ФЕДЕРАЛЬНОЕ АГЕНСТВО ВОЗДУШНОГО ТРАНСПОРТА

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ПРОФЕССИОНАЛЬНОГО ОБРАЗОВАНИЯ

МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ ГРАЖДАНСКОЙ АВИАЦИИ

Кафедра Экономики ГА

СТАТИСТИКА

Конспект лекций
(часть I)

Москва 2006 г.

Оглавление

Введение……………………………………………………………

Глава 1. МАССОВОЕ СТАТИСТИЧЕСКОЕ НАБЛЮДЕНИЕ……………

1.1 Формы организации статистического наблюдения……………….
1.2 План статистического наблюдения………………………………..
1.3 Виды статистического наблюдения………………………………..

Глава 2. СВОДКА И ГРУППИРОВКА МАТЕРИАЛОВ СТАТИСТИЧЕСКОГО НАБЛЮДЕНИЯ…………………………….

2.1 Виды статистических группировок………………………………
2.2 Выбор группировочных признаков и основные правила образования групп………………………………………………………

Глава 3. АБСОЛЮТНЫЕ И ОТНОСИТЕЛЬНЫЕ ВЕЛИЧИНЫ……………

3.1 Абсолютные величины…………………………………………….
3.2 Относительные величины…………………………………………..

Глава 4. СРЕДНИЕ ВЕЛИЧИНЫ…………………………………………

4.1 Средняя арифметическая и средняя гармоническая……………….
4.2 Мода и медина…………………………………………………………..

Глава 5. ПОКАЗАТЕЛИ ВАРИАЦИИ………………………………………..

5.1 Абсолютные и относительные показатели вариации………………
5.2 Правило сложения дисперсий……………………………………….
5.3 Коэффициент ассиметрии……………………………………………
5.4 Метод упрощенного вычисления средней арифметической, дисперсии и среднеквадратического отклонения………………..

Глава 6. ИНДЕКСЫ…………………………………………………………

6.1 Квалификация индексов……………………………………………
6.2 Индексы количественных показателей и их преобразование…..
6.3 Индексы качественных показателей и их преобразование……….
6.4 Взаимосвязь индексов………………………………………………

Глава 7. СТАТИСТИСЕСКОЕ ИЗУЧЕНИЕ СВЯЗИ ЯВЛЕНИЙ…………

7.1 Определение наличия корреляционной связи……………………
7.2 Определение параметров уравнения связи………………………
7.3 Показатели степени тесноты связи между признаками…………..
7.4 Понятие множественной корреляции………………………………

Глава 8. ПОКАЗАТЕЛИ ВЫБОРОЧНОГО НАБЛЮДЕНИЯ………………

8.1 Доля и средняя……………………………………………………..
8.2 Ошибки выборки…………………………………………………..
8.3 Способы формирования выборочной совокупности………………

Глава 9. РЯДЫ ДИНАМИКИ…………………………………………………

9.1 Классификация рядов динамики……………………………………
9.2 Показатели динамических рядов…………………………………….
9.3 Сглаживание временных рядов……………………………………
9.4 Изучение сезонной неравномерности………………………………
9.5 Приведение рядов динамики к общему основанию……………….

Введение

Слово «статистика» произошло от латинского слова «статус», что означает определенное состояние, положение вещей. В настоящее время этот термин употребляется в нескольких значениях.

Как правило, говоря о статистике, имеют в виду совокупность цифровых данных о той или иной сфере общественной жизни (статистика населения, статистика здравоохранения и т. д.) Под статистикой понимают особую отрасль общественной деятельности людей, направленную на сбор, обработку и анализ данных, характеризующих состояние экономики, культуры и т. д. И наконец статистикой называют общественную науку, занимающейся разработкой методов сбора, сводки, обработки и анализаи теоретическим обобщением **цифровых** данных о явлениях общественной жизни.

Между статистикой наукой и статистикой практикой существует тесная связь. Статистическая практика использует теоретические положения и методы, выработанные наукой. В свою очередь, статистическая наука опирается на материалы практики, обобщая практический опыт, разрабатывает новые положения.

Особенность статистики заключается прежде всего в том, что статистические данные сообщаются в количественной форме: статистика говорит языком цифр, дает цифровое освещение изучаемых ею общественных явлений. Статистику интересуют прежде всего те выводы, которые можно сделать на основе анализа собранных и обработанных цифровых данных.

Предметом статистики являются все общественные явления и процессы, происходящие в стране. Статистика показывает объем и изменение общественных явлений, взаимосвязи между этими явлениями.

Статистическая методология представляет собой совокупность приемов, правил и методов статистического исследования социально-экономических явлений: сбора информации, ее обработки, вычисления обобщающих показателей, анализ данных. Основными являются метод массового наблюдения, метод группировок и метод обобщающих показателей.

Массовое статистическое наблюдение представляет собой **первый этап** любого статистического исследования. Он характеризуется сбором первичного материала по заранее разработанной программе и необходим для того, чтобы статистика могла выявлять закономерности изучаемых явлений, характерные черты и тенденции их развития, вытекающие именно из сущности изучаемого явления и тех общих условий, в которых оно формируется.

Второй этап – группировка и сводка статистических данных – является одним из важнейших этапов статистического исследования и предопределяет правильность конечных результатов. Группировки дают возможность выделить из состава массы и раздельно изучить явления разного качества, показать особенности явлений, развивающихся в различных условиях. При помощи группировок изучают взаимосвязи общественных явлений, их зависимость от различных факторов.

После проведения группировки приступают к **третьему этапу –** к обработке данных наблюдения, т. е. к получению обобщающих показателей. Различают три основных вида статистических показателей:

1.Абсолютные величины, при помощи которых измеряют объемы (размеры) общественных явлений;

2.Относительные величины, которые выражают соотношение между частями или различными группами явления;

3. Показатели, отражающие характерные особенности отдельных однородных групп (средние величины, показатели вариации и т. д.).

На **четвертом этапе** статистического исследования собранные и обработанные данные подвергаются анализу и дальнейшему обобщению. Цель анализа – получение выводов о состоянии изучаемого общественного явления и закономерностях его развития. Эти выводы излагаются, как правило, в текстовой форме.

Важнейшими особенностями статистической методологии являются конкретность исследования; выяснение, главным образом сущности явления с учетом места и времени его развития; выделение однородных совокупностей, прежде всего социальных и экономических типов изучаемого явления; видоизменение приемов и методов исследования в связи с изменением сущности и форм изучаемых явлений и процессов общественной жизни; применение системы показателей, позволяющих дать всестороннюю характеристику изучаемых явлений и процессов, закономерностей их размеров и количественных соотношений.

Глава 1. **МАССОВОЕ СТАТИСТИЧЕСКОЕ НАБЛЮДЕНИЕ**

Статистическое наблюдение – это планомерный, научно организованный сбор данных о явлениях и процессах общественной жизни путем регистрации по заранее разработанной программе наблюдения существенных их признаков. Данные наблюдения представляют собой первичную статистическую информацию о наблюдаемых объектах, которая является основой для получения их обобщающих характеристик. Статистическое наблюдение выступает как один из главных методов статистики и как одна из важнейших стадий (этапов) статистического исследования.

1.1. **Формы сбора данных статистического наблюдения**

По способу организации сбора данных статистического наблюдения различают две его формы – отчетность и специально организованное статистическое наблюдение.

Отчетность – это такая форма организации статистического наблюдения, при которой сведения поступают в статистические органы от предприятий, организаций и учреждений в виде обязательных отчетов об их деятельности. Для нее характерны:

а) обязательность представления отчетов по заранее установленной программе (показателям) в установленные адреса и сроки;

б) обоснованность данных документами, которыми оформляются хозяйственные операции в оперативном и бухгалтерском учете;

в) юридическая сила отчетов, за достоверность показателей которых ответственность несут руководители предприятий.

Отчетность делится на бухгалтерскую и статистическую. Бухгалтерская отчетность основана на данных бухгалтерского учета. Статистическая отчетность в основном базируется на данных оперативного учета, ее формы разрабатываются и утверждаются Федеральной службой по статистике (Росстат).

По длительности отображаемого периода и частоте представления различают текущую отчетность, периодичность представление которой менее года (полгода, квартал, месяц, декада, неделя) и годовую, представляемую за год.

Отчеты представляются на унифицированных формах (бланках), содержащих следующие обязательные реквизиты:

1. номер формы и дату ее утверждения;
2. название формы;
3. адреса, в которые должен представляться отчет;
4. за какой период (на какую дату) и какого числа после отчетного периода должен представляться отчет;

5. наименование отчитывающегося предприятия, вышестоящих органов, которым оно подчинено, и администротивно-территориальной единицы, где оно расположено;
6. сообщаемые показатели;
7. подписи лиц, ответственных за достоверность и своевременность представления отчета.

Состав, порядок и способы представления отчетов для предприятий устанавливаются в табеле отчетности, представляющем собой полный перечень отчетности, по каждой из которых указываются ее номер, наименование, способ представления, периодичность, кем и кому представляется, когда высылается.

Отчетность является основной формой статистического наблюдения. Основываясь на документах первичного и бухгалтерского учета, обязательном и регулярном представлении данных, она позволяет удовлетворить потребности органов управления в статистических данных о важнейших сторонах деятельности, состояния и развития предприятия.

Часть данных учета требуется нерегулярно, а сведения о некоторых процессах воспроизводства населения, личного потребления, доходах и расходах семей и т.д. не регистрируются первичным и бухгалтерским учетом и не могут быть получены от предприятий. Поэтому для их сбора проводятся специально организованные статистические наблюдения.

Это, прежде всего, переписи, проводимые по единой программе единовременно на какой – либо территории с целью получения сведений, отсутствующих в отчетности (перепись населения, перепись не установленного оборудования перепись жилищного фонда и т. д.)

Другой формой специально организованного статистического наблюдения являются разного рода обследования. К ним, в частности, относятся обследования семейных бюджетов, цен и продаж на колхозных рынках, разного рода социологические обследования.

1.2 План статистического наблюдения.

Статистическое наблюдение необходимо проводить по строго определенному плану, который включает в себя:

1) определение цели и задач наблюдения;
2) установление объекта и единицы наблюдения;
3) определение вопросов программы статистического наблюдения;
4) заполнение статистических документов – инструкций и формуляров;
5) решение организационных вопросов.

При этом к статистическому наблюдению необходимо предъявлять следующие требования: данные должны быть объективными, своевременными и актуальными.

Рассмотрим каждый раздел плана наблюдения.

На первом этапе необходимо точно сформулировать цель и задачи наблюдения, т. к. от этого зависят выбор объекта, единицы наблюдения и программа.

Если, например, исследуют производительность труда, то следует собирать сведения о произведенной продукции и затратах труда, об установленных нормах выработки и их выполнении, об уровне механизации труда, составе рабочих и других показателях, влияющих на уровень производительности труда. Если же речь идет о себестоимости промышленной продукции, то в центре исследования будут производственные затраты и объем произведенной продукции, организация и оплата труда, нормы расходования материалов и другие показатели, влияющие на себестоимость продукции.

Объект наблюдения. Объектом статистического наблюдения называется совокупность единиц изучаемого явления, о которых должны быть собраны статистические сведения.

Для определения объекта обычно дают четыре его границы:

1) территориальная (граница места);
2) граница времени;
3) граница качественная;
4) граница количественная.

Территориальные границы объекта обычно совпадают с границами определенных административных делений. Различают наблюдения, охватывающее данное явление на всей территории (перепись населения России, оборудования и т. п.), и наблюдения, осуществляемые в пределах данной территориальной единицы – района, города и т. д. (например, перепись жилищного фонда города Москвы). Особую группу составляют наблюдения, которые ограничиваются рамками предприятий или отдельных участков производства.

Граница времени устанавливается в плане наблюдения различно в зависимости от особенностей объекта и содержания наблюдения. Различают объективное и субъективное время статистического наблюдения. Объективным временем (критический момент) называется время, к которому относятся данные наблюдения. Оно характеризует тот период или момент времени, по состоянию на который быили собраны, зарегистрированы признаки совокупности. Так, например. Данные о выпуске продукции, размерах потребления и т. д. можно получить только за определенный период времени. Сведения о численности населения,, врачей можно собрать только на определенную дату.

В некоторых случаях правильная постановка наблюдения требует определения не только даты, но и момента времени по состоянию на который производится регистрация явления. Этот момент называется критическим.

Надобность в нем возникает тогда, когда изучаемые явления изменяются не только по дням, но и по часам и минутам (численность населения).

Субъективное время наблюдения – это время производства наблюдения, т. е. период, в течение которого производится регистрация единиц совокупности (например, срок представления отчета предприятия за 1 квартал – 10 апреля; в этом случае объективное время составит 3 месяца, а субъективное – 10 дней, т. е. время, которое дается для составления отчета).

Наиболее трудным является определение качественных границ объекта наблюдения. Здесь требуется указать, какие именно части или категории изучаемого явления подлежат регистрации, и, кроме того, точно определить основания, которыми следует руководствоваться при отнесении отдельных случаев в категорию учитываемых.

Например, при переписи оборудования возникает вопрос, учитывать ли оборудование всех отраслей промышленности или оборудование определенной отрасли. При проведении хронометража указывают категорию подлежащих наблюдению рабочих и круг подлежащих хронометрированию операций и т. д.

Граница количественная означает, что только при определенных размерах учитывается объект наблюдения.

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

Наряду с определением объекта статистического наблюдения необходимо определить и единицу этой совокупности, а также установить единицу наблюдения.

Единица совокупности – это первичный элемент объекта статистического наблюдения, являющийся носителем признаков, подлежащих регистрации. От единицы совокупности следует отличать единицу наблюдения. **Единица наблюдения** – это та превичн6ая ячейка, от которой должны быть получены необходимые статистические сведения в процессе наблюдения.

Иначе говоря, единица совокупности – это то, что подвергается обследованию, а единица наблюдения – это источник получаемых сведений. Определение единицы совокупности важно при разработке программы статистического наблюдения, а определение единицы наблюдения – при решении вопросов организации сбора сведений.

Единицами наблюдения являются главным образом отдельные предприятия, учреждения и организации. В некоторых случаях единицами наблюдения выступают отдельная семья рабочего, служащего (при изучении семейного бюджета), отдельный человек (в переписи населения). Единица совокупности в зависимости от задач и характера работы могут быть самыми разнообразными. Они иногда совпадают с единицами наблюдения (перепись населения), но чаще всего не совпадают. Так, например при проведении переписи промышленного оборудования единицей наблюдения будет промышленное предприятие, от него получают сведения о промышленном оборудовании. Единицей совокупности будет отдельная единица оборудования (станок, например), т. к. признаки, регистрируемые в этой переписи (техническое состояние, его производственные мощности и др.) относятся не к предприятию в целом, а к каждой единице оборудования.

После того, как определены объект и единица наблюдения, ножно составить программу наблюдения. **Программа статистического наблюдения** – это перечень четко сформулированных вопросов, на которые необходимо получить ответы приданном статистическом наблюдении.

В программу должны включатся только те вопросы, на которые могут быть даны достоверные ответы. Поэтому в статистическую отчетность нужно включать только такие показатели, которые основываются на данных первичного учета предприятий или же могут быть установлены путем непосредственной регистрации фактов.

Для правильной организации статистического наблюдения и получения достоверных материалов большое значение имеет формулировка вопросов программы наблюдения. Они должны быть сформулированы таким образом, чтобы их содержание понимали все одинаково.

Степень полноты программы наблюдения зависит от числа охватываемых наблюдением признаков изучаемого явления. С одной стороны, и от выбора этих признаков – с другой.

Чем шире программа, чем полнее и разносторонее освящается изучаемое явление, тем более пригодны данные наблюдения для научного познания. Расширение программы наблюдения имеет, однако, свои границы, т. к. с расширением программы связано значительное увеличение работы по сбору сведений и их обработке, а это замедляет получение выводов и «удорожает» наблюдения. Содержание программы должно быть поэтому в каждом случае строго продумано, нужно избегать внесение в программу «лишних», выходящих за указанные рамки вопросов.

Выбор признаков, которые достаточно отражали бы стороны изучаемого явления, часто представляет значительные трудности.

В некоторых случаях (например, при характеристике возрастного состава населения) наглядно видны те признаки, которые непосредственно отражают интересующие нас стороны изучаемого явления. Во многих же других случаях об изучаемых особенностях явления приходится судить на основе косвенных признаков. Например, уровень квалификации рабочего не может быть прямо измерен, а выражается косвенным признаком – присвоенным рабочему разрядом.

Ответы на вопросы программы наблюдения собирают путем заполнения документов, называемыми статистическими формулярами наблюдения, к которым дается инструкция по их заполнению.

Формуляры наблюдения – это заранее заготовленные бланки (обычно отпечатанные), в которых в определенном порядке записаны как вопосы программы, так и заносятся ответы на эти вопросы.

Формуляры статистического наблюдения могут носить различные названия (форма отчетности, переписной лист, бланк обследования и т. д.) и иметь различную форму.

Обычно различают два вида статистических формуляров: индивидуальные формуляры (карточки) и списочные.

В индивидуальный формуляр заносятся данные по одной единице наблюдения, например, формы отчетности являются индивидуальными формулярами (в каждом отдельном отчете сообщаются сведения только по одному предприятию).

В списочный формуляр заносятся данные наблюдения нескольких единиц.

При разработке формуляров необходимо обращать внимание на четкую и ясную постановку вопросов; неясные формулировки могут повлечь за собой неправильные и нечеткие ответы на поставленный вопрос.

Формуляр заполняется сведениями в процессе наблюдения. Иногда, однако, он может содержать также разделы, в которые до производства наблюдения заносятся некоторые взятые из других наблюдений данные, характеризующие общие условия, при котрорых наблюдается единица (или ряд единиц) явления.

Так, при проведении хронометражных наблюдений, в формуляр наблюдения (хронокарту) заносятся сначала сведения о рабочем (разряд, средняя выработка), взятые из имеющихся документов, и об условиях выполнения данной операции; в процессе же наблюдения хронокарта заполняется сведениями о длительности выполнения отдельных элементов операции.

В инструкциях к проведению наблюдения и заполнению формуляров разъясняются основные вопросы программы, указываются, в какой форме и на основе каких данных даются ответы на эти вопросы и т. д. Инструкциями снабжаются все участники наблюдения.

