ВВЕДЕНИЕ

В современном обществе статистика является необходимым и эффективным инструментом государственного управления экономикой. Статистика − многоотраслевая наука. Общая теория статистики является наукой о наиболее общих принципах и методах статистического исследования общественных явлений. Крупными, относительно самостоятельными частями являются экономическая и социальная статистики, исследующие явления, которые характерны соответственно для хозяйственной (экономической) и социальной жизни общества.

Составными частями статистической науки являются отраслевые статистики: промышленная, сельскохозяйственная, транспортная и др.

Статистика железнодорожного транспорта − это отраслевая статистика. Она изучает железнодорожный транспорт как самостоятельную отрасль народного хозяйства, учитывая специфику его деятельности – осуществление процесса перевозок грузов и пассажиров, используя единые, выработанные статистикой принципы и методологию. К их числу относятся правила и методы представления статистической информации в виде таблиц и графиков, построения статистических рядов; определение сущности, правил применения и способов расчета обобщающих (сводных) показателей: абсолютных и относительных величин, средних показателей динамики; выявление взаимосвязей изучаемых явлений. Все названные обобщающие показатели и методы используются при любом статистическом исследовании, включающем три взаимосвязанные стадии: статистическое наблюдение, сводку и анализ исследуемых процессов на основе полученной информации. Прикладные статистические исследования в рамках конкретной отрасли осуществляются в той же последовательности и с применением тех же методов. Поэтому в данном пособии и рассмотрены эти вопросы в соответствующих разделах: основы теории обобщающих показателей и статистической методологии, основы железнодорожной статистики, анализ статистической информации о работе железной дороги.

1 ОСНОВЫ ТЕОРИИ ОБОБЩАЮЩИХ ПОКАЗАТЕЛЕЙ И СТАТИСТИЧЕСКОЙ МЕТОДОЛОГИИ

1.1 Предмет, метод и общие понятия статистической науки

Термин «статистика» происходит от латинского слова status, что в средние века означало политическое состояние государства.

Статистика − наука, изучающая положение дел в государстве. Это вид деятельности, направленный на получение, обработку и анализ информации, характеризующей количественные закономерности жизни общества во всем их многообразии (технико-экономические, социально-политические явления, культура) в неразрывной связи с ее качественным содержанием. Под статистикой понимается также сбор цифровых данных, их обобщение и обработка. Статистика необходима для изучения явлений общественной жизни посредством цифр и должна давать не произвольный материал, а те данные, которые ясно и понятно характеризуют различные явления.

Предмет статистики. Статистика изучает массовые общественные явления и их динамику при помощи статистических показателей, в постоянной связи с их содержанием, а также количественное выражение качественно определённых закономерностей общественного развития в конкретных условиях места и времени. Для изучения предмета статистики разработаны и применяются специфические приёмы, совокупность которых образует *методологию статистики.*

Статистика оперирует определёнными *категориями*, т.е. понятиями, отражающими существенные, всеобщие свойства и основные отношения явлений действительности. К таким понятиям относятся: статистическая совокупность, единица совокупности, признак, статистический показатель, система статистических показателей, статистическая закономерность.

Статистика изучает закономерности массовых явлений. *Объект* конкретного статистического исследования называют статистической совокупностью.

Статистическая совокупность − это совокупность объектов или явлений общественной жизни, объединенных общей связью. Объекты, входящие в статистическую совокупность, обладают несколькими общими признаками и могут различаться между собой рядом других, второстепенных, признаков. Специфическим свойством статистической совокупности является *массовость* единиц, поскольку явление характеризуется массовым процессом и всем многообразием определяющих его причин и форм.

Совокупности могут быть разнородными и однородными. Совокупность объектов, у которых один или несколько изучаемых существенных признаков являются общим, называется качественно *однородной*. И, наоборот, совокупность, в которую входят разные типы явлений, будет *разнородной*. Выделение качественно однородных статистических совокупностей является предпосылкой расчёта обобщающих показателей, статистического изучения вариации, связей между признаками.

Единица совокупности − это первичный элемент статистической совокупности, являющийся носителем признаков, подлежащих регистрации, и основой ведущегося при обследовании счёта.

Единицы статистической совокупности характеризуются общими свойствами, именуемыми в статистике признаками. *Признак* − показатель, характеризующий свойства, характерные черты или особенности объектов (явлений), которые могут быть охарактеризованы рядом статистических величин. Система признаков используется для составления программы статистического наблюдения и последующей группировки материалов. Так, например, для промышленных предприятий признаками будут: форма собственности, род (или вид) выпускаемой продукции, размеры производства и др.

Понятия признака и статистического показателя взаимосвязаны. *Показатель* выражает единство количественной стороны явления (его меру), признак — отличительные особенности или сходство объектов статистической совокупности.

