7} = 73{,}57$$

$$\overline{xy} = \frac{16352{,}02}{7} = 2336{,}00$$

$$\sigma_x = \sqrt{\frac{8369{,}6}{7} - 29{,}23^2} = 18{,}47$$

$$\sigma_y = \sqrt{\frac{39835{,}32}{7} - 73{,}57^2} = 16{,}68$$

Следовательно,

$$r_{XY} = \frac{\overline{yx} - \bar{y} \times \bar{x}}{\sigma_x \sigma_y} = \frac{2336{,}00 - 73{,}57 \times 29{,}23}{18{,}47 \times 16{,}68} = 0{,}602$$

Связь между признаками заметная и однонаправленная.

Определим коэффициент детерминации:

$$d = r^2 \times 100\% = 0{,}602^2 \times 100\% = 36{,}24\,\%$$

Вариация размера активов на 36,24 % зависит от величины заемных средств сельскохозяйственных организаций, на 63,76 % − от других факторов.

2. Определим коэффициент эластичности:

$$Э_{x_i} = a_i \times \frac{\bar{x}_i}{\bar{y}} = 0{,}544 \times \frac{29{,}23}{73{,}57} = 0{,}216.$$

Если размер заемных средств увеличить на 1 %, то размер активов вырастет на 0,216 %.
Найдем значение β-коэффициента:

$$\beta_i = a_i \times \frac{\sigma_{X_i}}{\sigma_Y} = 0{,}544 \times \frac{18{,}47}{16{,}68} = 0{,}602.$$

Если размер заемных средств увеличится на свое среднее квадратическое отклонение, то активы организаций вырастут на 0,602 своего среднего квадратического отклонения.