В организационном плане наблюдения должны быть указаны:

1) органы, которым поручено наблюдение;
2) форма организации наблюдения;
3) время наблюдения (критический момент и срок наблюдения);
4) вид наблюдения;
5) порядок проверки полученных данных;
6) смета расходов, связанных с выполнением данной работы.

1.2. Виды статистического наблюдения

Статистическое наблюдение делится на несколько видов по полноте охвата единиц совокупности и по учету факторов во времени. По этим признакам может быть составлена следующая классификация статистического наблюдения (таблица 1)

Таблица 1

Классификация статистического наблюдения

Признак классификации	Виды наблюдения	Разновидности
1	2	3
Полнота охвата единиц совокупности	1 Сплошное 2. Несплошное	2.1. Выборочное 2.2. Основного массива 2.3. Анкетное 2.4.Монографическое
Учет фактов во времени	1.Текущее (непрерывное) 2. Периодическое 3. Единовременное	- - -

Сплошным наблюдением называется такое статистическое наблюдение, при котором регистрации подлежат все без исключения единицы, входящие в состав изучаемого объекта..

Сплошное наблюдение применяется преимущественно там, где необходимо точно определить размеры изучаемых общественных явлений или собрать данные, необходимые для контроля выполнения плана. Сплошными являются единовременные наблюдения, если они предпринимаются с целью получения данных о количественном объеме изучаемого явления; в форме сплошного наблюдения проводится также и текущее наблюдение, на основе которого устанавливается число тех или иных событий за определенный отрезок времени. Например, отчетность о выполнении плана представляется каждым предприятием, т. е. является сплошным наблюдением.

Применение сплошного наблюдения, однако, не всегда возможно и не всегда целесообразно; его приходится часто заменять несплошным наблюдением.

Применение несплошных наблюдений возможно потому, что сводные статистические характеристики явлений становятся точными при достаточно большом числе наблюдений, даже если последними и не охвачены все единицы совокупности. В этом проявляется закон больших чисел.

Несплошные наблюдения имеют преимущества перед сплошными: они значительно меньше по объему, требуют намного меньше сил и средств и позволяют применять более совершенный способ учета фактов, быстрее подвести итоги обследования и, следовательно, првысить оперативность статистического материала.

В некоторых же случаях несполошное наблюдение является единственно возможным (при контроле качества продукции, изучения потребления населения и т. д.). Поэтому несплошные обследования в статистике существенно дополняют основные материалы, которые получают в результате сплошных наблюдений

Несплошное наблюдение организуется по- разному в зависимости от характера объекта и задачи исследования. Если на основе частичного исследования необходимо получить данные, характеризующую всю совокупность в целом, то несплошное наблюдение должно быть организовано как выборочное.

Выборочным наблюдением называется такое, при котором характеристика всей совокупности фактов дается по некоторой части, отобранной в случайном порядке. Выборочное наблюдение – наиболее распространенный вид несплошного наблюдения в статистике.

В основе выборочного наблюдения лежит случайный отбор единиц, подлежащих обследованию. Этим гарантируется независимость результатов выборки от воли лиц, производящих выборку.

Выборочное наблюдение широко применяют в различных отраслях экономической статистики. Так, в промышленности его применяют для контроля качества продукции, изучения использования рабочего времени, использования оборудования.

Вторым видом несплошного наблюдения является способ основного массива. При этой форме наблюдения отбирают наиболее крупные единицы наблюдения, в которых сосредоточена значительная часть всех подлежащих изучению фаектов.

Несполшное наблюдение может быть еще организовано и как анкетное. Суть анкетного наблюдения заключается в том, что лицам, от которых необходимо получить сведения, рассылают анкеты с просьбой заполнить их и прислать обратно. Обычно обратно получают значительно меньше анкет, чем рассылают, и наблюдения, таким образом, всегда получается несплошным, даже при условии, если анкеты были разосланы всем единицам наблюдения. Если произведенный таким образом отбор окажется случайным, то по результатам анкетное обследование будет соответствовать выборочному.

К анкетному способу прибегают редко и лишь в тех случаях, когда нет другого, более надежного способа получения данных. К этому способу, например, прибегают в статистике связи, чтобы выяснить своевременность доставки писем и телеграмм, на воздушном транспорте, чтобы узнать распределение пассажиропотоков по целям поездок.

Особым способом статистического наблюдения является способ монографического описания, который представляет собой детальное обследование отдельных единиц совокупности или небольших групп для целей изучения и популяризации передового опыта, анализа и разработки прогрессивных норм, предварительного изучения сложных явлений в порядке подготовки к сплошному или выборочному наблюдения, предварительного научного анализа общественных явлений.

В зависимости от задач статистического исследования и характера исследуемого явления учет фактов можно производить систематически, постоянно охватывая факты по мере их возникновения – текущее наблюдение, или регулярно, но не постоянно, а через определенные промежутки времени – периодические наблюдения, или, наконец, в порядке единовременного, разового наблюдения.

При текущем наблюдении регистрируются постоянно возникающие факты. Таковы, например, факты, характеризующие работу и результаты деятельности предприятий, или затраты, связанные с этой деятельностью. Текущее наблюдение позволяет с исчерпывающей полнотой учесть развитие того или иного процесса или явления во времени, что очень важно для планового ведения хозяйства. Так организована статистика продукции, издержек производства, статистика грузооборота и т. д. Статистическая отчетность по этим показателям имеет в своей основе непрерывный во времени учет фактов, хотя сами формы отчетности, в которых даются уже сводные итоги, можно представлять по укрупненным периодам (месяцам, кварталам и т. д.).

Периодическими и единовременными наблюдениями являются такие, которые проводятся не постоянно, а через определенные промежутки времени или одновременно. Чаще всего эти наблюдения характеризуют состояние каких – либо явлений на определенный момент времени (перепись запасов материалов, учет численности рабочих и служащих, перепись населения и т. д.).

Глава 2. СВОДКА И ГРУППИРОВКА МАТЕРИАЛОВ СТАТИСТИЧЕСКОГО НАБЛЮДЕНИЯ

В результате статистического наблюдения получают материалы, которые содержат данные о каждой единице совокупности. Дальнейшая задача заключается в том, чтобы привести эти материалы в определенный порядок, систематизировать их на этой основе дать сводную характерисмтику всей совокупности фактов при помощи обобщающих статистических показателей. Этого достигают при помощи статистической сводки.

Статистическая сводка – это научная обработка материалов статистического наблюдения для получения обобщающих показателей. Это вторая стадия статистического исследования.

Составными элементами, определяющими содержание сводки, являются:

а) статистические группировки;

б) разработка системы показателей, характеризующих совокупность в целом и ее отдельные группы;

в) подсчет групповых и общих итогов;

г) оформление конечных результатов сводки в статистических таблицах.

Значение правильной научной организации статистической сводки для статистического исследования очень велико.

В итоге статистического наблюдения может быть собран большой материал, правдиво и всесторонне характеризующий действительность. Но, если разработка этого материала будет произведена неправильно, то на основе собранного материала нельзя будет сделать объективных обобщающих выводов.

2.1. Виды статистических группировок.

Сводка является простой, если статистические сведения подытоживаются для всех единиц совокупности, и групповой, если сведения суммируются для однородных групп единиц совокупности. Выделение в совокупности однородных групп единиц называется группировкой. Группировка представляет собой расчленение совокупности на однородные, типичные группы по существенным для них признакам.

Группировка является приемом не только обобщения, но и анализа данных. С ее помощью решаются следующие аналитические задачи:

- в совокупности выделяются качественно однородные типы явлений;
- характеризуются структура явления и структурные сдвиги, происходящие в нем;
- выявляются связи и закономерности между явлениями.

Группировки, реализующие эти аналитические задачи, называются соответственно типологическими, структурными и аналитическими.

Типологическая группировка – группировка, с помощью которой в изучаемой совокупности явлений выделяются однокачественные в существенном отношении группы, прежде всего классы и социально – экономические типы. Типологическими являются, например, группировки предприятий по формам собственности, финансовому состоянию, населения – по общественным группам. Для выделения типов часто берутся не отдельные изолированные признаки, а совокупность признаков.

В структурной группировке производится расчленение однородной в качественном отношении совокупности единиц на группы, характеризующие строение совокупности, ее структуру. Например, можно группировать семьи по числу их членов и в динамике проследить тенденции изменение числа и удельного веса семей разной величины в их общей совокупности; состав населения по полу, возрасту, уровню образования, роду занятий и т. п. Сопоставление данных структурной группировки во времени дает представление о структурных сдвигах.

Аналитическая группировка – группировка, выявляющая связи между изучаемыми явлениями и их признаками. Взаимосвязанные признаки делятся на признаки факторные и признаки результативные. При этом группы образуются по факторному признаку, а для каждой выделенной группы рассчитывается либо среднее значение результативного признака, если он количественный, либо относительные величины, если признак альтернативный (разновидность качественного признака) взаимосвязь проявляется в систематическом изменении значений результативного признака в связи с изменением признака факторного. Так, с ростом стажа работы в большинстве случаев растет выработка рабочих. Количественно эту связь между стажем работы и выработкой можно охарактеризовать с помощью группировки, в которой отдельным значениям стажа работы (факторный признак) соответствует средняя выработка рабочих (результативный признак), имеющих этот стаж работы.

2.2. Выбор группировочных признаков и основные правила образования групп.

Для практического осуществления группировки необходимо выявить группировочные признаки, установить по каждому из них интервалы группировки, найти, какие из признаков должны комбинироваться друг с другом, определить для каждой группы показатели, которыми должны характеризоваться группы.

Признаки можно классифицировать следующим образом.

По форме признаки наблюдения делятся на количественные и качественные. Качественные признаки выражаются числом (возраст, заработная плата, стаж работы , разряд и т.д.) и характеризуют размеры явления. Качественные признаки характеризуют качественное состояние данного явления и не могут быть выражены числом (пол, специальность, образование, тип самолета и т. д.)

Разновидностью качественного признака является альтернативный признак, который имеет два противоположных значения. Например, учится – не учится, работает – не работает.

По характеру вариации (изменение) количественные признаки делятся на дискретные и непрерывные. Дискретными (прерывными) называются признаки, которые отличаются друг от друга на конечную величину (обычно целые числа), т. е. даны в виде прерывных чисел, между которыми не может быть никаких промежуточных значений. Например, число работающих членов семьи может быть выражено следующими значениями варьирующего признака: 0,1,2,3, и т. д., не может быть семей с дробным числом членов. Непрерывный признак: количество книг в библиотеке, число станций в метро имеет иной характер. Его варианты (значения) могут отличаться друг от друга на сколь угодно малую величину, принимать любое значение в определенных пределах и выражаться лишь приближенно. Примером может служить урожайность, производительность, рост человека и др.

По содержанию признаки делятся на факторные и результативные. Факторные признаки характеризуют условия, определяющие размер того или иного явления, а результативные – результат, вызванный влиянием факторных признаков. Один и тот же признак в одних взаимосвязях может быть факторным, в других – результативным. Например, стаж работы – факторный признак, а заработная плата – результативный, стаж работы – результативный, а разряд рабочего – факторный.

В зависимости от единиц измерения признаки бывают натуральные, стоимостные, трудовые. Натуральные признаки характеризуют явления в натуральном выражении (т., км, шт. и т. д.), стоимостные – в денежном выражении (руб., тыс. руб. и т. д.), трудовые – в чел.-часах и т. п.

По способу определения признаки делятся на первичные и вторичные. Признаки, непосредственно характеризующие каждую единицу совокупности, называются первичными, производные от них – вторичными. Например, объем отправок аэропорта – первичный признак, а средний объем отправок на один самолето-вылет производный вторичный признак.

Признаки единиц совокупности, положенные в основание группировки признаками статистического материала, называются группировочными.

В области экономических явлений выбор группировочных признаков в каждом конкретном случае должен быть обоснован экономической теорией. Только на основе правильного научного анализа законов развития явления могут быть определены необходимые признаки, которые должны быть положены в основание статистических группировок.

Можно сформулировать три основных правила для выбора группировочных признаков:

1) в основание группировки необходимо положить наиболее существенные признаки, отвечающие задачам исследования;
2) при этом следует исходить из тех конкретных исторических и территориальных условий, в которых протекает процесс развития изучаемого явления, т. к. с изменением конкретных условий могут меняться и группировочные признаки;
3) при изучении явлений, на которые воздействует несколько различных закономерностей, необходимо в основании группировки закладывать не один, а несколько признаков, взятых в комбинации, т. е. применять комбинационные группировки.

Группировка, произведенная по одному признаку, называется простой. Группировка, произведенная по сочетанию (комбинации) двух или более признаков, называется комбинационной.

2.3. Ряды распределения.

Простейшим примером группировки является ряд распределения.

Статистическим рядом распределения называют упорядоченное распределение единиц совокупности на группы по изучаемому признаку. В зависимости от группировочного признака ряды могут быть вариационные (количественные) и атрибутивные (качественные).

В ряду распределения различают следующие элементы: варианты и частоты, или частости. Вариантами (x) называют отдельные значения группировочного признака, которые он принимает в ряду распределения. Числа, которые показывают сколько раз (как часто) встречается в совокупности то или иное значение признака, или, что тоже самое, сколько единиц в совокупности обладает тем или иным значением признака, называют частотами f_i. Частость f_i - это относительная величина, определяющая долю частот от отдельных вариант в общей сумме частот. Сумма всех частостей равна единице Частости могут выражаться и процентах, тогда сумма всех частостей равна 100%. Термин «частость» применяется также и к показателю доли (удельного веса) единиц, обладающим определенным процентом признака, полученного по данным выборки.

Ряды распределения могут быть образованы как по атрибутивным признакам, так и по количественным. В соответствии с этим они делятся на атрибутивные и вариационные ряды распределения.

Вариационные ряды могут быть дискретными и интервальными. Дискретный ряд распределения – это ряд, в котором варианты выражены одним конечным числом. Интервальный ряд – это ряд, в котором значение признака заданы в виде интервала. При чем, интервалом называется разность между максимальным значением признака в каждой группе.

При построении интервальных рядов необходимо определить количество групп, величину интервала и какие взять интервалы (равные, неравные, открытые, закрытые). Эти вопросы решаются на основе Экономического анализа сущности изучаемых явлений, поставленной цели и характера изменений признака.

Величина интервала определяется следующим образом:

1) когда число групп (k) оговаривается исследованием

$$i = \frac{x_{\max} - x_{\min}}{k},$$

где: i - величина интервала;

$x_{\max}$ - максимальное значение признака;

$x_{\min}$ - минимальное значение признака.

2) когда число групп не оговаривается исследованием, ве личина интервала определяется по формуле, предложенной американским ученым Г.А. Стерджессом:

$$i - \frac{x_{\max} - x_{\min}}{1 + 3{,}322 \lg n},$$

где: n - число наблюдений признака (объем совокупности).

В основу группировки могут быть положены атрибутивные (качественные) признаки. В этом случае число групп чаще всего определяется числом значения признака.

Группировки по атрибутивным признакам часто называют классификациями, т. е. это такие группировки, в основании которых положены существенные признаки и которые имеют устойчивую номенклатуру групп и подгрупп (классификация отраслей, видов производства, товаров, профессий и др.)

Ряды распределения дают возможность судить о закономерности распределения и о границах варьирования совокупности. Различные обобщающие показатели, которые рассматриваются в следующих главах (средние, мода, медиана, дисперсия и т. д.) исчисляются на основе ряда распределения.

Ряды распределения удобнее всего изучить при помощи графического метода, предварительно составив ранжированный ряд, в котором отдельные значения величины варьирующего признака располагаются в порядке их возрастания или убыванию (по ранжиру).

Для облегчения анализа рядов распределения их изображают при помощи графиков в виде полигона и гистограмм.

Полигоном называется график, на котором ряд распределения изображается в вид линейной программы. По оси абсцисс откладываются значения варьирующего признака в порядке их возрастания или убывания, а по оси ординат – частоты. Соответствующие точки пересечения соединяются прямыми. Полигон применяется обычно для дискретных рядов распределения (рис. 1).

Гистограммой называется график, на котором ряд распределения изображается я виде смежных друг с другом столбиков. Гистограмма применяется для интервальных рядов. В этом случае нельзя считать, что единица совокупности каждого интервала имеет одно и тоже значение признаков, т. е. сосредотачивается в одной точке. Поэтому вариационные интервальные ряды изображаются в виде столбиковой диаграммы (рис.2). Высота столбиков просчитывается так, чтобы была пропорциональной частотам. Ширина столбика при равных интервалах одинакова, при неравных – неодинакова.

При построении гистограммы распределение вариационного ряда с неравными интервалами для устранения влияния величины интервала на распределение совокупности, высоту столбиков изображают пропорционально не частотам, а плотности распределения признака соответствующих интервалах.

Плотность распределения – это количество случаев, приходящихся на единицу ширины интервала варьирования признака. Если, например, в распределении магазинов по объему товарооборота товарооборот 40 магазинов находится в пределах от 900 – 1000 тыс. рублей, т. е. отношение $\frac{40}{1000-900} = 0{,}4$ и будет плотностью распределения данного интервала.

Форма полигона и гистограммы зависит от характера распределения. В статистике чаще всего приходится иметь дело с такими формами распределения:

- симметричное распределение, в котором частота средней варианты наибольшая, а убывание частот признака по обе стороны от средней происходит равномерно;
- ассиметричное распределение, в котором максимальная частота варианты или вершина кривой сдвинута влево либо вправо и, следовательно, частоты по обе стороны от максимальной изменяются неодинаково. Если вершина кривой такого распределения сдвинута влево, то ассиметрия называется правосторонней. Если же вершина кривой сдвинута вправо, то ассиметрия называется левосторонней.

ГЛАВА 3. АБСОЛЮТНЫЕ И ОТНОСИТЕЛЬНЫЕ ВЕЛИЧИНЫ

Первые два этапа статистического исследования – это статистическое наблюдение и сводка и группировка статистических данных. В результате статистического наблюдения получают статистические материалы, где каждая единица изучаемой совокупности характеризуется рядом признаков. При статистической сводке и группировке собранные материалы о каждой единице совокупности, то есть индивидуальные показатели, относящиеся к какой-то одной единице, обобщают. В результате получают сводные статистические таблицы, в которых различными показателями дается характеристика не каждой единицы, а совокупности единиц (фактов) в целом и по группам при помощи обобщающих статистических показателей. Показатели, которыми статистика характеризует совокупности единиц, соединенных в группы, или в целом, называются обобщающими показателями.