Признаки могут иметь количественное выражение или не иметь такового. Признаки, систематически принимающие различные значения у отдельных единиц совокупности, называются *варьирующими признаками* (размеры одежды и обуви, нормы выработки при однородных условиях и т.д.). Варьирующие признаки могут быть *количественными*, если их варианты выражаются числовыми значениями (возраст, стаж работы, оплата труда), и *атрибутивными*, представляющими собой смысловые понятия (профессия, социальная принадлежность).

Количественные признаки могут быть *дискретными* (принимающими обычно целые значения) и *непрерывными* (интервальными). В случае, когда варианты признака могут принимать одно из двух противоположных значений, говорят об *альтернативном признаке* (да, нет).

Признаки могут быть *основными*, определяющими, например, экономическое содержание процессов, и *второстепенными*, внешними по отношению к сущности изучаемых явлений, т.е. непосредственно не связанными с внутренней структурой процессов. Признаки бывают первичными и вторичными. *Первичные признаки* лежат в основе программы сбора первичных статистических материалов. *Вторичные признаки* (производные показатели) −

это признаки, получаемые в процессе обработки и анализа данных. Так, группировка предприятий по эффективности использования инвестиций основана на вторичных признаках, ибо для определения показателей эффективности необходимо знать первичные признаки предприятий (компаний, фирм и т. д.) − размер инвестиций, уровень себестоимости и др.

Признаки, характеризующие статистическую совокупность, взаимосвязаны между собой, поэтому следует различать факторные и результативные признаки. *Факторные признаки* – это независимые признаки, оказывающие влияние на другие, связанные с ними признаки. *Результативные признаки* – это зависимые признаки, которые изменяются под влиянием факторных признаков. Например, квалификация, стаж работы – факторные признаки, производительность труда – результативный.

Статистическая совокупность состоит из массы отдельных единиц, разрозненных фактов. Задача статистики – установить общие свойства единиц совокупности, изучить имеющиеся взаимосвязи и закономерности развития. Достигается это с помощью расчёта обобщающих статистических показателей и их анализа.

Статистический показатель − обобщенная количественная характеристика явлений и процессов в единстве с их качественной определенностью. Примером статистического показателя служат: численность населения; количество индивидуальных частных предприятий в общем; доля работающих граждан в общей численности населения; доля менеджеров, имеющих специальное управленческое образование, от их общего количества и т.п. Величина статистического показателя определяется в результате измерения объектов (элементов) и меняется в зависимости от методологических особенностей его построения, обусловленных, в свою очередь, степенью охвата изучаемых процессов.

Статистические показатели называются *натуральными,* когда они выражены в единицах счета или различных физических единицах измерения, и *денежными*, или стоимостными, когда они представляют денежную оценку экономических объектов или их элементов.

Статистические показатели также условно делятся на *объемные* и *качественные*. К первым относятся показатели, связанные с изменением величины совокупности объектов (элементов), например, численности рабочих на предприятии, компании или фирме, объем основных средств. К группе качественных статистических признаков относят признаки уровня развития явлений, например себестоимости единицы изделия, перевозок. Такие признаки полнее и ярче характеризуют качественные особенности явлений, закономерности их развития.

Поскольку отдельные свойства совокупности не изолированы, а связаны между собой, то и статистические показатели, характеризующие эти свойства, не являются разрозненными, а образуют систему показателей.

Система статистических показателей − это совокупность взаимосвязанных между собой статистических показателей, всесторонне отображающих процессы общественной жизни в определенных условиях места и времени. Показатели в системе могут быть связаны как жестко *детерминированной связью* (например, связь основных средств, числа работников и объёма продукции предприятия), так и не жёсткой, свободной, т.е. *стохастической связью* (например, зависимость себестоимости перевозок от качественных показателей работы подвижного состава).

Задача статистики: используя адекватную систему показателей, дать обобщающую характеристику объёма и состава совокупности, а также выявить и изучить статистические закономерности.

Закономерности, выявленные для той или иной совокупности, обнаруживаются при массовых наблюдениях благодаря действию закона больших чисел. *Закон больших чисел* – это объективный закон, согласно которому совместное действие большого числа случайных факторов приводит к результату, почти не зависящему от случая.

Случайное событие – событие, которое при заданной совокупности условий может произойти, а может и не произойти, но для которого определена вероятность его осуществления. *Случайность является формой проявления необходимости.*

Важнейшей категорией статистики является статистическая закономерность. Под закономерностью вообще принято понимать повторяемость, последовательность и порядок изменений в явлениях. *Статистическая закономерность* – количественная закономерность изменения в пространстве и во времени массовых явлений и процессов общественной жизни, состоящих из множества элементов (единиц совокупности). Она проявляется не в индивидуальном явлении, а в массе однородных явлений, при обобщении данных статистической совокупности, т.е. в среднем. Следовательно, это средняя закономерность массовых явлений и процессов.

Статистическая закономерность отражает относящиеся к определённому пространству и времени причинно-следственные связи, выражающиеся в последовательности, регулярности, повторяемости событий с достаточно высокой степенью вероятности. Статистическая закономерность устанавливается на основе анализа массовых данных, это обусловливает её взаимосвязь с законом больших чисел.

Метод статистики. Основной чертой статистической методологии является конкретность исследований, выражающаяся в неразрывной связи количественного анализа с установлением качественного своеобразия объектов в конкретно-исторических условиях места и времени. Применение в статистике конкретных методов предопределяется конкретными задачами и зависит от характера исходной информации.

Общей основой разработки и применения статистической методологии является диалектический метод познания, согласно которому общественные явления и процессы рассматриваются в развитии, взаимной связи и причинной обусловленности. Знание законов общественного развития создаёт фундамент, с помощью которого можно понять и правильно истолковать явления, подлежащие статистическому исследованию, выбрать надлежащую методику их изучения и анализа. При этом статистика опирается на такие диалектические категории, как *количество* и *качество*, *необходимость* и *случайность*, *причинность*, *закономерность*, *единичное* и *массовое*, *индивидуальное* и *общее*.

Статистические методы − это совокупность приемов, применяемых в процессе статистического исследования.

Статистическое исследование − процесс изучения явлений на основе статистических методов. Статистические исследования начинаются с подготовительных работ по организации исследований. Они делятся на взаимосвязанные и в большей мере самостоятельные этапы, как правило, обособленные друг от друга во времени, которые называются *стадиями*. Обычно выделяют *три основные стадии*: статистическое наблюдение, сводка и обработка материалов, анализ данных. *На первой стадии* с помощью первичного учета, систематической регистрации и других специальных форм статистического наблюдения собираются массовые статистические данные; *на второй стадии* эти данные сводятся в систему сводных таблиц с применением системы группировок и сводных величин (обобщающих показателей); *на третьей стадии* собранные данные анализируются, т.е. проводится сравнение фактов для разных периодов времени, для различных объектов, устанавливаются причины явлений, дается общее описание фактов и объяснение закономерностям, выделяемым с помощью статистических методов. *Статистический анализ* − завершающее звено статистического исследования, он имеет большое познавательное и практическое значение. Статистический анализ изучает статистические данные о явлении для выяснения его характерных признаков и присущих ему в данных конкретных условиях закономерностей.

Статистические методы разделяются на две основные группы: методы статистического наблюдения и методы обработки и анализа статистических данных.

1.2 Статистическое наблюдение

1.2.1 Понятие о статистическом наблюдении

Для проведения статистического исследования необходимо обладать статистической информацией. **Статистическая информация** (статистические данные) – первичный статистический материал о социально-экономических явлениях, формирующийся в процессе статистического наблюдения, который затем подвергается систематизации, сводке, анализу и обобщению.

Статистическое наблюдение − это научно организованный сбор количественных данных о явлениях и процессах, происходящих в различных областях деятельности, с помощью учета первичных данных о каждом отдельном случае или факте, относящемся к изучаемому явлению.

При проведении статистического наблюдения необходимо придерживаться следующих положений:

− статистическое наблюдение должно проводиться по тщательно разработанной программе;

− наблюдению должны подвергаться, прежде всего, те явления и процессы, благодаря которым осуществляется успешная коммерческая деятельность и решаются социальные проблемы;

− наблюдение должно проводиться по программе, соответствующей целям и задачам наблюдения, со строгим ограничением объекта и единицы наблюдения;

− наблюдение должно проводиться на научной основе и методами, обеспечивающими доступность, полноту и объективность получаемых сведений;

− система (форма), виды и способ наблюдения должны выбираться в соответствии с экономической сущностью изучаемого явления или процесса и отвечать конечной цели исследования;

− наблюдение должно обеспечивать сопоставимость регистрируемых данных с прогнозируемыми показателями и сопоставимость данных с предшествующими исследованиями.

Статистическое наблюдение − первый этап любой статистической работы, в результате проведения которого получают достоверные, сопоставимые друг с другом исходные цифровые данные, позволяющие обеспечить научно обоснованные выводы о характере и закономерностях развития изучаемого явления.

Следующим этапом статистической работы является сводка, группировка собранных данных в пределах каждой группы и по совокупности в целом. Обработка статистического материала производится путем построения рядов цифр, таблиц, графиков. Затем переходят к вычислению обобщающих показателей, которыми заполняют таблицы: относительных величин, средних величин, индексов, показателей вариации и т. д.

1.2.2 Формы статистического наблюдения

Статистическое наблюдение осуществляется в двух формах: путём предоставления отчётности и проведения специально организованных статистических наблюдений.

Отчётностью называют такую организованную форму статистического наблюдения, при которой сведения поступают в виде обязательных отчётов в определённые сроки и по утверждённым формам. Отчетность дает возможность получать исчерпывающие данные о деятельности предприятий, организаций, учреждений.

Отчетность подразделяется на срочную, текущую и годовую.