Обобщение – конечная задача статистики. Метод обобщающих показателей является характерным, специфическим для статистики, как и метод группировок и метод массовых наблюдений. Обобщающие показатели в статистике, как и индивидуальные, могут быть абсолютными и относительными, кроме того, обобщающие показатели могут быть и средними величинами.

Одним из видов обобщающих показателей являются **абсолютные величины**. Их получают непосредственно в результате суммирования статистического материала. На основе таких абсолютных величин исчисляют относительные и средние величины.

Абсолютные величины выражают объем или размеры тех или иных общественных явлений, взятых в определенных границах времени и места.

Например, численность населения России по последней переписи в 2002 году составляла 145,2 млн. чел., за 2005 год воздушным транспортом России было перевезено 35,2 млн.чел.

Абсолютные величины всегда имеют единицу измерения.

При получении итога по всем группам могут быть два случая:

- когда все итоги даны в одинаковых единицах измерения (перевозка грузов на воздушном транспорте – в тоннах, площадь – в кв.м.);
- когда отдельные итоги даются не в одинаковых или в неравноценных единицах измерения. Например, 10 т угля и 12 т нефти. Итог определяется в ккал. через удельную теплоемкость. Итоги могут быть неравноценны с точки зрения выполненной производственной работы. Например, за год железнодорожным транспортом было перевезено 1616,9 млн.т груза. За этот же период автомобильным транспортом было перевезено 6474,7 млн.т груза. С точки зрения выполненной работы эти оценки неодинаковы, т.к. среднее расстояние перевозки одной тонны груза транспортом 805 км, а автомобильным – 11,2 км. Следовательно,объем работы в ткм составит:

3202 млрд ткм - на ж/д транспорте,

76,8 млрд. ткм – на автомобильном транспорте.

Например, 150 автомобилей работало 20 дней, а 40 – 15 дней. Здесь применяется комбинированная единица – машинодни.

Однако для анализа и получения выводов абсолютных единиц недостаточно, т.к. они не могут отразить качественные стороны их явлений. Их необходимо сопоставить с другими абсолютными величинами.

Относительная величина в статистике – обобщающий показатель, который дает числовую меру соотношения двух сопоставляемых абсолютных статистических величин.

Относительные величины широко применяют в экономических исследованиях, при анализе работы отраслей экономики и отдельных предприятий.

В зависимости от характеристики связи между двумя сравниваемыми величинами относительные величины имеют следующий вид:

1) относительные величины выполнения плана;

2) относительные величины структуры;
3) относительные величины координации и сравнения;
4) относительные величины динамики;
5) относительные величины интенсивности.

Относительные величины выполнения плана дают количественную характеристику выполнения плановых заданий. Их широкое применение в экономическом анализе обусловлено практикой планирования в отраслях экономики. Способы расчета относительных величин выполнения плана зависят от характера показателей, выражающих плановое задание.

Так, для экономических явлений, которым свойственно поступательное развитие во времени, плановыми заданиями обычно устанавливается достижение в предстоящих периодах тех или иных абсолютных (или средних) уровней. Относительные величины выполнения плана определяются для них как процентное отношение фактически достигнутой в отчетном периоде абсолютной величины уровня к абсолютной величине уровня планового задания:

$$\text{Относительная величина выполнения плана (\%)} = \frac{\text{фактическое выполнение}}{\text{плановое задание}} \times 100\%.$$

Например, предприятия машиностроения за отчетный период должны выпустить продукции на сумму 63,0 млн. у.е., фактический выпуск составил 66,4 млн. у.е.

Тогда относительная величина выполнения плана выпуска продукции в отчетном году $= \frac{66{,}4}{63} \times 100 = 105{,}4\%$

т.е. процент выполнения плана равен 105,4%.

Для некоторых явлений задания плана предусматривают не рост, а снижение уровней на ту или иную величину. Относительные величины выполнения плана в таких случаях определяются путем сравнения фактически достигнутого и запланированного снижения уровня:

$$\text{Относительная величина выполнения плана (\%)} = \frac{\text{фактическая величина снижения}}{\text{плановая величина снижения}} \times 100\%$$

Например, ремонтный завод планирует снизить себестоимость продукции на 10 у.е., фактическое снижение составило 10,5 у.е.

Следовательно,

относительная величина выполнения плана по снижению себестоимости продукции (%) $= \frac{10{,}5}{10} \times 100\% = 105\%$

В экономическом анализе плановое задание может быть выражено и в форме относительной величины, то есть в виде коэффициента роста или прироста уровня в планируемом периоде по сравнению с уровнем базисного периода. В этом случае относительная величина выполнения плана определяется из процентного сопоставления коэффициента фактического роста явления с плановым коэффициентом:

$$\text{Относительная величина выполнения плана (\%)} = \frac{\text{коэффициент фактического роста}}{\text{коэффициент планового задания}} \times 100.$$

Например, в авиакомпании планом предусматривалось увеличить в отчетном году объем перевозок по сравнению с предыдущим годом на 2,8%. Фактический рост перевозок составил 5%. Следовательно, % выполнения плана по росту перевозок составил

$$\frac{105}{102{,}8} \times 100\% = 102{,}1\%$$

Относительная величина структуры характеризует долю (удельный вес) составных частей целого в их общем итоге и обычно выражается в виде коэффициентов (доли единиц) или процентах.

Важное значение относительных величин структуры в экономической статистике состоит в том, что они применяются для изучения состава (строения) статистической совокупности. При определении относительных величин структуры сравниваемыми величинами могут быть или численности отдельных групп совокупности, или объемы их признаков. За основание (базу) сравнения принимается общий итог статистической совокупности.

Например, имеются следующие данные по России о перевозке грузов всех видов транспорта общего пользования (тыс.т) за 2003 год.

Следует определить относительные величины структуры перевозки грузов отдельных видов транспорта в 2003 году.

Структура перевозки грузов по видам транспорта общего пользования (млн.т)

Виды транспорта	2003г.	Показатели структуры,%
Транспорт всего	2731,8	100
В том числе: Железнодорожный	1161	42,50
Автомобильный	471	17,24
Трубопроводный	975	35,69
Морской	24	0,88

Внутренний водный	100	3.66
Воздушный	0,8	0,03

Важное значение при анализе экономических явлений имеют **относительные величины координации**. В отличие от относительных величин структуры, выражающих удельные веса частей в целом, относительные величины координации характеризуют соотношение частей изучаемой совокупности, которое показывает, во сколько раз сравниваемая часть явления больше или меньше части, принимаемой за основание (базу) сравнения. Относительные величины координации выражаются в кратных отношениях.

Например, по переписи населения в 2002 году городское население составило 106,4 млн. чел., сельское – 38,8 млн.чел. Для определения относительной величины координации численности городского и сельского населения следует сопоставить между собой исходные данные, приняв за базу сравнения численность сельского населения:

$\frac{106{,}4}{38{,}8} = 2{,}74$, т.е. в 2002 г. в России на каждые 100 человек сельских жителей приходилось 274 чел. городских жителей.

В экономической статистике часто приходится сопоставлять значения одноименных признаков по нескольким совокупностям. В результате получают **относительные величины сравнения**.

Например, в августе месяце по воздушной линии Москва-Сочи перевезли 18 тыс.чел., а по воздушной линии Москва-Анапа – 22 тыс.чел., следовательно, относительная величина сравнения равна

$$\frac{22}{18} = 1{,}22$$

Относительная величина динамики характеризует развитие изучаемого явления во времени. Они позволяют при анализе данных, характеризующих развитие явления по времени, выявлять направление развития и измерять темпы роста.

Относительные величины динамики (темпы роста) исчисляются как отношения абсолютных (или средних) уровней и выражаются в форме коэффициентов или процентов.

Например, имеются данные о численности студентов принятых в высшие и средние специальные заведения (тыс.чел.).

Годы	Число студентов	Изменение численности по сравнению с 1999г.	
		%	Коэффициент
1999	1059	100,0	1,00
2000	1293	122,0	1,22

2001	1462	138,0	1,42
2002	1504	142,0	1,42
2003	1643	155,0	1,55

Относительные величины интенсивности характеризуют степень насыщенности изучаемым явлением определенной среды. Они выражают соотношение разноименных, но связанных между собой величин и исчисляются как отношение величины изучаемого явления к объему той среды, в которой происходит развитие явления.

Относительные величины интенсивности являются именованными числами и могут выражаться в кратных отношениях, процентах и других формах.

Например, имеются следующие данные по району за 2005 год:

Число родившихся за год детей - 2701

Среднегодовая численность населения - 94980

Требуется определить относительную величину интенсивности, характеризующую рождаемость детей.

Для этого следует определить коэффициент рождаемости детей в 2005 году.

$$\text{Коэффициент рождаемости} = \frac{\text{число родившихся}}{\text{среднегодовая численность населения}} = \frac{2701}{94980} = 0{,}028,$$

т.е. рождаемость детей в районе в расчете на 1000 человек населения составляла 28 человек.

Глава 4. СРЕДНИЕ ВЕЛИЧИНЫ

Среди обобщающих показателей, которыми статистика характеризует общественные явления и свойственные им закономерности, большое значение имеют средние величины. Это объясняется тем, что статистика изучает совокупности по варьирующим признакам, среди которых важное место занимают признаки, у которых вариация проявляется в форме различных количественных значений.

Например, нужно определить уровень месячной заработной платы, характерный для рабочих одного предприятия. Нельзя на основе заработной платы какого – либо одного рабочего сказать, что она характеризует этот уровень, т. к. заработная плата других рабочих обычно отличается от этой заработной платы. Не может характеризовать этот уровень и общая сумма заработной платы, начисленная за месяц рабочим предприятия, т. к. она зависит от численности рабочих. Однако можно исключить влияние численности рабочих, если общую сумму месячной зарплаты разделить на число рабочих. В результате такого деления получим показатель средней заработной платы, который и будет обобщенной характеристикой всей совокупности рабочих по этому признаку.

Средней величиной в статистике называется обобщенная характеристика совокупности однотипных явлений по какому-либо варьирующему признаку, которая показывает уровень признака, отнесенный к единице совокупности.

Существует несколько видов средних величин:

1) средняя арифметическая;
2) средняя гармоническая;
3) средняя квадратическая;
4) средняя геометрическая.

Существуют еще описательные характеристики распределения варьирующего признака – мода (наиболее часто встречающаяся варианта) и медиана (средняя варианта). Их иногда называют структурными средними.

Средние квадратическая и гармоническая будут рассмотрены в других главах.

4.1.Средняя арифметическая и средняя гармоническая.

Важным средством анализа являются средние величины, в обобщенной форме характеризующие типичный уровень того или иного признака изучаемой совокупности.

Наиболее часто средний уровень значений признака исчисляется по формулам средней арифметической и средней гармонической.

Средняя величина исчисляется как средняя арифметическая в тех случаях, когда имеются данные об отдельных значениях признака и о числе единиц совокупности, обладающих этим значением.

Пример. Одним из признаков, характеризуемых при изучении семей, является число членов семьи. Распределение 100 семей (включая одиночек) по числу членов семьи приведено в таблице 4.1.

Таблица 4.1.

Распределение семей по числу членов семьи

Число членов семьи (x)	Число семей (f)	Общее число людей в группе семей (fx)
1	9	9
2	15	30
3	22	66
4	30	120
5	16	80
6	5	30
7	2	14

8	1	8
Итого	$\sum f = 100$	$\sum fx = 357$

Среднее число членов семьи может быть получено, если общее число людей, составляющих, 100 семей, поделить поровну между семьями. Математическое выражение этой операции в приведенных обозначениях

$$\bar{x} = \frac{\sum xf}{\sum f} = \frac{357}{100} = 3{,}57 \text{ (члена семьи).}$$

Числа f представляют собой частоты или веса значений, признака. Поэтому формула (1) получила название средней арифметической взвешенной. Если бы любое значение признака имело одинаковый вес, то формула (1) стала бы формулой средней арифметической простой

$$\bar{x} = \frac{\sum x}{n},$$

где n – число значений признака.

Если в результате сводки имеются показатели $\sum fx$ и $\sum f$, то расчет средней арифметической по формуле (1) сводится к делению первого показателя на второй.

Во многих случаях для расчета средней имеются не три ряда данных: x, f, xf, а только два: x и xf. Тогда ее расчет производится по формуле, являющейся преобразованием формулы (1), путем подстановки в нее известных величин x и xf. Обозначим xf через w, откуда

$$f = \frac{w}{x} \text{ и } \bar{x} = \frac{\sum xf}{\sum f} = \frac{\sum x\frac{w}{x}}{\sum \frac{w}{x}} = \frac{\sum w}{\sum \frac{w}{x}}.$$

Полученная формула называется средней гармонической взвешенной.

Таким образом, средняя арифметическая и средняя гармоническая характеризуют один и тот же уровень значения признака, исчисленный, однако, исходя из разной исходной информации: в первом случае из x и f, а во втором – из x и xf. Поэтому при практических расчетах важно учитывать, что же дано – частоты или произведения (частоты, умноженные на значения усредняемого признака).

Рассмотрим пример для средней арифметической и средней гармонической для рядов распределения.

Пример №1. имеются следующие данные:

Группы рабочих по стажу, лет	Число рабочих f	Середина интервала x^1	$x^1 f$
3-5	10	4	40
5-7	30	6	180
7-9	40	8	320
9-11	15	10	150
11-13	5	12	60
Итого	100		750

В данном ряду варианты представлены не одним числом, а в виде интервала. Таким образом каждая группа ряда распределения имеет нижнее и верхнее значения вариант, иначе закрытые интервалы. Расчет проводится по средней арифметической взвешенной:

$$\bar{x} = \frac{\sum x^1 f}{\sum f},$$

где: x^1 - центральное значение признака в каждой группе (середина интервала).

$$\bar{x} = \frac{\sum x^1 f}{\sum f} = \frac{750}{100} = 7{,}5 \text{ лет}$$

Преобразуем рассмотренный выше ряд распределения в ряд с открытыми интервалами.

Группы рабочих по стажу, лет	Число рабочих
До 5	10
5-7	30
7-9	40
9-11	15
11 и более	5
Итого	100

В таких рядах условно величина интервала в первой группе принимается равной величине интервала в последующей группе, а величина интервала в последней группе – величине интервала в предыдущей группе. Дальнейший расчет аналогичен изложенному выше.

Пример №2. рассчитать среднюю дальность перевозки 1 т груза по трем автобазам:

Автобаза	Средняя дальность перевозки 1 т/км «х»	Грузооборот, тыс.ткм «W»
1	120	126
2	90	90
3	10	120

Средняя дальность перевозки = грузооборот/количество перевезенного т груза.

Следовательно, грузооборот равен произведению средней дальности на количество перевезенного груза, т. е. $W = xf$. Таким образом, в условии задачи дано «W» и «х», а «f» отсутствует.

В этом случае следует использовать среднюю гармоническую:

$$\bar{x} = \frac{\sum W}{\sum \frac{W}{x}} = \frac{126 + 90 + 100}{\frac{126}{120} + \frac{90}{90} + \frac{120}{10}} = \frac{316}{3{,}25} = 97 \text{ км.}$$

Средняя арифметическая в отличие от других средних величин обладает следующими свойствами:

1) произведение средней на сумму частот всегда равно сумме произведений вариант на частоты:

$$\bar{x}\sum f = \sum xf;$$

2) если каждую варианту увеличить или уменьшить на какое-либо произвольное число A, то новая средняя увеличится или уменьшится на то же самое число:

$$\frac{\sum (x \pm A) f}{\sum f} = \bar{x} \pm A;$$

3) если каждую варианту разделить или умножить на какое-либо произвольное число A, то средняя уменьшится или увеличится во столько же раз:

$$\frac{\sum \frac{x}{A} f}{\sum f} = \frac{\bar{x}}{A}; \qquad \frac{\sum xAf}{\sum f} = \bar{x}A;$$

4) если все частоты (веса) разделить или умножить на какое-либо число A , то средняя арифметическая от этого не изменится

$$\bar{x} = \frac{\sum xf}{\sum f} = \frac{\sum x\frac{f}{A}}{\sum \frac{f}{A}} = \frac{\sum xfA}{\sum fA};$$

5) сумма отклонений вариант всегда равняется нулю

$$\sum (x - \bar{x})f = 0, \qquad \sum (x - \bar{x}) = 0.$$

Перечисленные свойства средней арифметической используются в основном для упрощения расчета средней, основанного на способе условных моментов, который будет подробно рассмотрен ниже.

4.2. Мода и медиана

Достоинством средней как обобщающего показателя является то, что она одной величиной характеризует целую совокупность различных величин. Но для всесторонней характеристики совокупности, так и для решения некоторых практических задач, нужны и такие обобщающие показатели, которые характеризуют особенности распределения единиц совокупности по величине изучаемого признака. К таким показателям относятся мода и медиана, которые называются структурными средними.

Модой в статистике называется наиболее часто встречающееся значение признака в вариационном ряду, т. е. варианта, у которой частота (вес) наибольшая.

В дискретном вариационном ряду будет варианта, имеющая наибольшую частоту.

В интервальном вариационном ряду, в котором, как известно, не приводятся все значения признака, а указываются их группы в виде интервалов, мода исчисляется по следующей формуле:

$$M_0 = x_{m_0} + i\frac{f_{m_0} - f_{m_0 \mp 1}}{(f_{m_0} - f_{m_0 \mp 1}) + (f_{m_0} - f_{m_0 \pm 1})},$$

где M_0 - мода;

x_{m_0} -нижняя граница модального интервала;

f_{m_0} - частота модального интервала

$f_{m_0 -1}$ - частота интервала, предшествующая модальному;

$f_{m_0 +1}$ -частота интервала, следующего за модальным;

i - величина модального интервала.

Мода имеет важное значение для решения некоторых задач, например, какой размер обуви чаще всего встречается у мужчин и женщин, какое время дня является «пиковым» для работы предприятий общественного питания, городского транспорта и др.

Медианой в статистике называется варианта, расположенная в середине вариационного ряда.

Если ряд распределения дискретный и имеет нечетное число членов, то медианой будет варианта, находящаяся в середине упорядоченного ряда (упорядоченный ряд – это ряд, в котором значения признака расположены в порядке возрастания или убывания). Например, стаж пяти рабочих составил 2, 4, 7, 8, и 10 лет. В таком ряду медиана равна 7 годам.