Срочная отчетность действует в тех случаях, когда необходимо получить информацию по важнейшим показателям хозяйственной деятельности. В случае коммерческого торгового предприятия это: объем реализации, товарные запасы, издержки обращения и т. д. Срочная отчетность характеризуется короткой периодичностью ее предоставления: пятидневная, декадная, пятнадцатидневная.

Текущая отчетность более подробная, т.е. имеет более развернутый круг показателей. Она предоставляется за месяц, квартал. Месячная отчетность дополняет квартальную и позволяет анализировать основные показатели хозяйственной деятельности не только в целом за квартал, но и по отдельным месяцам.

Годовая отчетность дает полную, подробную, законченную характеристику состояния деятельности предприятий, фирм, компаний различных сфер деятельности и форм собственности, организаций и учреждений. Годовая отчетность уточняет данные месячной и квартальной отчетности, а круг ее показателей позволяет дать углубленный (детализированный) анализ работы различных коммерческих и некоммерческих предприятий, учреждений, организаций.

Второй формой статистического наблюдения являются специально организованные статистические наблюдения.

Специально организованные статистические наблюдения проводятся в тех случаях, когда необходимо получить сведения по показателям, не предусмотренным статистической отчетностью. Например, подробная характеристика структуры потребления по отдельным категориям населения. Посредством специально организованного наблюдения получают дополнительную информацию для уточнения данных статистической отчетности. К числу специально организованных статистических обследований относятся разного рода переписи и учеты.

1.2.3 Виды статистического наблюдения

Статистическое наблюдение подразделяется на в и д ы: по *времени регистрации данных* и по *степени охвата единиц* наблюдения.

По **времени регистрации фактов** различают текущее наблюдение, периодическое (регистрация по мере надобности) и единовременное. *Текущее наблюдение* ведется систематически, непрерывно, по мере возникновения явлений. Например, регистрация рождаемости и смертности осуществляется загсами, учет выпуска продукции, явки и неявки на работу − предприятиями, компаниями, фирмами и т. д.

Единовременное наблюдение проводится один раз для решения какой-либо задачи или повторяется эпизодически через неопределенный промежуток времени, по мере надобности. Примером может быть перепись жилого фонда.

По **степени охвата единиц** совокупности различают сплошное и несплошное наблюдения. *При сплошном наблюдении* регистрируются все без исключения единицы совокупности. Оно применяется при переписи населения, отчетности, охватывающей все государственные и негосударственные предприятия, фирмы, компании, учреждения, организации и т.д. В статистике сплошное наблюдение является одним из основных источников получения необходимых данных.

Одновременно в современной статистике, в условиях рыночной экономики, используется в широких масштабах *несплошное наблюдение*. Несплошным называется такое наблюдение, при котором обследованию подвергаются не все единицы изучаемой совокупности, а только их часть, на основе которой можно получить обобщающую характеристику всей совокупности. Несплошное наблюдение имеет ряд преимуществ перед сплошным: за счёт уменьшения числа обследуемых единиц совокупности оно требует значительно меньше материальных и трудовых затрат, может быть проведено в более короткие сроки, позволяет применять более совершенные способы учета фактов, что повышает оперативное значение статистических материалов.

Несплошное наблюдение подразделяется на способ основного массива, монографическое и выборочное.

Способ основного массива состоит в том, что обследованию подвергается та часть единиц совокупности, у которой величина изучаемого признака является преобладающей во всём объёме. Часть совокупности, о которой заведомо известно, что она не играет большой роли в характеристике совокупности, исключается из наблюдения.

Монографическое наблюдение. Оно заключается в подробном описании и исследовании небольшого числа отдельных, характерных в каком-либо отношении единиц совокупности для их углублённого изучения. Монографическое наблюдение широко используется научными учреждениями для глубокого и всестороннего изучения существенных особенностей объектов. Так, особый интерес представляет монографическое исследование какой-либо эффективно работающей фирмы с целью изучения способов и методов ее коммерческой деятельности.

При *выборочном наблюдении* обследованию подвергается отобранная в определенном порядке часть единиц совокупности, а полученные результаты распространяются на всю совокупность. Таким образом, в основе выборочного наблюдения лежит случайный отбор некоторой части единиц изучаемой совокупности и распространение полученных в результате наблюдения сводных характеристик (средних и относительных величин) на всю совокупность.

В любом статистическом исследовании для получения первичных данных могут быть использованы непосредственные наблюдения, документальный учёт, опрос.

Непосредственным является такое наблюдение, при котором сами регистраторы путём замера, взвешивания или подсчёта устанавливают факт, подлежащий регистрации, и на этом основании производят записи в формуляре наблюдения. Так, при учёте остатков товаров в торговле за основу берётся их инвентаризация, при переписи оборудования сведения заносятся в формуляр на основе личного осмотра машин.