Если упорядоченный ряд состоит из четного числа членов, то медиана будет определяться как средняя арифметическая из двух вариант, расположенных в середине ряда. Например, пусть будет не пять, а шесть человек со стажем 2, 4, 6, 8, и 10 лет. В этом случае медиана будет определяться как $\frac{6+7}{2}$, т. е. 6, 5 лет.

Если же в дискретном ряду значения признаков повторяются, то для расчета медианы используется сумма накопленных частот. Пример: имеется следующий дискретный ряд распределения рабочих по стажу работы. Определить медиану.

Стаж работы, лет	Число рабочих	Сумма накопленных частот
10	2	2
12	6	8(6+2)
15	16	24(16+8)
17	12	
20	4	
Итого	40	

Для определения медианы надо подсчитать сумму накопленных частот ряда. Наращивание итога продолжается до получения суммы частот немногим больше половины В нашем примере сумма равна 40, ее половина – 20. Накопленная сумма частот ряда получилась равной 24. Варианта, соответствующая этой сумме, т. е. 15 лет, и есть медиана данного ряда.

Смысл полученного результата следующий: одна половина рабочих имеет стаж меньше 15 лет, а другая – больше 15 лет.

Если же сумма накопленных частот против одной из вариант будет равна точно половине суммы частот, то медиана определяется как средняя арифметическая из этой варианты и последующей.

В интервальном ряду распределения медиана распределяется по следующей формуле:

$$Me = x_{Me} + i\frac{\frac{\sum f}{2} - S_{Me-1}}{f_{Me}},$$

где: Me - медиана;

x_{Me} - нижняя граница медианного интервала;

i - величина медианного интервала;

S_{Me-1} - сумма накопленных частот, предшествующих медианному интервалу;

f_{Me} - частота медианного интервала.

Хотя мода и медиана не получили в статистике такого широкого применения, как средние величины, их использование в некоторых экономико-статистических и других расчетах может иметь важное значение. Например, при исследовании производительности труда, норм выработки и т. д. Мода позволяет выявить наиболее распространенный размер или уровень изучаемого явления, а медиана характеризует деление совокупности явления на две половины, одна из которых ниже, а другая выше определенного уровня.

Медиана, кроме того, обладает следующим свойством: сумма отклонений значений признака от медианы по модулю является минимальной величиной, т. е.

$$\sum |x_i - Me| f = \min.$$

Это свойство имеет важное значение для решения некоторых практических задач. Например, определение мест строительства общественных зданий (школ, больниц, детских садов) с точки зрения их приближения к местам проживания населения, расчет самого короткого пути из всех возможных для транспорта и т. д.

Рассмотрим пример расчета моды и медианы.

Распределение предприятий по численности работающих характеризуется следующими данными:

Группировка предприятий по числу работающих	Число предприятий	Сумма накопленных частот
100-200	1	1
200-300	3	4(3+1)
300-400	7	11(7+4)
400-500	30	41(30+11)

500-600 600-700 700-800	19 15 5	
Итого	80	

В этой задаче наибольшее число предприятий (30) имеет численность работающих от400 до 500 чел. Следовательно, этот интервал является модальным интервала ряда распределения. Введем обозначения:

$$x_{Mo} = 400$$
$$i_{Mo} = 100$$
$$f_{Mo} = 30$$
$$f_{Mo-1} = 7$$
$$f_{Mo+1} = 19$$

$$Mo = 400 + 100\frac{30-7}{(30-7)+(30-19)} = 400 + 100\frac{23}{23+11} = 468 чел,$$

Таким образом, среди 80 предприятий наиболее часто встречаются предприятия с численностью работников равной 468 чел.

Чтобы определить моду необходимо сумму накопленных частот чуть больше половины суммы накопленных частот. Против варианты 400-500 сумма накопленных частот равна 41, отсюда следует, что *Me* будет находиться в этом интервале. Ее величина будет равна:

$$Me = 400 + 100\frac{40-11}{30} = 497 чел.$$

Следовательно, половина предприятий имеет численность меньше 417 чел., а половина – более 497 чел.

ГЛАВА 5. ПОКАЗАТЕЛИ ВАРИАЦИИ

Средние величины дают обобщающую характеристику уровня количественного признака по совокупности в целом. В одних случаях отдельные значения признака могут очень незначительно отличаться от средней арифметической, в других случаях, наоборот, отдельные значения далеко отстоят от средней.

Изменение (колеблемость) величины количественного признака от одной единицы однородной совокупности до другой принято называть вариацией. Размеры вариации позволяют судить, насколько однородна изучаемая группа и, следовательно, насколько характерна средняя по группе. Изучение отклонений от средних имеет большое практическое и теоретическое значение, поскольку в отклонениях проявляется развитие явления; небольшие количественные изменения, постоянно нарастая, могут в дальнейшем привести к существенным, качественным сдвигам.

5.1. Абсолютные и относительные показатели

Для характеристики размера вариации используются специальные показатели колеблемости (вариации): размах вариации, среднее линейное отклонение, среднее квадратическое отклонение, коэффициент вариации.

Размах вариации равен разности между наибольшим и наименьшим значениями варьирующего признака в данном ряду распределения. Это наиболее простой, но зато и наименее точный показатель вариации. При его расчете не учитываются колеблемость тех значений признака, которые заключены между его крайними значениями, и частоты разных значений признака. Наименьшее и наибольшее значение признака могут оказаться случайными, нехарактерными для совокупности и существенно отличными от других его значений.

Эти недостатки устраняются, если применить другие показатели вариации.

Среднее линейное отклонение можно рассчитать по той же формуле, что и среднюю арифметическую, но в качестве варианты выступают абсолютные отклонения (модули) значений признака от его среднего значения.

Соответственно формула расчета среднего линейного отклонения ($\overline{d}$) имеет вид:

$$\overline{d} = \frac{\sum \left| x - \overline{x} \right| \times f}{\sum f}$$

Если исчислить среднюю не из абсолютных отклонений значений признака от средней, а из квадратов этих отклонений, то получим показатель дисперсии (σ^2), квадратный корень из которого называют средним квадратическим отклонением (σ):

$$\sigma^2 = \frac{\sum \left(x - \overline{x} \right)^2 \times f}{\sum f}$$ - дисперсия;

$$\sigma = \sqrt{\frac{\sum \left(x - \overline{x} \right)^2 \times f}{\sum f}}$$ - среднее квадратическое отклонение.

Среднее линейное отклонение также как и среднее квадратическое отклонение показывают, насколько в среднем отличаются индивидуальные признаки от среднего его значения.

Выше названные показатели являются **абсолютными показателями вариации**. Эти величины – именованные, они зависят от масштаба измерения признака, поэтому не всегда пригодны для сравнения.

Абсолютными показателями вариации нельзя непосредственно пользоваться в двух случаях:

- для сравнения степени вариации двух различных признаков в одной и той же группе;
- для сравнения вариации по одному и тому же признаку, но в двух различных группах с разным уровнем средних, т.е. значительно чаще отличающихся по объему совокупности.

Поэтому чаще используется относительный показатель вариации, который носит название коэффициента вариации (V) и рассчитывается по формуле:

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

Коэффициент вариации показывает, на сколько процентов в среднем индивидуальные значения признака отличаются от его среднего значения. В известной степени коэффициент вариации является критерием надежности средней: если он велик (более 40%), то это свидетельствует о большой колеблемости в величине признака у отдельных единиц данной группы, а, следовательно, средняя недостаточно надежна.

Рассмотрим расчет этих показателей на основе данных, представленных в таблице:

Группировка рабочих по стажу, лет	Число рабочих	$x^!$	$x^! \cdot f$	$x - \bar{x}$	$\lvert x - \bar{x} \rvert \cdot f$	$(x - \bar{x})^2 \cdot f$
До 2	22	1	22	-3	66	198
2-4	26	3	78	-1	26	26
4-6	32	5	160	1	32	32
6-8	20	7	140	3	60	180
Итого	100		400		184	436

$$\bar{x} = \frac{\Sigma x^! \cdot f}{\Sigma f} = \frac{400}{100} = 4 \text{ года}$$

$$\bar{d} = \frac{\sum\left|x - \bar{x}\right| \times f}{\sum f} = \frac{\sum\left|\bar{d}\right| \cdot f}{\sum f} = \frac{184}{100} \approx 1{,}8 \quad \text{года}$$

$$\sigma = \sqrt{\sigma^2} = \sqrt{\frac{\sum\left(x - \bar{x}\right)^2 \cdot f}{\sum f}} = \sqrt{\frac{436}{100}} = 2{,}1 \text{ года}$$

$$v = \frac{\sigma}{\bar{x}} \cdot 100\% = \frac{2{,}1}{4} \cdot 100\% = 52{,}5\%$$

На основе коэффициента вариации можно сделать вывод, что средняя в данной совокупности не надежна, т.е. совокупность неоднородна.

Дисперсия обладает рядом свойств:

1. дисперсия постоянной величины равна нулю;
2. если все значения вариант (признака) увеличить или уменьшить на некоторую постоянную величину А, то значение дисперсии не изменится;
3. дисперсия увеличится или уменьшится в A^2 раз, если все значения признака умножить или разделить на одну и ту же постоянную величину А;
4. дисперсия меньше среднего квадрата отклонений вариант от произвольной величины А на квадрат разности между средней арифметической и произвольной величиной А;
5. дисперсия суммы независимых случайных величин равна сумме их дисперсий.

5.2. Коэффициент асимметрии

Одной из важнейших задач анализа вариационных рядов распределения является выявление закономерности распределения и определение ее характера. При **симметричном распределении** частот в вариационном ряду обобщающие характеристики ряда – средняя арифметическая, мода и медиана – равны между собой.

Однако в экономической статистике такое распределение встречается крайне редко. Чаще всего наблюдается **асимметрия** распределения, которая может быть правосторонней или левосторонней в зависимости от расположения частот. Если в ряду распределения преобладают варианты с меньшим, чем средняя арифметическая, значением признака, то вершина кривой распределения сдвинута влево и правая часть кривой оказывается длиннее. Такая **асимметрия называется**

правосторонней (положительной). Если же преобладают варианты с большим, чем средняя арифметическая, значением признака – вершина кривой распределения сдвинута вправо и, следовательно, левая часть кривой получается длиннее правой. Такая **асимметрия называется левосторонней** (отрицательной).

Степень асимметрии может быть определена с помощью коэффициента асимметрии:

$$K_A = \frac{\bar{x} - M_0}{\sigma}, \text{ где}$$

$\bar{x}$ - средняя арифметическая ряда распределения;

M_0 - мода;

σ - среднее квадратическое отклонение.

При симметричном (нормальном) распределении $\bar{x} = M_0$, следовательно, коэффициент асимметрии (K_A) равен нулю. Если $K_A > 0$, , то $\bar{x} > M_0$, следовательно, имеется правосторонняя асимметрия.

Если $K_A < 0$, то $\bar{x} < M_0$, следовательно, имеется левосторонняя асимметрия.

Коэффициент асимметрии может изменяться от -3 до +3.

Пример. На примере следующего ряда распределения рассчитать коэффициент асимметрии:

Группировка рабочих по стажу, лет	Число Рабочих	x'	$x' \cdot f$	$\lvert x - \bar{x} \rvert$	$(x - \bar{x})^2$	$(x - \bar{x})^2 \cdot f$
1-2	1	1,5	1,5	3,4	11,56	11,56
2-3	3	2,5	7,5	2,4	5,76	17,28
3-4	7	3,5	24,5	1,4	1,96	13,72
4-5	30	4,5	135	0,4	0,16	4,8
5-6	19	5,5	104,5	0,6	0,36	6,84
6-7	15	6,5	97,5	1,6	2,56	38,4
7-8	3	7,5	22,5	2,6	6,76	20,28
Итого	80		393			112,88

Средний стаж рабочих равен:

$$\bar{x} = \frac{\sum x' \cdot f}{\sum f} = \frac{393}{80} = 4{,}9 \text{ лет}$$

Среднее квадратическое отклонение:

$$\sigma = \sqrt{\frac{\sum (x' - \bar{x})^2 \cdot f}{\sum f}} = \sqrt{\frac{112{,}88}{80}} = 1{,}2 \text{ года}$$

Преобразовав интервальный ряд в дискретный, можно принять M_0 =4,5 года, т.к. данная варианта наиболее часто встречается в ряду распределения (f = 30).

Следовательно, коэффициент асимметрии будет равен:

$$K_A = \frac{\bar{x} - M_0}{\sigma} = \frac{4{,}9 - 4{,}5}{1{,}2} = 0{,}33$$

Таким образом, в данном ряду распределения имеется незначительная правосторонняя (положительная) асимметрия.

5.3. Правило сложения дисперсий

Если совокупность разбита на группы по изучаемому признаку, то для такой совокупности могут быть исчислены следующие виды дисперсий: общая, групповые (частные), средняя из групповых (частных) и межгрупповая.

Общая дисперсия (σ^2) равна среднему квадрату отклонений отдельных значений признака (x) от общей средней ($\bar{x}$). Она может быть исчислена как простая средняя или взвешенная соответственно по формулам:

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

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

Общая дисперсия отражает вариацию признака за счет всех условий и причин, действующих в совокупности.

Групповая дисперсия (σ_i^2) равна среднему квадрату отклонений отдельных значений признака внутри группы от средней арифметической этой группы ($\overline{x}_i$). Она может быть исчислена как простая средняя или взвешенная соответственно по формулам:

$$\sigma^2 = \frac{\sum\left(x - \overline{x_i}\right)^2}{n};$$

$$\sigma^2 = \frac{\sum\left(x - \overline{x}_i\right)^2 \times f}{\sum f}$$

Эта дисперсия отражает вариацию признака за счет условий и причин, действующих внутри группы.

Средняя из групповых дисперсий ($\overline{\sigma}_i^2$) – это средняя арифметическая из дисперсий групповых:

$$\overline{\sigma}_i^2 = \frac{\sum \sigma_i^2}{n}; \qquad \overline{\sigma}_i^2 = \frac{\sum \sigma_i^2 \cdot f}{\sum f}$$

Межгрупповая дисперсия (∂^2) равна среднему квадрату отклонений групповых средних от общей средней:

$$\partial^2 = \frac{\sum\left(\overline{x}_i - \overline{x}\right)^2}{n} \qquad \partial^2 = \frac{\sum\left(\overline{x}_i - \overline{x}\right)^2 \cdot f}{\sum f}$$

Межгрупповая дисперсия характеризует вариацию результативного признака за счет признака группировочного.

Между указанными видами дисперсий существует определенное соотношение: **общая дисперсия равна сумме средней из групповых дисперсий и межгрупповой дисперсии**:

$$\sigma^2 = \overline{\sigma}_i^2 + \partial^2$$

Это соотношение называют **правилом сложения дисперсий**.

Правило сложения дисперсий используется в статистике для определения степени тесноты связи между изучаемыми признаками.

С этой целью рассчитывается **коэффициент детерминации** (η^2), который представляет собой отношение межгрупповой дисперсии к общей, и показывает, какую часть общей вариации изучаемого признака составляет вариация межгрупповая, т.е. обусловленная группировочным признаком.

Корень квадратный из коэффициента детерминации называется **эмпирическим корреляционным отношением:**

$$\eta = \sqrt{\frac{\partial^2}{\sigma^2}}$$

Оно характеризует степень тесноты связи между взаимосвязанными признаками. По его абсолютной величине судят о тесноте связи или степени зависимости признака результативного от одного признака факторного (группировочного) или нескольких.

Эмпирическое корреляционное отношение может изменяться от 0 до 1. Чем ближе его величина к единице, тем связь теснее. Знак указывает на характер, направление связи.

5.4. Метод упрощенного вычисления средней арифметической, дисперсии и среднего квадратического отклонения.

Для совокупности с большим объемом совокупности, а также если значения признака выражены большими числами, процесс вычисления $\bar{x}$, σ, σ^2 значительно усложняется и становится нередко громоздким и трудоемким. Поэтому в статистике разработаны приемы упрощенного вычисления этих показателей. Одним из наиболее эффективных способов упрощения является способ условных моментов, который основан на использовании свойств средней арифметической и дисперсии.

Условным моментом называется средняя арифметическая из отклонений отдельных значений признака от некоторой постоянной величины A, называемой условным началом ($A \neq 0$, $A \neq \bar{x}$). Обычно за условное начало A принимают варианту наиболее часто встречающуюся в совокупности, т.е. моду.

В зависимости от степени, в которой берутся отклонения, моменты бывают разных порядков. Если средняя рассчитывается из отклонений первой степени, получается условный момент первого порядка (M_1), из отклонений второй степени – условный момент второго порядка (M_2) и т.д., т.е.

$$M_1 = \frac{\sum \left(\frac{x - A}{i} \right) \cdot f}{\sum f}; \qquad M_2 = \frac{\sum \left(\frac{x - A}{i} \right)^2 \cdot f}{\sum f}$$

Используя способ условных моментов, можно рассчитать $\bar{x}$, σ, σ^2, по формулам:

$$\bar{x} = A + i + M_1;$$

$$\delta^2 = i^2 \cdot (M_2 - M_1^2);$$

$$\delta = i\sqrt{M_2 - M_1^2}$$

Пример: на основе ряда распределения, характеризующего время горения электроламп, определить среднее время горения электроламп, среднее квадратическое отклонение и дисперсию времени горения электролампы, пользуясь способом условных моментов.

Группы электроламп по времени горения (час)	Число Ламп	x'	$x' - A$	$\frac{x' - A}{i}$	$\left(\frac{x' - A}{i}\right) \cdot f$	$\left(\frac{x' - A}{i}\right)^2 \cdot f$
800-10000	20	900	-400	-2	-40	80
1000-1200	80	1100	-200	-1	-80	80
1200-1400	160	1300	0	0	0	0
1400-1600	90	1500	200	1	90	90
1600-1800	40	1700	400	2	80	160
1800-2000	10	1900	600	3	30	90
Итого	400				80	500

Данный ряд распределения является интервальным.

Для расчета среднего времени горения электролампы используется формула средней арифметической взвешенной:

$$\overline{x} = \frac{\sum x^{!} \cdot f}{\sum f}$$

где $x^{!}$ - центральное значение признака в каждой группе, или, пользуясь способом условных моментов:

$$\overline{x} = A + i \cdot M_1 = 1300 + 200 \cdot \frac{80}{400} = 1340 \text{ ч.},$$

где A - условное начало, т.е. A = 1300ч (варианта, которая наиболее часто повторяется;

M_1 - момент первого порядка.

Дисперсия времени горения электролампы равна:

$$\sigma^2 = i^2 (M_2 - M_1^2);$$

где M_2 - момент второго порядка.

$$\sigma^2 = 200^2 \left[\frac{500}{400} - \left(\frac{80}{400} \right)^2 \right] = 200^2 (1{,}25 - 0{,}04) = 48400 \text{ ч.}$$

Среднее квадратическое отклонение:

$$\sigma = \sqrt{\sigma^2} = i \cdot \sqrt{M_2 - M_1^2} = \sqrt{48400} = 220 \text{ ч.}$$

ГЛАВА 6. ИНДЕКСЫ

Среди обобщающих статистических показателей одно из важных мест принадлежит индексам, т. е. относительным показателям сравнения общественных явлений во времени, в пространстве или в сопоставлении с планом.

По форме выражения индекс представляет относительную величину. Но если всякий индекс является относительной величиной, то не всякую относительную величину можно назвать индексом. Различие в данном случае состоит в следующем: относительная величина характеризует изменение какого-либо единичного явления или такого массового явления, отдельные элементы которого можно суммировать; индекс же характеризует относительное изменение какого-либо сложного явления, состоящего из качественно однородных, но не поддающихся суммированию элементов. Кроме того, отдельные виды относительных величин могут выражать соотношение как одноименных, так и разноименных явлений. Индекс же всегда характеризует соотношение только одноименных явлений, от которых он и получает свое название: индекс производительности труда, индекс продукции, индекс себестоимости и т. д.

Таким образом, **под индексом в статистике понимают количественный показатель соотношения во времени, в пространстве или в сравнении с планом массовых общественных явлений, состоящих из качественно однородных, но непосредственно не поддающихся суммированию элементов.**

Например, нельзя суммировать тонны хлопка, предположим, с тоннами железа, нефти и т. д. Но надо знать обобщенную характеристику общего объема произведенной продукции, чтобы выяснить ее динамику или сравнить эти общие объемы в разных странах. Такую характеристику приходится давать при помощи специально построенных показателей – индексов физического объема произведенной (проданной или потребленной) продукции.

С такого же рода совокупностями приходится сталкиваться и тогда, когда возникает потребность дать сводную характеристику изменения общего уровня цен, т. к. и в этом случае цены отдельных товаров нельзя складывать. Приходится рассчитывать специальные показатели – индексы цен. К такого рода показателям относятся индексы себестоимости производства продукции, индексы производительности труда.

Чтобы измерить, как изменился объем (количество) разнородной продукции, нужно принять для сравниваемых периодов одинаковые цены, а чтобы измерить, как изменился уровень цен по группе разнородной продукции, необходимо в индексе исключить изменение ее количества. В этом и заключаются специфические приемы индексного метода.

Кроме использования индекса в качестве сводного показателя, индексный метод применяется в статистике для оценки роли отдельных факторов в изменении сложного явления. К таким сложным явлениям может быть отнесена стоимость произведенной продукции. Ее изменение обусловлено изменением количества произведенной продукции и изменением цен. Предполагая, что цены неизмененные, мы узнаем, как повлияло изменение количеств на общую динамику стоимости продукции, а принимая неизменными количества, узнаем как повлияло изменение цен. Таким образом, для анализа роли факторов применяются не отдельные индексы, а система взаимосвязанных индексов.

6.1. Классификация индексов.

Индексы классифицируются по ряду признаков:

1. В зависимости от объекта исследования различают:

а) **индексы количественных показателей** (индексы физического объема продукции, розничного товарооборота, потребления и т.д.). Во всех этих индексах количества оцениваются в одинаковых, неизменных ценах, обычно в ценах базисного периода;

б) **индексы качественных показателей** (индексы цен, себестоимости, производительности труда). Все эти индексы вычисляются на базе одинаковых, неизменных количеств продукции, обычно взятых из отчетного периода. В эту группу входят показатели, которые рассчитываются вторичным способом. Например, чтобы определить себестоимость продукции, следует затраты разделить на количество производственной продукции.

2. **С точки зрения охвата элементов совокупности индексы разделяются на:**

а) индивидуальные;

б) общие;

в) групповые.

Индивидуальные индексы дают сравнительную характеристику отдельных элементов той или иной совокупности.

Например, во 2 квартале авиакомпанией было перевезено 463,5 тыс.чел., а в 1 квартале – 391 тыс.чел. Следовательно, во 2 квартале пассажиров было перевезено в 1,18 раза больше (463,5/391,0), чем в 1 квартале. 1,18 – это индивидуальный индекс объема перевозок пассажиров.

Общие индексы характеризуют изменение совокупности в целом. Если индекс охватывает не все элементы совокупности, а только какую-то часть, какую-то группу их, то такие индексы называются **групповыми.**

Например, в одном из городов России в 2005 году по сравнению с 2004 годом выпуск товаров народного потребления увеличился на 8%. При общем увеличении товаров народного потребления на 8% выпуск товаров культурно-бытового и хозяйственного назначения снизился за этот же период на 2%. Таким образом, 1,08 – общий индекс, а 1,02 – групповой.

В отличии от индивидуальных индексов расчеты общих и групповых индексов сложны.

3. В зависимости от методологии расчета общие и групповые индексы подразделяются на:

а) агрегатные;

б) средние из индивидуальных.

Агрегатные индексы показывают изменение сложного явления и являются основной формой экономических индексов, а средние из индивидуальных индексов – производными, которые получаются в результате преобразования агрегатных индексов.

4. В зависимости от базы сравнения различают:

а) цепные индексы;

б) базисные индексы.

Цепные индексы исчисляются путем сопоставления величины явления в каждом последующем периоде с величиной в предшествующем ему периоде, а **базисные индексы** исчисляются путем сопоставления величины явлений в каждом периоде с величиной в каком-то одном периоде, принятом за базу сравнения (обычно это начальный период).

Введем общепринятые обозначения нескольких показателей.

Показатель	Условное обозначение в базисном периоде	Условное обозначение в отчетном периоде
1.Количество вырабатываемой (потребленной) продукции	q_0	q_1
2.Цена за единицу продукции	p_0	p_1
3.Себестоимость единицы продукции	z_0	z_1
4.Трудоемкость единицы продукции	t_0	t_1
5.Индивидуальный индекс	i	
6.Агрегатный индекс	I	

Приведем расчетные формулы индексов:

$$I_q = \frac{\sum q_1 \cdot p_0}{\sum q_0 \cdot p_0}$$ - агрегатный индекс физического объема.

При определении индекса физического объема продукции в качестве коэффициента соизмерения могут выступать:

Себестоимость, тогда $I_q = \dfrac{\sum q_1 \cdot z_0}{\sum q_0 \cdot z_0}$. Цена, тогда $I_q = \dfrac{\sum q_1 \cdot p_0}{\sum q_0 \cdot p_0}$.

$$I_z = \frac{\sum z_1 \cdot q_1}{\sum z_0 \cdot q_1}$$ - агрегатный индекс себестоимости продукции,

$$I_p = \frac{\sum p_1 q_1}{\sum p_0 q_1}$$ - агрегатный индекс цены ,

$i = \frac{q_1}{q_0}$; $\quad i = \frac{z_1}{z_0}$; $\quad i = \frac{p_1}{p_0}$ - индивидуальные индексы (соответственно, продукции, себестоимости, цены).

$$\left. \begin{array}{l} I_q = \dfrac{\sum q_1 \cdot p_0}{\sum q_0 \cdot p_0} = \dfrac{\sum i \cdot q_0 \cdot p_0}{\sum q_0 \cdot p_0} \\[2ex] I_z = \dfrac{\sum z_1 \cdot q_1}{\sum z_0 \cdot q_1} = \dfrac{\sum \cdot z_1 \cdot q_1}{\sum \dfrac{z_1}{i} \cdot q_1} \end{array} \right\}$$

Преобразованные из агрегатных: средние из индивидуальных индексов (соответственно арифметический и гармонический).

При расчете агрегатных индексов качественных показателей (цены, себестоимости, производительности труда и т.д.) в зависимости от цели исследования можно использовать: индексы качественных показателей переменного состава (индекс средних переменного состава):

$$I_Z^{ПЕР} = \frac{\bar{z}_1}{\bar{z}_0} = \frac{\sum z_1 q_1}{\sum q_1} : \frac{\sum z_0 q_0}{\sum q_0};$$

Индексы качественных показателей постоянного (фиксированного) состава:

$$I_Z^{ФИКС} = \frac{\sum z_1 q_1}{\sum z_0 q_1}$$

Эти два индекса между собой связаны следующим образом:

$$I_Z^{ПЕР} = I_Z^{ФИКС} \cdot K \,,$$

где $I_Z^{ПЕР}$ - индекс себестоимости переменного состава, который показывает изменение средней себестоимости под воздействием двух факторов (индивидуальной себестоимости и удельного веса продукции);

К - индекс структуры, который показывает влияние на изменение средней себестоимости удельного веса продукции каждого предприятия в общем ее объеме;

$I_Z^{ФИКС}$ - индекс себестоимости фиксированного состава, который показывает влияние на изменение средней себестоимости индивидуальной себестоимости продукции каждого предприятия.

Расчет таким способом используется в том случае, если на разных предприятиях производится продукция одного и того же вида.

Расчет индексов постоянного и переменного состава покажем на примере индекса себестоимости продукции.

Пример № 1. Изделие А производится на двух заводах. В базисном периоде на заводе №1 было произведено 500 тыс. штук изделий, себестоимость единицы равнялась 5 у.е. На заводе №2 за этот же период было произведено 200 тыс. штук изделий А при себестоимости 4 у.е. В отчетном периоде объем производства на заводе №1 остался тот же, а себестоимость составила 4,8 у.е. На заводе №2 объем производства увеличился до 500 тыс. изделий, а себестоимость составила 3,9 у.е.

Представим исходные данные в таблице.

Заводы	Производство продукции тыс. штук		Себестоимость продукции, у.е.	
	Базисный период	Отчетный период	Базисный период	Отчетный период
1	500	500	5	4,8
2	200	500	4	3,9
	q_0	q_1	z_0	z_1

Индекс себестоимости изделия А на заводе №1 равен:

$$i_1 = \frac{4.8}{5} = 0.96$$

На заводе №2 - $i_2 = \frac{3.9}{4} = 0.975$

Таким образом, себестоимость единицы продукции в отчетном периоде по сравнению с базисным снизилась на заводе №1 на 4%, а на заводе №2 – на 2,5%.

Для оценки изменения средней себестоимости по двум заводам вместе вычисляется индекс переменного состава:

$$I_{\overline{Z}}^{ПЕР} = \frac{\overline{z}_1}{\overline{z}_0} = \frac{\sum q_1 z_1}{\sum q_1} : \frac{\sum q_0 z_0}{\sum q_0} = \frac{500 \cdot 4.8 + 500 \cdot 3.9}{1000} : \frac{500 \cdot 5 + 200 \cdot 4}{700} = \frac{4.35}{4.714} = 0.923$$

(92,3%)

Итак, средняя себестоимость изделия по двум заводам в отчетном периоде по сравнению с базисным снизилась на 7,7% (100-92,3), т.е. значительно больше, чем по каждому заводу в отдельности. Это объясняется тем, что на динамику средней себестоимости изделия влияют два фактора:

1) снижение себестоимости единицы изделия на каждом заводе в отдельности (индивидуальный индекс себестоимости);

2) изменение удельных весов продукции отдельных заводов в общем объеме производства этого изделия.

В данном случае в отчетном периоде значительно вырос вес продукции завода №2 в общем объеме производства, имеющего более низкую себестоимость по сравнению с заводом №1. Это и привело к дополнительному снижению средней себестоимости по обоим заводам.

Для оценки величины снижении средней себестоимости без учета влияния структурных сдвигов рассчитывается индекс постоянного (фиксированного) состава.

$$I_Z^{ФИКС} = \frac{\sum z_1 \cdot q_1}{\sum z_0 q_1} = \frac{4.8 \cdot 500 + 3.9 \cdot 500}{5 \cdot 500 + 4.0 \cdot 500} = 0.967 \quad (96{,}7\%)$$

Таким образом, среднее снижение себестоимости составило 3,3%, что не превышает индивидуальные индексы себестоимости.

Расчет индекса влияния изменения структуры на динамику средней себестоимости производится путем деления индекса переменного состава на индекс фиксированного состава. В нашем примере он равен:

$$K = \frac{I_{\bar{Z}}^{ПЕР}}{I_Z^{ФИКС}} = \frac{0.923}{0.967} = 0.955 \quad (\ 95{,}5\%\ )$$

Следовательно, изменение структуры привело к дополнительному снижению средней себестоимости на 4,5%.

Возможен и другой способ расчета индекса структурных сдвигов. Для этого необходимо определить структуру выпущенной продукции в обоих периодах (по удельным весам продукции отдельных заводов в общем объеме производства) и эти структурные показатели взвесить по себестоимости базисного периода.

Удельные веса продукции заводов в общем выпуске составляют, %:

В базисном периоде завод №1 - (500/700) 100%=71%;
Завод №2 - (200/700) 100%=29% ;
В отчетном периоде завод №1 - (500/1000) 100%=50%;
Завод №2 - (500/1000) 100%=50%

Тогда индексы структурных сдвигов (К) будет равен:

$$K = \frac{50 \cdot 5 + 50 \cdot 4}{71 \cdot 5 + 29 \cdot 4} = 0.955 \qquad (95,5\%).$$

6.2. Взаимосвязь индексов

Многие экономические индексы тесно связаны между собой и образуют индексные системы. Различают систему индексов во времени и систему индексов по факторам.

Система индексов во времени характеризуется взаимосвязью цепных и базисных индексов. Эта связь проявляется в следующем:

- произведение цепных индексов дает соответствующий базисный индекс;
- отношение последующего базисного индекса к предыдущему базисному индексу дает цепной индекс.

Эта связь характерна для всех индивидуальных индексов независимо от объекта исследования (цена, продукция, себестоимость и т. д.).

Для агрегатных индексов эта связь обнаруживается только для индексов количественных показателей, т. к. в них используются постоянные соизмерители (взятые из базисного периода).

Пример №2. Представлены следующие данные о перевозке пассажиров по воздушной линии:

Года	1-й	2-й	3-й
Количество перевезенных пассажиров, тыс. чел.	320	350	360
Обозначения	q_0	q_1	q_2

Цепные индексы:

$$i_{1/0} = \frac{q_1}{q_0} = \frac{350}{320} = 1.094; \qquad i_{2/1} = \frac{q_2}{q_1} = \frac{360}{350} = 1.028$$

Базисные индексы:

$$i_{1/0} = \frac{q_1}{q_0} = \frac{350}{320} = 1.094; \qquad i_{2/0} = \frac{q_2}{q_0} = \frac{360}{320} = 1.125$$

Произведение цепных индексов:

$$i_{1/0} \cdot i_{2/1} = 1.094 \cdot 1.028 \approx 1.125, \text{ т.е. } i_{1/0} \cdot i_{2/1} = i_{2/0}$$

Соотношение базисных индексов:

$$\frac{i_{2/0}}{i_{1/0}} = \frac{1.125}{1.094} \approx 1.028, \quad \text{т.е.} \quad \frac{i_{2/0}}{i_{1/0}} = i_{2/1}$$

Для агрегатных индексов количественных показателей эта связь выглядит следующим образом:

Цепные индексы:

$$I_{q1/0} = \frac{\sum q_1 p_0}{\sum q_0 p_0}; \qquad I_{q2/1} = \frac{\sum q_2 p_0}{\sum q_1 p_0}$$

Базисные индексы:

$$I_{q1/0} = \frac{\sum q_1 p_0}{\sum q_0 p_0}; \qquad I_{q2/0} = \frac{\sum q_2 p_0}{\sum q_0 p_0}$$

$$\frac{\sum q_1 p_1}{\sum q_0 p_0} \cdot \frac{\sum q_2 p_0}{\sum q_1 p_1} = \frac{\sum q_2 p_0}{\sum q_0 p_0}$$

Пользуясь **системой индексов по факторам**, можно выявить влияние каждого фактора в отдельности на обобщенный показатель.

$$I_{qz} = \frac{\sum q_1 z_1}{\sum q_0 z_0}$$ - индекс затрат.

На индекс общих затрат влияют два фактора: объем выпушенной продукции и ее себестоимость. Влияние первого фактора можно показать, используя индекс физического объема продукции, влияние второго – индекс себестоимости.

$$I_{q\,q} = \frac{\sum q_1 z_0}{\sum q_0 z_0}; \qquad I_z = \frac{\sum z_1 q_1}{\sum z_0 q_1}$$

$$I_{qz} = \frac{\sum q_1 z_0}{\sum q_0 z_0} \cdot \frac{\sum z_1 q_1}{\sum z_0 q_1} = \frac{\sum z_1 q_1}{\sum z_0 q_0} = I_q \cdot I_z$$

Таким образом, индекс общих затрат равен произведению двух индексов: индекса физического объема продукции и индекса себестоимости.

То же можно отнести и к индексу общей стоимости продукции (I_{qp}), на величину которой влияет изменение объема выпущенной продукции и изменение ее цены, т. е.

$$I_{qp} = I_q \cdot I_p = \frac{\sum q_1 p_0}{\sum q_0 p_0} \cdot \frac{\sum p_1 q_1}{\sum p_0 q_1} = \frac{\sum p_1 \cdot q_1}{\sum p_0 q_0}$$

Пример №3. Машиностроительный завод производит поставку трех видов продукции на экспорт. Данные представлены в таблице:

Вид продукции	Объем поставок, тыс. шт.		Цена за ед. прод., у.е.	
	I квартал	II квартал	I квартал	II квартал
A	200	220	100	102
B	410	390	340	325
C	100	120	408	410
	q_0	q_1	p_0	p_1

Определить изменение общей суммы выручки от поставок продукции на экспорт во II квартале по сравнению с I (в абсолютном выражении и в %), в том числе за счет изменения физического объема поставок и изменения цен (в абсолютном выражении и в %).