При *документальном* учёте фактов источником сведений служат соответствующие документы. Этот способ наблюдения используется при составлении предприятиями и учреждениями отчётности на основе документов первичного учёта и обеспечивает большую точность сведений.

Опрос – это наблюдение, при котором ответы на изучаемые вопросы записываются со слов опрашиваемого, например, перепись населения.

Опрос может быть по-разному организован. В статистике применяются следующие основные способы опроса: экспедиционный (устный) опрос, анкетный, саморегистрации и корреспондентский способ.

Экспедиционный способ заключается в том, что специально выделенное лицо − регистратор − опрашивает обследуемое лицо и с его слов заполняет бланк обследования. При этом он одновременно контролирует правильность получаемых сведений. Этот способ обеспечивает достаточно точные результаты, но он весьма дорогостоящий. По этой причине его применяют при наиболее важных статистических обследованиях населения (например, переписи населения).

Анкетный способ состоит в том, что разработанная анкета рассылается определенному кругу лиц и после заполнения возвращается статистическим органам. Таким образом, данный способ основан на принципе добровольного заполнения специальных опросных бланков (анкет), рассылаемых лицам, от которых желательно получить сведения, с просьбой их заполнить и прислать обратно.

При способе *саморегистрации* соответствующие документы заполняют сами опрашиваемые. Обязанность счётчиков (регистраторов) состоит в раздаче бланков наблюдения опрашиваемым, их инструктаже и затем в сборе и проверке заполненных формуляров.

Корреспондентский способ заключается в том, что статистические и другие органы рассылают специально разработанные бланки и инструкции к их заполнению отдельным организациям или специально подобранным лицам, давшим согласие периодически заполнять их и присылать статистическому или другому органу в установленные сроки.

1.2.4 Программно-методологическое обеспечение статистического наблюдения

К **программно-методологическим вопросам статистического наблюдения** относятся:

− установление цели и задач наблюдения;
− определение объекта и единицы наблюдения;
− разработка программы наблюдения;
− выбор вида и способа наблюдения.

Цель статистического наблюдения определяется исходя из общих задач, поставленных перед статистическим изучением явлений. *Задача* наблюдения непосредственно вытекает из задач статистического исследования и предопределяет его программу и формы организации. В зависимости от цели выбирается объект статистического наблюдения.

Объект статистического наблюдения − это определенное явление, которое подлежит наблюдению. Следует определить, что входит в состав объекта, а что не входит. Установить объект наблюдения − значит точно определить состав и границы совокупности. Например, объектом переписи населения является совокупность всех живущих в данной стране лиц, объектом наблюдения при изучении промышленности по производству безалкогольных напитков − совокупность фирм, компаний, предприятий и т.д., производящих безалкогольные напитки.

Определение объектов наблюдения представляет собой сложную и ответственную задачу, потому что различные явления тесно связаны между собой и взаимно переплетаются. Недостаточно указать объект исследования, нужно дать ему четкое научное определение, которое позволило бы отграничить данный объект от смежных с ним. Определение объекта наблюдения должно содержать точные указания на его главные признаки и свойства. Например, мало сказать, что объектом наблюдения являются сельскохозяйственные предприятия и хозяйства, необходимо четко определить, к каким формам собственности они относятся.

Следовательно, совокупность вопросов, которые необходимо выяснить в объекте наблюдения, должна быть точно определена, чтобы результаты наблюдения отвечали поставленной цели.

Единица наблюдения − единица, о которой записываются данные, составляющие программы статистического изучения. В каждом конкретном статистическом исследовании объектов наблюдения, а также в зависимости от тех задач, которые нужно разрешить в процессе наблюдения, определяется, сколько единиц наблюдения должно быть обследовано (одна или несколько). При переписи населения, например, единицей наблюдения является человек; если же изучению подлежат также и семьи, то устанавливаются две единицы наблюдения: отдельный человек и семья. Правильное определение единицы наблюдения имеет существенное значение не только для проведения самого наблюдения, но и для последующих стадий статистического исследования.

От единицы наблюдения следует отличать *единицу совокупности*, т.е. первичный элемент объекта статистического наблюдения, признаки которого подлежат регистрации и который является основой ведущегося счета. Например, при переписи оборудования единицей наблюдения является каждое предприятие, а единицей совокупности − станок и т.д. Таким образом, единица наблюдения является источником сведений, которые получают в результате наблюдения, а единицы совокупности − носителем признаков, подлежащих наблюдению.

Следует отметить, что единица совокупности и единица наблюдения могут совпадать. Так, например, при переписи населения единицей совокупности и единицей наблюдения является каждый житель страны, но при изучении спроса населения на различные продукты единицей совокупности будет каждый зарегистрированный случай спроса, как удовлетворенного, так и неудовлетворенного, а единицей наблюдения будет торговая фирма (предприятие, компания и т.д.), в которой это наблюдение производится. Четкое определение единицы совокупности и единицы наблюдения является важным элементом научной организации статистического наблюдения.

Исходя из содержания объекта, цели и задач статистического наблюдения разрабатывается программа наблюдения.

Программа наблюдения − перечень вопросов (показателей), по которым регистрируются единицы наблюдения и на которые должны быть получены правильные, исчерпывающие ответы.

Вопросы программы статистического наблюдения и ответы на них находят отражение в статистическом формуляре. *Формуляр* статистического наблюдения − это специальный документ, в котором регистрируются ответы на вопросы программы наблюдения. Он представляет собой разграфленный лист бумаги, в котором содержится перечень вопросов программы, свободные места для записи ответов (с указанием шифров и кодов) на них. Формуляр наблюдения состоит из двух частей: титульной и адресной. Титульная часть формуляра наблюдения включает: наименование статистического наблюдения и органа, его проводящего, а также дату и наименование органа, утвердившего данный формуляр. Адресная часть формуляра содержит запись точного адреса единицы или совокупности единиц наблюдения, их соподчиненность, иногда − сроки и место рассылки заполненных формуляров.

В статистике различают две системы статистического формуляра: индивидуальную (формуляр-карточка) и списочную (формуляр-список). *Индивидуальный формуляр* − это формуляр, предназначенный для регистрации в нем ответов на вопросы программы наблюдения только об одной единице наблюдения. *Списочный формуляр* − это формуляр, предназначенный для регистрации в нем ответов на вопросы программы наблюдения о нескольких единицах наблюдения.

К статистическим формулярам составляется *инструкция*, в которой подробно разъясняется, как следует заполнять статистический формуляр.

Для успешного проведения статистического наблюдения разрабатывается *организационный* план. В нём указываются: органы наблюдения, время наблюдения, сроки наблюдения, необходимые подготовительные работы, порядок проведения наблюдения, приёма и сдачи материалов, получения и предоставления предварительных и окончательных итогов.

Для правильной характеристики изучаемого объекта важное значение имеет установление времени наблюдения. В статистике различают объективное и субъективное время наблюдения. *Объективным временем* называется время, к которому относится данное наблюдение. *Субъективное время наблюдения* − это время производства наблюдения, т.е. период, в течение которого производится регистрация единиц совокупности.

Срок (период) наблюдения − это время, в течение которого производится заполнение статистических формуляров, т.е. осуществляется регистрация единиц наблюдения по установленной программе. Срок наблюдения определяется рядом факторов. В первую очередь он зависит от специфики и особенностей объекта наблюдения. Так, чем крупнее объект наблюдения, тем, при прочих равных условиях, требуется больше времени для проведения статистического наблюдения над ним. Срок диктуется также программой наблюдения, ее объемом и сложностью признаков, подлежащих регистрации. Срок наблюдения, как правило, предполагает указание даты начала и завершения статистического наблюдения.

Критический момент статистического наблюдения − это момент времени (конкретный год, день и час), по состоянию на который производится регистрация собираемых сведений в процессе статистического наблюдения.

1.2.5 Ошибки статистического наблюдения

Всякое статистическое наблюдение ставит задачу получения таких данных, которые точнее бы отражали действительность. Точность и достоверность собираемой статистической информации – важнейшая задача статистического наблюдения.

Расхождение между установленными статистическим наблюдением и действительными значениями изучаемых величин называется *ошибками наблюдения.*

В статистике различают о ш и б к и регистрации и ошибки репрезентативности. *Ошибки регистрации* могут возникнуть как вследствие неправильного установления факта, так и вследствие неправильной записи. В результате проверки статистических данных могут быть обнаружены ошибки случайные и систематические. *Случайные ошибки регистрации* − это ошибки, которые возникают вследствие различных случайных причин. *Систематические ошибки регистрации* − это неточности, возникающие в силу определенных и постоянно действующих на протяжении всего статистического наблюдения в одном направлении факторов.

Систематические ошибки могут быть преднамеренные и непреднамеренные. *Преднамеренные систематические ошибки регистрации* − это ошибки, являющиеся результатом того, что опрашиваемый умышленно, сознательно сообщает регистратору неправильные данные. К преднамеренному искажению данных относятся занижения величины прибыли в формах отчетности некоторых коммерческих структур. Непреднамеренные систематические ошибки регистрации – это ошибки, которые носят нечаянный характер, допускаются неумышленно. Такие ошибки могут быть результатом недостаточной квалификации работников, их небрежности в работе. Сюда относятся пропуски в записях, использование неисправных измерительных приборов и т. д.

Непреднамеренные ошибки чаще всего возникают в результате небрежности или недостаточной квалификации счетного аппарата, плохой постановки первичного учета. *Ошибки регистрации* – это расхождение между зафиксированным при наблюдении статистическом значении признака и действительным его значением в результате неправильной, ошибочной регистрации ответа на вопрос статистического формуляра. Этот вид ошибок может быть и при сплошном, и при несплошном наблюдении.