Изменение общей суммы выручки можно показать с помощью индекса общей стоимости поставки (выручки):

$$I_{qp} = \frac{\sum q_1 p_1}{\sum q_0 p_0} = \frac{220 \cdot 102 + 390 \cdot 325 + 120 \cdot 410}{200 \cdot 100 + 410 \cdot 340 + 100 \cdot 408} = \frac{198390}{200200} = 0.991$$

$$\Delta_{qp} = \sum q_1 p_1 - \sum q_0 p_0 = 198390 - 200200 = -1810 \quad \text{тыс.у.е.}$$

Следовательно, выручка снизилась во II квартале по сравнению с I на 0,9% или на 1810 тыс. у. е.

Изменение выручки за счет изменения физического объема поставок можно определить с помощью индекса количественного:

$$I_q = \frac{\sum q_1\, p_0}{\sum q_0\, p_0} = \frac{220 \times 100 + 390 \times 340 + 120 \times 408}{200200} = \frac{203560}{200200} = 1{,}016$$

$\Delta_q^{qp} = 203560 - 200200 = 3360$ тыс. у. е.

Таким образом, объем поставок увеличился на 1,6%, вследствие чего выручка увеличилась на 3360 тыс. у.е.

Изменение цен повлияло на изменение выручки следующим образам:

$$I_p = \frac{\sum p_1\, q_1}{\sum p_0\, q_{1_1}} = \frac{198390}{203560} = 0{,}975$$

или $I_p = \dfrac{I_{pq}}{I_q} = \dfrac{0{,}991}{0{,}016} = 0{,}975$

$\Delta_p^{qp} = 192390 - 203560 = 5170$ тыс. у. е.

Следовательно, цена снизилась на 2,5%, в результате чего выручка снизилась на 5170 тыс. у. е.

Пример №4. Общие затраты на производство продукции увеличились во II квартале по сравнению с I на 3,8 %, объем продукции снизился на 2,0%. Определить, на сколько % изменилась в среднем себестоимость продукции.

Индекс затрат $I_{qz} = \dfrac{\sum q_1\, z_1}{\sum q_0\, z_0} = 1{,}038$

Индекс продукции $I_q = \dfrac{\sum q_1\, z_0}{\sum q_0\, z_0} = 0{,}98$

Индекс затрат можно представить:

$I_{qz} = I_q\, I_z$, следовательно,

$$I_z = \frac{I_{qz}}{I_q} = \frac{1{,}038}{0{,}98} = 1{,}059$$

Таким образом, себестоимость увеличилась на 5, 9%.

ГЛАВА 7. СТАТИСТИЧЕСКОЕ ИЗУЧЕНИЕ СВЯЗЕЙ МЕЖДУ ПРИЗНАКАМИ

Явления общественной жизни взаимосвязаны и взаимообусловлены. Поэтому анализ объекта управления должен включать характеристику если не всех, то хотя бы основных связей между его разными элементами. При анализе общественных явлений характеристика связей должна основываться на изучении всей совокупности, а не отдельных единиц.

Хотя связи общественных явлений многообразны и сложны, все же в каждом конкретном случае одни признаки влияют на другие и обуславливают их изменение, а другие зависят от них. Зависимые признаки (показатели) называют результативными, а независимые, влияющие на них, - факторными.

По характеру зависимости между факторными и результативными признаками связи делятся на функциональные (полные) и стохастические (вероятные, неполные).

При функциональной связи каждой определенной системе значений факторных признаков соответствуют одно или несколько определенных значений результативного признака. Изменение факторов приводит к строго определенному изменению результативного признака. Так, если даны численность работников и производительность их труда (факторные показатели), то результативный показатель – объем продукции – будет иметь строго определенное значение, равное произведению численности работников и производительности труда. В экономическом анализе функциональные связи изучаются с помощью многих методов. Среди них рассмотренный выше индексный метод.

При стохастической связи каждой определенной системе значений факторных признаков соответствует некоторое множество значений результативного признака. Изменение факторных признаков приводит не к строго определенному изменению результативного признака, как в случае функциональной связи, а к изменению только распределения его значений. Так, известно, что урожайность зависит от количества удобрений. Однако при одной и той же дозе удобрений на разных почвах и в разные годы урожайность будет неодинакова. Разным будет и прирост урожайности при увеличении внесенных удобрений на одну и туже величину. Стохастическая связь, следовательно, является нестрогой, неполной вероятностной. Для анализа стохастической связи применяется ряд методов: аналитические группировки, параллельные ряды, корреляционный, факторный и дисперсионный анализы.

Стохастическая связь называется корреляционной, если при изменении значений факторных признаков меняется средняя величина результативного признака. Уравнение, характеризующее изменение средней величины результативного признака в зависимости от изменений значений фактороного признака (фактроных признаков), называется уравнением корреляционной связи или уравнением регрессии. Корреляционный анализ предназначен для изучения тесноты связи между факторкыми и результативными признаками, а регрессионный анализ – для нахождения уравнений корреляционной связи (регрессий), оценки их точности и надежности.

7.1. Определение наличия корреляционной связи.

При статистическом изучении корреляционной связи между двумя признаками исходными данными являются данные об индивидуальных значениях признаках этих признаков в изучаемой совокупности единиц.

Корреляционные связи бывают прямолинейные и криволинейные. Под прямолинейной корреляционной связью понимают такую связь, при которой с возрастанием одного признака происходит непрерывное возрастание /или убывание/ другого признака в среднем на постоянную величину. Эта связь описывается уравнением прямой. При криволинейной связи между признаками имеется не постоянное, а меняющееся соотношение /результативный признак то увеличивается, то уменьшается с различной степенью интенсивности/. Эта связь описывается уравнением какой-либо кривой.

Применяются несколько способов выявления наличия связи между показателями и ее формы:

1. Метод параллельных рядов /параллельного сопоставления/.
2. Графический способ.
3. Способ группировки и выведения средних по группам.

Сущность способа параллельных рядов или параллельного сопротивления заключается в том, что факториальный признак (x) располагают в порядке возрастания и против каждого его значения записывают соответствующее значение результативного признака ($\overline{y}$). Если с увеличением одного признака другой возрастает (убывает), то между ними имеется связь. Этот анализ может предсказать и форму связи – является ли эта связь прямолинейной или более сложной.

Сущность графического метода заключается в построении поля корреляции, представляющий собой точечный график, для построения которого в прямоугольной системе координат по оси абсцисс откладывают значения факториального признака, а по оси ординат – результативного признака. Получают в поле графика точки, соответствующие этим значениям. По тому, как располагаются эти точки, судят о наличии и форме связи.

Эти два метода наиболее простые, но громоздкие. Их целесообразно использовать в случае, если объем совокупности небольшой.

Более отчетливо корреляционная связь проявляется при использовании метода группировок и расчета средних по группам. Сущность этого метода заключается в том, что совокупность разбивают на группы по факториальному признаку и для каждой группы рассчитывают средние значения результативного признака. Благодаря исчислению этих средних (групповых средних) влияние прочих случайных причин, взаимопогашается и проявляется воздействие именно факториального признака. Средние значения результативного признака ($\bar{y}$) наносят на график, соединив точки, которые им соответствуют, получают эмпирическую линию связи или линию регрессии.

Когда изменение величины факториального признака в определенном направлении вызывает изменение величины результативного признака в том же направлении, то такая корреляционная связь считается прямой (положительной связью). Если же увеличение одного признака обуславливает уменьшение величины другого, находящегося с ним в корреляционной связи, или, наоборот, уменьшение величины одного признака вызывает увеличение другого, то такая корреляционная связь считается обратной (отрицательной). Этот метод является основным для выявления наличия и формы корреляционной связи.

7.2. Определение параметров уравнения связи.

Общий вид корреляционного уравнения прямой линии регрессии, т. е. уравнения прямолинейной корреляционной связи, выглядит следующим образом

$$y = a + bx,$$

Где x и y - индивидуальные значения соответственно факториального и результативного признаков;

a и b - параметры уравнения прямолинейной и корреляционной связи.

Найти теоретическое уравнение связи – значит в данном случае определить параметры прямой.

Эти численные значения / параметры уравнения/ определяются на основе имеющихся данных наблюдения способом наименьших квадратов. Его сущность заключается в следующем. Теоретическая линия регрессии должна изображать изменение средних /рассчитанных как среднее арифметическое/ величин результативного признака «y» по мере изменения величин факториального признака x при условии полного взаимопогашения всех прочих, случайных по отношению к фактору x, причин. Поэтому теоретическая линия регрессии должна обладать основными свойствами средней арифметической, т. е. должна быть проведена так, чтобы сумма отклонений точек эмпирической линии регрессии от соответствующих точек теоретической линии регрессии = 0, а сумма квадратов этих отклонений была бы минимальной величиной.

Если обозначить ординаты фактических точек эмпирической линии регрессии, т. е. индивидуальные значения результативного признака, через «y_i», а ординаты теоретической линии регрессии – через «$\overline{y_x}$», следовательно, эти свойства можно представить как:

$$\sum_i^n (y_i - \overline{y_x})^2 = \min \qquad \begin{cases} \sum (y_i - \overline{y_x}) = 0 \\ \sum (y_i - \overline{y_x})^2 = \min \end{cases}$$

Это условие и лежит в основе способа наименьших квадратов.

Т.к. $\overline{y_x} = a + bx, mo \sum (y_i - a - bx)^2 = \min$.

Рассчитывая первую производную по *«a»* и первую производную по *«b»* от этой функции и приравнивая каждую из производных 0, получим возможность определить те значения *«a»* и *«b»*, при которых $\sum (y - \overline{y})^2 = \min,$, т. е. необходимо решить следующую систему нормальных уравнений

$$\begin{cases} \sum y = an + b\sum x \\ \sum xy = a\sum x + b\sum x^2 \end{cases}$$

Параметр «*b»* играет решающую роль при определении характера связи. Он показывает, насколько возрастет *«y»* при каждом возрастании *«x»* на единицу. Если зависимость положительная прямолинейная, то и параметр *«b»* является положительной величиной, если зависимость обратная, то этот параметр является отрицательной величиной.

Определив параметры *a* и *b* и подставив их значения в уравнения связи, найдем количественную характеристику связи между результативным и факториальным признаками. Подставляя в это уравнение индивидуальные значения факториального признака, определяют средние величины результативного признака для соответствующих значений факториального признака. По полученным величинам средних значений результативного признака составляют теоретическую линию регрессии, характеризующую форму корреляционной связи между изучаемыми признаками.

Криволинейная связь может быть весьма разнообразной. Наиболее часто встречается в экономическом анализе уравнение криволинейной зависимости:

Уравнение гиперболы $y = a + \frac{b}{x}$

Эта форма связи характерна для определения зависимости себестоимости продукции от производительности труда.

Для нахождения параметров уравнения гиперболы способ наименьших квадратов дает следующую систему двух уравнений:

$$\begin{cases} \sum y = an + b\sum \frac{1}{x} \\ \sum \frac{y}{x} = a\sum \frac{1}{x} + b\sum \frac{1}{x^2} \end{cases}$$

Уравнение параболы второго порядка:

$y = ax^2 + bx + c$ - используется при выявлении фактора, влияющего на выполнения плана продукции.

Параметры этого уравнения находят, решая систему трех уравнений:

$$\begin{cases} \sum y = a\sum x^2 + b\sum x + nc \\ \sum xy = a\sum x^3 + b\sum x^2 + c\sum x \\ \sum x^2 y = c\sum x^4 + b\sum x^3 + c\sum x^2 \end{cases}$$

7.3. Показатели степени тесноты связи между признаками

Степень тесноты связи можно измерить следующими показателями:

1) в случае линейной формы связи применяется линейный коэффициент корреляции τ или корреляционное отношение η (см. тему «Показатели вариации»);

2) в случае криволинейной связи рассчитывают лишь корреляционное отношение.

Линейный коэффициент корреляции τ построен на сопоставлении отклонений варьирующих признаков от их среднего значения. Он имеет следующую исходную формулу:

$$\tau = \frac{\sum (x - \bar{x})(y - \bar{y})}{n\sigma_x \sigma_y},$$

где σ_x, σ_y - среднее квадратическое отклонение соответственно по факториальному и результативному признакам.

Преобразование этой формулы приводит ее к виду, наиболее удобному для вычислений:

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

В отличие от корреляционного отношения коэффициент корреляции показывает не только тесноту, но и направление связи. Его значение изменяется от -1 до +1. Если коэффициент корреляции имеет знак плюс, то связь прямая. Близость к единице в том и другом случае характеризует близость к функциональной зависимости.

В случае линейной формы связи факт совпадения или несовпадения η и τ используется для оценки формы связи. Установлено, что если разность между η и τ не превышает 0,1, то гипотезу о прямолинейной форме связи можно считать подтвержденной. Если разность между η и τ больше 0,1, то на различие следует обратить внимание, т. е. связь не является абсолютно прямолинейной.

Оценка существенности коэффициента корреляции определяется на основании критерия его надежности *t*, который рассчитывается по формуле:

$$t = \frac{|\tau|\sqrt{n-1}}{1-\tau^2},$$

где n - объем совокупности.

В математической статистике доказано, что если критерий надежности численно меньше, чем величина 2,56, то связь между коррелируемыми признаками признается несущественной. В этом случае считается, что признак-фактор не оказывает существенного влияния на результативный признак. Если t > 2,56, то связь между признаками считается существенной, а, следовательно, факториальный признак оказывает существенное влияние на результативный признак.

7.4. **Понятие множественной корреляции.**

В отличие от парной, множественной называется корреляция, при которой производится анализ влияния на результативный признак y двух или более признаков – факторов / *x, z, v* и т. д./, т. е.:

$y = f(x, z, v...m)$,

Например, себестоимость перевозок зависит от объема выполненных перевозок, использования основных фондов, величины расходов горючего и т. д.

Задача статистики сводится к выявлению влияния каждого фактора в отдельности и совместного влияния всех факторов на результативный признак.

В исследованиях с применением множественной корреляции обычно применяется линейная прямая зависимость между результативным признаком и признаками - факторами, например, двумя x и z она может быть записана как:

$$y=a+bx+cz.$$

Параметры *a, b, c* как и в случае парной корреляции, находятся по способу наименьших квадратов, т. е. решается система следующих трех уравнений:

$$\begin{cases} \sum y = na + b\sum x + c\sum z \\ \sum xy = a\sum x + b\sum x^2 = c\sum xz \\ \sum zy = a\sum z = b\sum xz = c\sum z^2 \end{cases}$$

Параметры уравнения определяют путем подстановки или при помощи определителя.

Показателем тесноты связи в случае ее линейного характера служит совокупный /общий/ коэффициент корреляции, который определяется по формуле:

$$R_{y(x,z)} = \sqrt{\frac{\tau^2{}_{xy} + \tau^2{}_{zy} \mp 2\tau_{xy}x_{zy}\tau_{xz}}{1-\tau^2{}_{xz}}},$$

где $\tau_{xy,}\tau_{zy},\tau_{xz}$ - линейные коэффициенты парной корреляции между соответствующими признаками.

$$\tau_{xy} = \frac{\sum dxdy}{n\sigma_x\sigma_y};\tau_{zy} = \frac{\sum dzdy}{n\sigma_z\sigma_y};\tau_{zx} = \frac{\sum dxdz}{n\sigma_z\sigma_x}.$$

Так же как и коэффициент парной корреляции τ, совокупный коэффициент корреляции R изменяется от 0 до1. Степень приближения его к единице характеризует приближение корреляционной зависимости y от факторов x и z ,взятых вместе к функциональной.

ГЛАВА 8. ВЫБОРОЧНЫЕ НАБЛЮДЕНИЯ

Выборочным называется такое наблюдение, при котором характеристика всей совокупности единиц дается по некоторой их части, отобранной в случайном порядке. Выборочное наблюдение – наиболее распространенный вид несплошного наблюдения в статистике. Оно дает возможность, не прибегая к сплошному обследованию, получить обобщающие показатели, которые правильно отражают характеристики всей совокупности единиц.

Вся совокупность единиц называется генеральной совокупностью, а та часть совокупности единиц, которая подвергается выборочному обследованию, - выборочной совокупностью. Задача выборочного наблюдения – получить правильное представление о показателях всей генеральной совокупности на основе изучения выборочной совокупности.

8.1. **Доля и средняя**

При выборочном наблюдении дело имеют с двумя категориями обобщающих показателей: с относительными и средними величинами.

Относительные величины применяют для сводной характеристики совокупностей по альтернативному признаку. Под альтернативным понимается такой статистический показатель, который принимает одно из двух взаимоисключающих значений (пол – мужской или женский; изделие – годное или негодное; план по выпуску продукции – выполнен или не выполнен; заказ – выполнен менее чем на 90% или более чем на 90% и т.д.). Как видим, конкретное содержание альтернативного признака, устанавливается самим исследователем. Обычно считают, что если признак x принял интересующее нас значение, то его величина равна 1, в противном случае x =0. В результате в n_1 наблюдениях имеем интересующее нас явление (когда x =1), а когда в n_2 случаях оно отсутствует (когда x =0).

Во всех случаях, когда речь идет о вариации альтернативных признаков, имеют дело с обобщающими показателями в виде относительной доли единиц, составляющих какую-то часть всей совокупности. Этот сводный показатель для генеральной совокупности называется долей или долей в генеральной совокупности, а для выборочной совокупности – выборочной долей.

Следовательно, задача выборочного наблюдения в данном случае – дать на основе выборочной доли правильное представление о доли в генеральной совокупности.

Следовательно, среднее значение альтернативного показателя будет определяться следующим образом:

$$\bar{x}_p = \frac{1 \cdot n_1 + 0 \cdot n_2}{n_1 + n_2} = \frac{n_1}{n} = p \ ,$$

Т.е. среднее значение альтернативного признака равно частоте его появления ($p = \frac{n_1}{n}$).

Аналогично будет определена и дисперсия альтернативного признака:

$$\delta_p^2 = \frac{(1-p)^2 + (0-p)^2 \cdot n_2}{n_1 - n_2} = \frac{(1-p)^2 \cdot n_1 + (0-p)^2 \cdot (n - n_1)}{n} =$$

$$= (1-p)^2 + p^2 \cdot (1-p) = p - 2 \quad p^2 + p^3 + p^2 - p^3 = p - p^2 =$$

$$= p \cdot (1-p)$$

т.е. дисперсия альтернативного признака равна произведению частоты его появления на частоту отсутствия в изучаемой совокупности.