При несплошном наблюдении зачастую возникают ошибки репрезентативности. *Ошибки репрезентативности* − это расхождение между значениями изучаемого признака или показателя в отобранной и обследованной части совокупности (выборочной) и его значениями во всей исходной (генеральной) совокупности. Причина возникновения данного рода ошибок заключается в том, что отобранная и обследованная часть изучаемой совокупности недостаточно точно отражает состав всей совокупности в целом. Ошибки репрезентативности могут быть случайными и систематическими. Случайные ошибки репрезентативности − это ошибки, возникающие в силу несплошного характера статистического наблюдения, когда совокупность отобранных на основе принципа беспристрастного, непреднамеренного, случайного отбора единиц наблюдения недостаточно полно и точно воспроизводит совокупность в целом. Систематические ошибки репрезентативности – это ошибки, возникающие вследствие нарушения принципов беспристрастного, непреднамеренного отбора единиц изучаемой (генеральной) совокупности, которые должны быть подвергнуты наблюдению.

Для выявления и устранения допущенных при регистрации ошибок может применяться счётный и логический контроль собранного материала.

Счётный контроль заключается в проверке точности арифметических расчётов, применявшихся при составлении отчётности или заполнении формуляров обследования (проверка итоговых данных по графам или строкам, нарастающих итогов, вычисленных относительных, средних и других величин). Например, следует проверить, правильно ли вычислена сумма товарооборота за квартал по данным месячной отчетности, годового товарооборота по данным квартальной отчетности и т.п.

В результате статистического наблюдения данные могут быть представлены исчерпывающе, арифметические действия произведены правильно, однако логически факты противоречат друг другу. В этих случаях применяется логический контроль. *Логический контроль* заключается во взаимном сопоставлении ответов на вопросы программы наблюдения путём их логического осмысления или путём сравнения полученных данных с другими источниками по тому же вопросу.

1.3. Сводка и группировка статистических данных

1.3.1 Сводка статистических данных

Собранный в процессе статистического наблюдения материал (статистическая информация) представляет собой разрозненные первичные цифровые сведения об отдельных единицах изучаемого явления (объекта). В таком виде материал еще не характеризует явления в целом, так как он слишком разрознен и не классифицирован. Из него не видно ни состава, ни численности, ни существа связей одного явления с другими. Указанные признаки могут быть получены лишь в процессе обработки материалов наблюдения. Это и является задачей *второго этапа* статистической работы − *сводки и группировки результатов статистического наблюдения.*

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

Если производится подсчёт только общих итогов по изучаемой совокупности единиц наблюдения, то сводка называется *простой*.

По технике или способу выполнения сводка может быть *ручной* либо *механизированной.*

Статистическая сводка должна проводиться по программе и плану.

Программа статистической сводки предусматривает следующие э т а п ы:

− выбор группировочных признаков;

− определение порядка формирования групп;

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

– разработка макетов статистических таблиц для представления результатов сводки.

План статистической сводки содержит указания о последовательности и сроках выполнения отдельных частей сводки, её исполнителях и порядке изложения и представления результатов.

В сводке статистического материала отдельные единицы статистической совокупности объединяются в группы при помощи метода группировок.

Группировка является методом исследования сущности явлений путем расчленения совокупности на группы по определенным признакам.

Выявление связей между явлениями и их признаками – основная задача группировки статистического материала. *Статистическая группировка* – это расчленение изучаемой совокупности на группы и подгруппы по определенным характерным достаточным признакам для глубокого и всестороннего изучения явлений. Например, группировка предприятий по формам собственности.

Особым видом группировок является *классификация*, представляющая собой устойчивую номенклатуру классов и групп, образованных на основе сходства и различия единиц изучаемого объекта.

Метод статистических группировок позволяет разрабатывать первичный статистический материал. На основе группировки рассчитываются сводные показатели по группам, появляется возможность их сравнения, анализа причин различий между группами, изучения взаимосвязей между признаками. Расчёт сводных показателей в целом по совокупности позволяет изучить её структуру.

1.3.2 Задачи и виды группировок

Метод группировок применяется для решения задач, возникающих в ходе научного статистического исследования:

– выявления социально-экономических типов явлений;

– изучения структуры явления и структурных сдвигов, происходящих в нём;

– выявления связей и зависимостей между отдельными признаками явлений.

Для решения этих задач применяются, соответственно, **три вида группировок:** типологические, структурные и аналитические (факторные).

Типологическая группировка – это расчленение разнородной совокупности на отдельные качественно однородные группы, социально-экономические классы и типы в соответствии с правилами научной группировки и выявление на этой основе экономических типов явлений. При построении типологической группировки в качестве группировочных признаков могут выступать как количественные, так и атрибутивные (качественные) признаки.