Например, при изучении качества продукции определяют относительную долю тех ее единиц, которые не выдерживают установленного стандарта качества, т.е. относятся к браку. При изучении совокупности студентов нас может интересовать доля в этой совокупности студентов-отличников или студентов, получающих стипендию, и т.д.

Кроме измерения доли перед выборочным наблюдением может стоять задача измерения среднего значения варьирующего признака во всей совокупности. В этом случае имеют дело с признаками, вариация которых проявляется в разных количественных значениях у отдельных единиц совокупности.

Среднее значение варьирующего признака во всей совокупности называется генеральной средней ($\bar{x}$), а среднее значение признака у единиц, которые подверглись выборочному наблюдению – выборочной средней ($\approx x$). Задача заключается в том, чтобы на основе выборочной средней дать правильное представление о средней генеральной.

8.2. **Ошибка выборки**

Выборочные показатели будут отличаться от генеральных в силу двух причин:

1) при любом статистическом наблюдении возникает ошибка точности или регистрации;

При несплошном наблюдении ошибка точности значительно меньше, чем при сплошном (из-за меньшего объема единиц).

2) ошибка репрезентативности – она свойственна только несплошным наблюдениям.

Возникает она только из-за того, что обследованию подвергались не все единицы. Они могут иметь различный знак и абсолютную величину, носят случайный характер и поэтому к ним применимы закон больших чисел и теоремы вероятности.

Выборку можно организовать таким образом, что ошибка может быть минимальной. Она зависит от объема выборки и способа отбора единиц, от вариации признака внутри совокупности.

Пользуясь положениями теории вероятности, можно определить предельную возможную ее величину.

$$\overline{x} = \approx{x} \pm \Delta_x; \qquad \overline{p} = \approx{x} \pm \Delta_p$$

Для расчета ошибки выборки используются теоремы Чебышева и Ляпунова.

Теорема Чебышева: с вероятностью как угодно близкой к единице при достаточно большом числе независимых наблюдений (случай бесповторного наблюдения) можно утверждать, что отклонение выборочной средней от генеральной будет сколь угодно малым.

Она указывает на принципиальную возможность по выборочной средней определить генеральную среднюю.

Эта теорема, дополняясь центральной теоремой Ляпунова, дает возможность определить ошибку.

Теорема Ляпунова: с вероятностью, равной интегралу Лапласа, можно ожидать, что ошибка средней по абсолютной величине не будет больше некоторой величины $t\mu$.

!!!!!!!!

$$P = \left(\left|\overline{x} - \approx{x}\right| \leq t\mu\right) = \phi(t); \quad \phi(t) = \frac{2}{\sqrt{2\pi}}$$

т.е. предельная ошибка выборки (Δ_x) связана со средней ошибкой следующим равенством:

$$\Delta = t\mu,$$

где μ - средняя ошибка выборки;

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

Приведем краткую выдержку из таблицы значений функции Лапласа при разных значениях t:

t	$D(t) = P$ (вероятность)
1,00	0,683
1,10	0,729
1,50	0,866
1,70	0,911
1,90	0,943
1,96	0,950
2,00	0,954
2,50	0,987
2,58	0,990
3,00	0,997

Существуют две формулы для определения средней ошибки выборки, которые действительны для повторной выборки.

Доказательство и вывод этих формул даны в курсе математической статистики.

Когда выборочное обследование ставит своей задачей измерить среднее значение количественно варьирующего признака, то

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

где μ_x - средняя ошибка выборки;

σ_x^2 - дисперсия признака;

n - численность единиц выборочной совокупности.

Для измерения доли альтернативного показателя выборочной совокупности:

$$\mu_p = \sqrt{\frac{\sigma_p^2}{n}} = \sqrt{\frac{p\left(1 - p\right\}}{n}}$$

p - доля признака в выборочной совокупности (например, если % брака в общем объеме продукции = 0,5%, то p = 0,005;

n – численность единиц выборочной совокупности.

Повторную выборку применяют очень редко. Обычно выборку организуют по схеме так называемой бесповторной выборки, при которой та или иная единица совокупности, попавшая в выборку, в дальнейшем уже в выборе не участвует. Таким образом, при бесповторной выборке численность единиц генеральной совокупности уменьшается в процессе выборки.

Для средней – $\mu_x = \sqrt{\frac{\sigma^2}{n}\left(1 - \frac{n}{N}\right)};$

Для доли – $\mu_p = \sqrt{\frac{p\left(1 - p\right)}{n}\left(1 - \frac{n}{N}\right)}$

где N - численность единиц генеральной совокупности.

Т.к. $n < N$, то дополнительный множитель $\left(1 - \frac{n}{N}\right)$ всегда будет <1. Отсюда следует, что величина ошибки выборки при бесповторном отборе всегда будет меньше, чем при повторном отборе.

Пример №1. На заводе с числом рабочих 1000 чел. было проведено 5%-ное выборочное обследование возраста рабочих методом случайного бесповторного отбора. В результате обследования получены следующие данные:

Возраст рабочих, лет	До 30	30-40	40-50	50-60	60 и более
Число рабочих	8	22	10	6	4

С вероятностью 0,997 определить пределы, в которых находится средний возраст рабочих завода.

Решение $\bar{x} = \tilde{\tilde{x}} \pm \Delta_x$

$$\tilde{\tilde{x}} = \frac{\sum x\, f}{\sum f} = \frac{25 \cdot 8 + 35 \cdot 22 + 45 \cdot 10 + 55 \cdot 6 + 65 \cdot 4}{50} \approx 40 \text{ лет.}$$

$$\Delta_x = t \cdot \eta = t \cdot \sqrt{\frac{\sigma_x^2}{n}\left(1 - \frac{n}{N}\right)}$$

$$\sigma_x^2 = \frac{\sum \left(x - \tilde{\tilde{x}}\right) \cdot f}{\sum f} = \frac{(25-40)^2 \cdot 8 + (35-40)^2 \cdot 22 + (45-40)^2 \cdot 10 +}{50}$$

$$+ \frac{(55-40)^2 \cdot 6 + (65-40)^2 \cdot 4}{} = 129 \text{ лет.}$$

$$\Delta_x = 3\sqrt{\frac{129}{50}}(1 - 0{,}05) = 4{,}5 \text{ лет.}$$

$$40\text{-}4{,}5 \le \bar{x} \le 40+4{,}5$$

$$3{,}5 \le \bar{x} \le 44{,}5$$

Пример №2. для определения доли рабочих, не выполняющих норму выработки, была произведена 10%-ная типическая выборка рабочих с отбором числа рабочих пропорционально численности типических групп. Внутри типических групп применяется метод случайного бесповторного отбора. Результаты выборки представлены в следующей таблице:

Цех	Число рабочих в выборке	Доля рабочих, не выполняющих норму выработки, %
Основной	120	5
Вспомогательный	80	2

С вероятностью 0,954 определить пределы, в которых находится доля рабочих, не выполняющих норму выработки.

Решение:

$$\bar{p} = \tilde{\tilde{p}} \pm \Delta_p; \qquad \tilde{\tilde{p}} - \Delta_p \le \bar{p} \le \tilde{\tilde{p}} \pm \Delta_p$$

$$\tilde{\tilde{p}} = \frac{\sum \tilde{\tilde{p}}_i \cdot n_i}{n} = \frac{0{,}05 \cdot 120 + 0{,}02 \cdot 80}{200} = 0{,}038$$

$$\Delta_p = t \cdot \mu_p = t \sqrt{\frac{\tilde{\tilde{p}}\left(1 - \tilde{\tilde{p}}\right)}{n}\left(1 - \frac{n}{N}\right)}$$

$$\Delta_p = 2 \cdot \sqrt{\frac{0{,}038(1 - 0{,}038)}{200} \cdot 0{,}9} = 0{,}025$$

$$0{,}038\text{-}0{,}025 \leq \overline{p} \leq 0{,}038+0{,}025$$

$$1{,}3\% \leq \overline{p} \leq 6{,}3\%$$

Пример №3. На заводе, где работает 10000 рабочих, необходимо установить их средний стаж работы методом случайного бесповторного отбора. Предварительным обследованием установлено, что среднее квадратическое отклонение стажа рабочих равно 5 годам. Определить необходимую численность выборки при условии, что с вероятностью 0,997 ошибка выборки не превысит 1,0 года.

$$n = \frac{\sigma_x^2 \cdot t^2 \cdot N}{N \cdot \Delta_x^2 + \sigma_x^2 \cdot t^2} = \frac{25 \cdot 9 \cdot 10000}{1 \cdot 10000 + 25 \cdot 9} = 220 \text{ чел.}$$

8.3. **Способы проведения выборочного наблюдения**

Выборочная совокупность должна быть образована на основе случайного отбора. Применяют следующие типы отбора:

1. простая случайная выборка или собственно случайная выборка;
2. случайная выборка с механическим отбором единиц;
3. серийная выборка (гнездовая);
4. расслоенная выборка (типическая).

Случайный отбор представляет такую организацию выборочного наблюдения, при которой отбор единиц из генеральной совокупности производится совершенно случайно, наугад или по жребию и при этом обеспечивается равная вероятность каждому элементу генеральной совокупности попасть в выборку.

Простая случайная выборка осуществляется в двух видах: повторная и бесповторная.

В первом случае единица, попавшая в выборку, возвращается в генеральную совокупность, во втором – не возвращается.

Организация выборки может быть осуществлена также и совсем по другому принципу. Вместо того, чтобы отбирать наугад отдельные единицы, попавшие в выборку, можно для отбора единиц установить определенный порядок, по которому производится отбор. В этом случае подлежащие выборочному обследованию единицы отбираются механически в определенной последовательности, например, соблюдается определенное расстояние по порядковому номеру.

Допустим, в отделе кадров какого-либо завода имеется пачка анкет всех рабочих, служащих, ИТР. Поставлена задача установить выборочным путем процент работников, получивших полное среднее образование. Допусти, анкеты строго пронумерованы. Тогда можно условиться, что в выборку попадает каждая десятая анкета. Такой способ носит название механического.

При серийном отборе в случайном порядке произвольно выбираются определенные районы, пункты (гнезда), внутри которых производится сплошное наблюдение.

При типическом отборе вся генеральная совокупность разбивается на группы однородные в качественном отношении, а затем уже внутри каждой такой группы производится случайный отбор, т.е. произвольно. Так, при выборочном отборе анкет работников предприятия все анкеты разбивают на группу рабочих, группу ИТР, группу служащих, а затем уже в пределах каждой такой однородной группы производят отбор наугад, целиком полагаясь на случай. Типический отбор может быть повторным и бесповторным.

Средняя ошибка выборочной средней при бесповторной типической выборке определяется по формуле:

$$\mu = \sqrt{\frac{\overline{\sigma}_i^2 (1 - \frac{n}{N})}{n}};$$

где $\overline{\sigma}_i^2$ - средняя из выборочных дисперсий типических групп,

$$\Delta = t \cdot \mu \cdot$$

Пример №4. Для выявления затрат времени на обработку деталей рабочими разной квалификации на заводе была проведена 10% - ная типическая выборка пропорционально численности групп (внутри групп проведен случайный типический отбор). Результаты обследования представлены в таблице.

Группы рабочих по квалификации	Число рабочих в выборке	Средние затраты времени на обработку одной детали, мин.	Среднее квадратическое отклонение, мин.
I	60	10	1
II	120	14	4
III	80	20	2
IV	40	25	6
	300		

С вероятностью 0,954 определить пределы, в которых находятся средние затраты времени на обработку деталей рабочими всего завода.

Решение.

$$\overline{x} = \tilde{\tilde{x}} \pm \Delta_x$$

$$\tilde{\tilde{x}} - \Delta_x \leq \overline{x} \leq \tilde{\tilde{x}} + \Delta_x$$

$$\Delta_x = t \cdot \mu = t \cdot \sqrt{\frac{\sigma_x^2 (1 - \frac{n}{N})}{n}}$$

$$\overline{\sigma_i^2} = \frac{\sum \sigma_i^2 \cdot n_i}{\sum n_i} = \frac{1^2 \cdot 60 + 4^2 \cdot 120 + 2^2 \cdot 80 + 6^2 \cdot 40}{300} = 12{,}5 \text{ мин.}$$

$$\tilde{\tilde{x}} = \frac{\sum \tilde{\tilde{x}}_i \cdot n_i}{\sum n_i} = \frac{10 \cdot 60 + 14 \cdot 120 + 20 \cdot 80 + 25 \cdot 40}{300} = 16{,}3 \text{ мин.}$$

$$\Delta_x = 2 \cdot \sqrt{\frac{12{,}5(1-0{,}1)}{300}} = 0{,}4 \text{ мин.}$$

$$16{,}3\text{-}0{,}4 \leq \overline{x} \leq 16{,}3+0{,}4$$

$$15{,}9 \leq \overline{x} \leq 16{,}7$$

ГЛАВА 9. РЯДЫ ДИНАМИКИ

Все явления общественной жизни находятся в непрерывном развитии. Изменение общественных явлений во времени статистика изучает с помощью построения и анализа временных рядов, иначе называемых рядами динамики. Каждый ряд состоит из двух граф: в одной указываются периоды (или даты) времени, во второй – числовая характеристика изучаемого явления в эти периоды. Отдельные члены второй графы временного ряда носят название уровней ряда.

9.1. **Классификация рядов динамики**

В зависимости от характера отражаемых показателей различают три вида динамических рядов:

1. динамический ряд, состоящий из абсолютных чисел;
2. динамический ряд, состоящий из средних величин;
3. динамический ряд, состоящий из относительных величин.

В свою очередь, временные ряды, состоящие из абсолютных величин, могут быть двух видов: интервальные и моментные.

В интервальном ряду приводятся данные, характеризующие изучаемые явления за данный период времени.

Примером такого ряда может служить ряд, приведенный в табл. 4.

Таблица 4

Количество перевезенных пассажиров ВТ России

Год	2001	2002	2003	2004	2005
Количество перевезенных пассажиров, млн.чел.	25,07	26,52	29,42	33,78	35,08

Особенностью интервальных рядов динамики является то, что данные этих рядов можно суммировать и получать новые численные значения, относящиеся к более длительным периодам времени. Так, по данным табл. 4 можно получить количество перевезенных пассажиров за 5 лет; он составляет 150,49 млн.чел.

Моментный ряд динамики состоит из показателей, характеризующих состояние (объем) явления на определенные моменты (даты) времени. Примером моментного ряда может служить ряд, приведенный в табл. 5.

Таблица 5

Остатки груза на складе предприятия

Дата	На 1.01.	На 1.02.	На 1.03.	На 1.04.
Остатки груза на складе, тыс.т	7,8	8,6	10,2	12,8

Уровни моментных рядов складывать нельзя. Так как слагающие явление единицы последовательно повторяются в различных уровнях ряда, поэтому их сумма не имеет смысла.

Важнейшим условием правильного построения временных рядов является сопоставимость статистических данных, входящих в состав ряда динамики. Только при соблюдении этого требования может быть обеспечена правильность тех выводов, которые являются результатом анализа ряда. Основным условием сопоставимости статистических показателей является методология их определения.

Другим условием сопоставимости данных является одинаковая полнота охвата различных частей явления, представленного временным рядом. Например. При характеристике динамики объема перевозок груза по годам нельзя использовать в одни годы данные для всего автомобильного транспорта, а в другие – только для автомобильного транспорта общего пользования.

Сравниваемые в рядах динамики уровни должны относиться к периодам одинаковой длительности, к одной и той же территории.

9.2. **Показатели динамических рядов**

Для сравнения между собой отдельных уровней временного ряда рассчитываются следующие показатели: абсолютные приросты, коэффициенты роста, относительные приросты (темпы приростов) и абсолютные значения одного процента прироста.

Абсолютный прирост характеризуется размером увеличения или уменьшения изучаемого явления за определенный период времени. Он определяется как разность между данным уровнем и предыдущим или первоначальным. Уровень, который сравнивается, называется текущим, а уровень, с которым производится сравнение, называется базисным, так как он является базой для сравнения.

Если каждый уровень ряда сравнивается с предыдущим, то получаются цепные показатели. Если же все уровни ряда сравниваются с одним и тем же первоначальным уровнем, то полученные показатели называются базисными.

Абсолютный прирост (Δ) определяется по формулам:

Цепной – $\Delta_i = y_i - y_{i-1}$;

Базисный – $\Delta_i = y_i - y_0$,

где y_i - текущий уровень ряда;

y_{i-1} - уровень, предшествующий текущему;

y_0 - начальный уровень ряда.

Абсолютный прирост может быть положительным, отрицательным и близким к нулю. По величине абсолютного прироста нельзя судить об интенсивности динамики явления, так как величина его зависит от величины уровней, поэтому наряду с абсолютными показателями исчисляют относительные показатели роста и прироста.

Коэффициент роста (К) показывает. Во сколько раз текущий уровень больше (или меньше) уровня базисного периода, т.е. рассчитывается делением текущего уровня на уровень, принятый за базу сравнения. Коэффициенты роста могут быть исчислены с переменной и постоянной базой сравнения (соответственно цепные и базисные коэффициенты) по следующим формулам:

Цепной - $k = \frac{y_i}{y_{i-1}}$;

Базисный - $k = \frac{y_i}{y_0}$.

Они могут быть больше единицы, равны единице и меньше единицы. Коэффициенты роста, выраженные в процентах, носят название темпов роста.

Если цепные коэффициенты роста характеризуют изменение явления от срока к сроку, то базисные – непрерывную линию развития.

Относительный прирост (или прирост в процентах) показывает, на сколько процентов текущий уровень больше (или меньше) уровня базисного периода. Он вычисляется как процентное отношение абсолютного прироста к базисному уровню или как разность между коэффициентом роста, выраженным в процентах, и 100 процентами и носит название темпа прироста.

Формулы расчета темпа прироста (Т) следующие:

при переменной базе:

$$T = \frac{\Delta_i}{y_{i-1}} \times 100 \qquad \text{или} \qquad T = K \times 100 - 100;$$

при постоянной базе:

$$T' = \frac{\Delta_i}{y_0} \times 100 \qquad \text{или} \qquad T' = K' \times 100 - 100.$$

Темп прироста может быть величиной как положительной, так и отрицательной.

Темпы роста и темпы прироста необходимо рассматривать в тесной связи с теми абсолютными величинами, изменение которых они характеризуют, поэтому необходимо исчислять абсолютное значение одного процента прироста.

Абсолютное значение одного процента прироста (А%) определяется делением абсолютного прироста на соответствующий относительный прирост либо путем деления предшествующего уровня на 100, а именно:

$$A = \frac{\Delta}{T} \qquad \text{или} \qquad A = \frac{y_{i-1}}{100}.$$

Для характеристики особенностей развития явления в течение всего рассматриваемого периода рассчитываются средние показатели динамики за единицу времени периода. К их числу относятся: средний уровень ряда, средний абсолютный прирост (средняя скорость роста), средний коэффициент роста, средний темп прироста и средняя величина одного процента прироста.

Средний уровень динамического ряда рассчитывается различно в зависимости от вида ряда.

Для интервального ряда он определяется по формуле простой средней арифметической:

$$\bar{y} = \frac{\sum_{t}^{n} y_i}{n},$$

где n - число уровней ряда.

Для моментного ряда динамики с равными интервалами средний уровень рассчитывается по формуле средней хронологической:

$$\overline{y} = \frac{1/2\, y_0 + y_1 + \ldots + y_{n-1} + 1/2\, y_n}{n-1}.$$

Использование средней хронологической в данном случае может быть объяснено следующим. При расчете средней величины за квартал необходимо использовать средние величины запаса по месяцам, которые могут быть получены как средняя арифметическая из данных на начало и конец месяца. Средняя за квартал представляет собой среднюю из средних месячных и нетрудно заметить, что эти расчеты приведут к формуле средней хронологической.

Средний уровень моментного ряда с неравными интервалами рассчитывается по формуле средней арифметической взвешенной, где в качестве весов выступают промежутки времени (t), в течение которых уровень признака оставался неизменным, т.е.:

$$\overline{y} = \frac{\sum_{t}^{n} y_i t_i}{\sum_{i}^{n} t_i}.$$

Средний абсолютный прирост (средняя скорость роста) определяется по формуле простой средней арифметической из отдельных приростов, рассчитанных с переменной базой сравнения, т.е.

$$\overline{\Delta} = \frac{\sum \Delta_{цепн}}{n-1} = \frac{y_n - y_0}{n-1}.$$

Средний коэффициент роста исчисляется по формуле средней геометрической из отдельных коэффициентов роста, т.е.

$$\overline{k} = \sqrt[n-1]{k_1 k_2 \cdots k_{n-1}},$$

где $k_1, k_2, \dots k_{n-1}$ - коэффициенты роста, рассчитанные с переменной базой сравнения за отдельные периоды времени.

Эта формула может быть преобразована, если значения коэффициентов заменить величинами, из которых они исчислены, т.е.

$$\overline{k} = \sqrt[n-1]{\frac{y_1}{y_0} \cdot \frac{y_2}{y_1} \dots \frac{y_n}{y_{n-1}}} = \sqrt[n-1]{\frac{y_n}{y_0}}$$

Средний темп роста – это средний коэффициент роста, выраженный в процентах, т.е.

$$\overline{T} = \overline{K} \cdot 100 .$$

Как видно, величина среднего темпа, вычисленного по формуле средней геометрической, зависит, по существу, только от соотношения крайних уровней ряда (конечного и начального). Поэтому для того, чтобы среднегодовой темп отразил действительную интенсивность изменения явления во времени, его надо вычислить для промежутка времени с однородным направлением в развитии явления.

Средний темп прироста исчисляется исходя из среднего коэффициента роста, т.е. $\overline{T} = \overline{K} \cdot 100 - 100$.

Средняя величина 1% прироста исчисляется путем деления средней скорости роста на средний темп прироста:

$$\overline{A} = \frac{\overline{\Delta}}{\overline{T}} .$$

9.3. **Сглаживание временных рядов**

При изучении временных рядов большое значение имеет выявление той или иной закономерности в развитии изучаемого явления. Основная закономерность изменения явления во времени часто бывает скрыта незначительными случайными отклонениями в ту или другую сторону.

Для обнаружения общей тенденции изменения ряда часто используют графический метод. В этом случае строят обычную линейную диаграмму. Кроме того, можно использовать методы преобразования рядов. Одним из таких методов является укрупнение интервалов, за которые представлены сведения. Укрупненные интервалы должны быть одинаковой продолжительности.

Другим методом обработки ряда динамики является сглаживание ряда переменной или скользящей средней. Сущность сглаживания заключается в замене абсолютных уровней ряда средними значениями; при этом колебания отдельных уровней ряда динамики взаимопогашаются в средних величинах и закономерность изучаемого явления выступает более отчетливо. Периоды, за которые исчисляют среднюю, могут быть различной продолжительности. Чем больше интервал, за который исчислена средняя, тем больше полученный ряд сглаживается и отличается от фактического (эмпирического) ряда.

При сглаживании по методу переменной средней для получения первого члена выравненного ряда следует складывать первые три уровня (если продолжительность периода принята равной 3) и поделить их сумму на три. Для получения следующего члена ряда складывают последующие три уровня и также находят их среднюю и т.д. Это можно представить следующими расчетами:

$$\overline{y}_1 = \frac{y_0 + y_1 + y_2}{3}; \quad \overline{y}_2 = \frac{y_3 + y_4 + y_5}{3}; \quad \overline{y}_3 = \frac{y_6 + y_7 + y_8}{3} \text{ и т.д.}$$

При сглаживании по методу скользящей средней также рассчитывают средние из уровней ряда. Первая средняя исчисляется так же как и в случае переменной средней. При определении второй средней исключается начальный член ряда и заменяется очередным, т.е. при расчете средней за три периода включается уровень четвертого периода при исключении уровня первого периода.

Расчет при этом может быть представлен таким образом:

$$\overline{y}_1 = \frac{y_0 + y_1 + y_2}{3}; \quad \overline{y}_2 = \frac{y_3 + y_4 + y_5}{3}; \quad \overline{y}_3 = \frac{y_6 + y_7 + y_8}{3} \text{ и т.д.}$$

Более сложным приемом обработки ряда динамики является аналитическое выравнивание, которое может быть проведено прямой или какой-либо другой линии, выраженной в виде функции времени:

$$\overline{y}_t = f(t).$$

Форма выравнивания должна устанавливаться на основе теоретического анализа сущности данного явления и законов его развития.

Если, скажем, теоретический анализ подсказывает, что данное явление развивается с относительно стабильными абсолютными приростами, то для выравнивания подходит прямая; если абсолютные приросты по периодам изменяются (ускоряются или замедляются), то нужно подбирать более сложную кривую (например, параболу 2-го, 3-го и т.д. порядков, гиперболу). Конечно, подбор функции для выравнивания всегда несколько условен, поскольку ни одно общественное явление в своем развитии строго не укладывается ни в какую математическую формулу.

Рассмотрим применение метода аналитического выравнивания по прямой для выражения основной тенденции.

Пример: В таблице приведены исходные и расчетные данные о выплавке чугуна на металлургическом комбинате на 10 дней февраля текущего года.

Таблица 6

Дни Месяца	Выплавка y чугуна, тыс.т	Условные числа февраля, t	y^t	t^2	$\overline{y}_t$	$y-\overline{y}_t$	$(y-\overline{y}_t)^2$
1	33,8	-5	-169.0	25	33,2	0,6	0,36
2	31,8	-4	-127,2	16	33,5	1,7	2,89
3	34,9	-3	-104,7	9	33,8	1,1	1,21
4	33,0	-2	-66,0	4	34,1	-1,1	1,21
5	35,8	-1	-35,8	1	34,5	1,3	1,69
6	35,3	+1	35,3	1	35,1	0,2	0,04
7	35,1	+2	70,2	4	35,4	-0,3	0,09
8	37,1	+3	111,3	9	35,8	1,3	1,69
9	35,7	+4	142,8	16	36,1	-0,4	0,16
10	35,6	+5	178,0	25	36,4	-0,8	0,64
	348,1		34,9	110	348,9		9,98

Для выравнивания ряда динамики по прямой используется уравнение:

$$\overline{y}_t = a_0 + a_1 t \ ,$$

где $\overline{y}_t$ - средние значения выравненного ряда;

a_0, a_1 - параметры уравнения;

t - показатели времени(дни, месяцы, годы и т.д.).

Для нахождения параметров уравнения используется способ наименьших квадратов, который дает систему двух нормальных уравнений:

$$\begin{cases} \sum y = a_0 n + a_1 \sum t \\ \sum yt = a_0 \sum t + a_1 \sum t^2 \end{cases}$$

В рядах динамики техника расчета параметров уравнения может быть упрощена. Для этой цели показателям времени t придают такие значения, чтобы их сумма была равна 0, т.е. $\sum t$ =0.

Если число уровней в ряду динамики нечетное, то центральное значение уровня ряда принимают равным 0.

При этом уравнения системы примут следующий вид:

$$\begin{cases} \sum y = a_0 n \\ \sum yt = a_1 \sum t^2 \end{cases}$$

Отсюда $a_0 = \dfrac{\sum y}{n}$; $a_1 = \dfrac{\sum y\, t}{\sum t^2}$.

$a_0 = \dfrac{348{,}1}{10} = 34{,}81$; $a_1 = \dfrac{34{,}9}{110} = 0{,}32$.

В результате получаем следующее уравнение основной тенденции выплавки чугуна на комбинате:

$$y_t = 34{,}81 + 0{,}32 \cdot t .$$

Подставляя в уравнение принятые обозначения t, вычислим выравненные уровни ряда динамики:

$\overline{y}_1 = 34{,}81 + 0{,}32(-5) = 33{,}2$;

$\overline{y}_2 = 34{,}81 + 0{,}32(-4) = 33{,}5$ и т.д.

Преимущество этого метода заключается в том, что он позволяет найти значение недостающего члена ряда. Такой способ называется интерполяцией рядов динамики, но, как правило, этот способ нет необходимости использовать, т.к. текущая статистика полностью обеспечивается информацией за все периоды времени.

Другой прием, основанный на выравнивании рядов динамики, называется экстраполяцией. Этот прием заключается в том, что пользуясь уравнением, находят значения ряда, лежащие вне его, как бы предсказывая дальнейшее развитие явления.

Например, выплавка чугуна на 11 февраля будет следующей:

$$\overline{y} = 34{,}81 + 0{,}32 \cdot 6 = 36{,}7 \text{ тыс.т.}$$

Но экстраполяцией можно пользоваться в том случае, если не произошло резких изменений в условиях изменений развития явления.

Основная тенденция (тренд) показывает, как воздействуют систематические факторы на уровень ряда динамики, а колеблемость уровней около тренда служит мерой воздействия остаточных факторов. Ее можно найти по формуле среднего квадратического отклонения:

$$\sigma_e = \sqrt{\frac{\sum(y - \overline{y}_t)^2}{n}} = \sqrt{\frac{9{,}98}{10}} \approx 1{,}0 \text{ тыс.т.}$$

Относительной мерой колеблемости является коэффициент вариации, который рассчитывается по формуле:

$$V = \frac{\sigma_t}{\overline{y}}.$$

В нашем примере $\overline{y} = \frac{\Sigma y}{n} = \frac{348{,}1}{10} = 34{,}81$,

Следовательно, $V = \frac{1{,}0}{34{,}81} = 0{,}0287$ или 2,87%.

9.4. **Изучение сезонной неравномерности**

Колебания уровней временного ряда могут иметь либо случайный, либо систематический (закономерный) характер. Случайные колебания не имеют регулярного характера и обычно оказываются сравнительно небольшими.

Колебания, имеющие систематический, закономерный характер, повторяются через приблизительно одинаковые промежутки времени и вызываются определенными причинами. К последним относятся так называемые сезонные колебания (сезонная неравномерность).

Сезонная неравномерность – это сравнительно устойчивые внутренние колебания, т.е. когда из года в год в одни месяцы уровень явления повышается, а в другие – снижается. Она обуславливается специфическими условиями, влиянием многочисленных факторов. Сезонные колебания отрицательно сказываются на экономических показателях работы предприятий, использовании основных фондов и трудовых ресурсов. Анализ сезонных колебаний необходим для улучшения работы предприятий.

Перед статистикой стоит задача выявить сезонную неравномерность и измерить ее величину. Наличие сезонной неравномерности выявляется, прежде всего, с помощью графического метода. Применяют в этом случае линейные диаграммы, на которые данные об объеме явления по месяцам наносят не менее, чем за три года. Чтобы исключить влияние различной продолжительности отдельных месяцев, предварительно исчисляют средние суточные объемы за каждый месяц.

Для измерения сезонной неравномерности (сезонной волны) используются индексы сезонности, которые рассчитываются двумя методами в зависимости от характера динамики: методом простых средних, методом помесячных отношений.

Метод простых средних применяется в тех случаях, когда годовой уровень сезонного явления остается из года в год относительно неизменным. Сущность метода состоит в том, что определяются процентные отношения среднего уровня каждого месяца к средней месячной за весь исследуемый период.

В табл. 7 представлены данные об объеме грузовых перевозок предприятия по месяцам за три года, а также данные о среднесуточном объеме.

Таблица 7

Перевозка грузов

Месяцы	Объем грузовых перевозок, т			Среднесуточный объем перевозок, т				Индексы сезонности, %
	2003	2004	2005	2003	2004	2005	Ср.сут.	
1	2	3	4	5	6	7	8	9
1	250	320	360	8,1	10,3	11,6	10,0	79,4
2	240	280	400	8,6	10,0	14,3	11,0	87,3
3	280	300	410	9,0	9,6	13,3	10,6	84,1
4	300	340	440	10,0	11,3	14,6	12,0	95,2
5	318	370	480	10,3	11,9	15,5	12,6	100,0

6	306	480	512	10,2	16,0	17,1	13,4	106,1
7	312	495	508	10,1	16,0	16,4	14,2	112,7
8	327	500	511	10,5	16,2	16,5	14,4	114,3
9	320	470	495	10,6	15,7	16,5	14,3	113,5
10	300	472	498	9,7	15,2	16,0	13,7	108,7
11	294	460	450	9,8	15,3	15,0	13,4	106,3
12	260	400	430	8,4	12,9	13,9	11,7	92,7

Средний суточный объем перевозок исчисляется путем деления общего объема перевозок за месяц на число календарных дней в месяце. Так, для января 2003 года средний суточный объем перевозок составил 8,1 т (250/31), а для февраля этого же года – 8,6 т (240/28).

Расчет индексов сезонности осуществляется в следующей последовательности:

Сначала исчисляются средние для каждого месяца по данным за три года, что дает возможность избавиться от случайных колебаний месячных уровней: эти средние представлены в гр. 8 табл. 7.

Они исчислялись следующим образом:

Январь - $\overline{y}_1 = \frac{8,1 + 10,3 = 11,6}{3} = 10$;

Февраль - $\overline{y}_2 = \frac{8,6 + 10,0 + 14,3}{3} = 11$ и т.д.

Затем рассчитывается средняя арифметическая за весь исследуемый период:

y_0 = (10,0 x 31 + 11,0 x 28 + 10,6 x 31 + 12 x 30 + 12,6 x 31 + 13,4 x 30 + 14,2 x 31 + 14,4 x 31 + 14,3 x 30 + 13,7 x 31 + 13,4 x 30 + 11,7 x 31) : 365 = 12,6 т.

На последнем этапе исчисляются индексы сезонности (J_s) как отношение средних для каждого месяца к среднему уровню за весь период, т.е.

Январь - $J_{s_1} = \frac{10}{12,6} \times 100\% = 79,4\%$;

Февраль -

$$J_{s_2} = \frac{11}{12,6} \times 100\% = 87,3\%$$ и т.д.

Индексы сезонности проставлены в графе 9.

Индексы сезонности показывают, что среднесуточный объем перевозок в январе меньше среднесуточного годового объема на 20,6% (79,4% - 100), а в августе превышает его на 14,3% (114,3 – 100).

Помимо индексов сезонности можно рассчитать коэффициент сезонной неравномерности, который определяется как отношение максимальной (пиковой) величины объема перевозок к средней за весь рассматриваемый период, т.е.

$$K_{сез.нер.} = \frac{\overline{y}_{\max}}{\overline{y}_0} = \frac{14,4}{12,6} = 1,143 .$$

Метод помесячных отношений применяется в том случае, когда уровни сезонного явления имеют тенденцию к росту или снижению. Он заключается в том, что вначале вычисляются по каждому году процентные отношения уровня каждого месяца к уровню предшествующего месяца (цепные коэффициенты роста в процентах), а затем из полученных отношений исчисляется средняя арифметическая (средние помесячные отношения).

9.5. **Приведение рядов динамики к общему (единому) основанию**

При изучении рядов динамики возникает необходимость получения сравнительных характеристик направления и интенсивности роста одновременно развивающихся во времени явлений. Это достигается путем приведения рядов динамики к общему единому основанию.

По исходным уровням нескольких рядов динамики определяют относительные величины – базисные коэффициенты роста. Принятый при этом за базу сравнения период времени (обычно начальный) выступает в качестве постоянной базы расчета коэффициентов роста для каждого из изучаемых рядов динамики.

Пример. Объем перевозок пассажиров ВТ России по международным и внутренним линиям характеризуется следующими данными (млн.чел.):

Годы	2001	2002	2003	2004	2005
Перевозки пассажиров	25,07	26,52	29,42	33,78	35,08
В том числе Международные	10,04	11,08	12,31	14,90	15,88
Внутренние	15,03	15,44	17,11	18,88	19,20

Различные значения абсолютных уровней приведенных выше рядов динамики затрудняют выявление особенностей в развитии международных и внутренних перевозок пассажиров. Приведем абсолютные уровни рядов к общему (единому) основанию, приняв за постоянную базу сравнения уровни 2001г. В результате получают данные в процентах к 2001г.

Перевозки Пассажиров	2001	2002	2003	2004	2005
Международные	100	1,104	1,226	1,484	1,582
Внутренние	100	1.027	1,138	1,256	1,277

Из этих данных видно, что международные перевозки непрерывно и быстро возрастают, значительно превосходя темпы роста внутренних перевозок. Если в 2005 г. Международные перевозки возросли по сравнению с 2001 г. Более чем в 1,5 раза, то внутренние за это время возросли почти в 1,3 раза.

Сопоставив базисные темпы роста международных и внутренних перевозок пассажиров на воздушном транспорте России получим показатель, который носит название коэффициент опережении ($K_{опер.}$):

$$K_{опер.} = \frac{1,582}{1,277} = 1,24$$

т.е. международные перевозки за 2001-2005 гг. развивались в 1,24 раза быстрее, чем внутренние.