Структурной называется группировка, в которой происходит разделение выделенных с помощью типологической группировки типов явлений, однородных совокупностей на группы, характеризующие их структуру по какому-либо варьирующему признаку. К структурным относится группировка населения по размеру среднедушевого дохода, группировка хозяйств по объёму продукции. Анализ структурных группировок, взятых за ряд периодов или моментов времени, показывает изменение структуры изучаемых явлений или *структурные сдвиги*. В изменении структуры общественных явлений отражаются важнейшие закономерности их развития.

Одной из задач группировок является исследование связей и зависимостей между изучаемыми явлениями и их признаками. *Аналитическая группировка* – это группировка, выявляющая взаимосвязи и взаимозависимости между изучаемыми социально-экономическими явлениями и признаками, их характеризующими.

В основе аналитической группировки лежит факторный признак, каждая выделенная группа характеризуется средними значениями результативного признака. Так, группируя достаточно большое число рабочих по факторному признаку x – квалификации (разряду) с указанием их заработной платы, можно заметить прямую зависимость результативного признака y – средней месячной заработной платы рабочих – от квалификации: чем выше квалификация, тем выше и средняя месячная заработная плата.

Используя в аналитических группировках методы математической статистики, можно определить показатель тесноты связи между изучаемыми признаками.

Группировочный признак. Каждая единица совокупности обладает рядом признаков, которые изучаются посредством проведения группировок.

Важнейший вопрос группировки − оптимальный выбор группировочного признака. Он зависит от характера изучаемых явлений и целей группировки. Признаки, по которым производится распределение единиц изучаемой совокупности на группы, называются *группировочными признаками* или *основанием группировки*. Признак должен быть существенным, а не второстепенным или малозначительным.

При выборе группировочных признаков необходимо помнить, что одни и те же признаки могут иметь различные значения в зависимости от конкретных условий, места и времени. Поэтому с изменением в развитии изучаемого явления видоизменяются приемы группировки, берутся другие группировочные признаки. Группировочные признаки условно различаются на качественные и количественные. Группировочный признак называют количественным, если он выражается числом. Примером группировок по количественным признакам может служить группировка работников по стажу работы, возрасту и т. п. Группировочный признак может иметь качественное выражение: группировка работников по полу, образованию. Признак, который характеризует свойство, качество данного явления и не имеет количественного выражения, называется *атрибутивным*.

Группировки по одному признаку называются *простыми*. Если для формирования групп берут два и более признаков, т.е. группы, образованные по одному признаку, подразделяются на подгруппы по другому, а полученные в результате этого подгруппы подразделяются (каждая в отдельности) еще на подгруппы и т.д., то такие группировки называются *комбинационными.*

По сравнению с простыми комбинационные группировки обладают дополнительными аналитическими свойствами. Они помогают выявить такие различия в связи между исследуемыми признаками, которые нельзя обнаружить при простой группировке. Комбинационные группировки имеют особенно важное значение при изучении сложных явлений и процессов, представляющих собой взаимодействие ряда элементов.

Отбор группировочных признаков проходит следующие стадии: вначале определяется цель, познавательная задача предполагаемой группировки, затем − специфическое содержание признаков, которые должны быть положены в основание группировки, устанавливаются число групп и количественные границы признаков. Все эти вопросы решают, исходя из существа изучаемого явления. Определение числа групп и количественных границ признаков зависит от цели группировки и от того, с какими признаками приходится иметь дело. При группировке по атрибутивным (качественным) признакам статистическая совокупность распределяется на столько групп, сколько разновидностей имеет признак (по полу − на две группы, по национальному составу − на столько групп, сколько имеется национальностей).

Если атрибутивный признак имеет большое количество разновидностей (профессии, наименование выпускаемой продукции, оборудования, товаров), то для обоснованного объединения их в группы разрабатываются номенклатуры и классификации. *Номенклатура* − это твердо установленный полный подробный перечень отдельных видов изучаемой совокупности.

Количественным называется признак, характеризующий размеры, величину совокупности и дающий возможность расчленить ее на группы по величине индивидуальных значений группировочного признака.

Признак пространства – это адресный признак (адрес предприятий, фирмы). При изучении изменений явлений во времени группируют по признаку времени. *Признак времени* позволяет установить хронологию событий (даты, годы).

Признаки также бывают первичными и вторичными. *Первичные* признаки характеризуют абсолютные размеры изучаемых явлений (численность сотрудников компании, издержки производства, издержки обращения, стоимость основных фондов и т.д.), *вторичные* являются производными от первичных и показывают структуру группируемых явлений (фондовооруженность, производительность труда, себестоимость единицы продукции и т. д.).

Группировки по количественному признаку. При группировке изучаемых явлений по одному признаку, а тем более при комбинации двух-трех признаков можно получить значительное число групп. Для решения вопроса о числе групп необходимо сначала выяснить положение и роль отдельных групп, тенденции их развития и затем выделить характерные, типичные группы, вытекающие из анализа изучаемого явления.

Если признак изменяется в широких пределах и имеет много различных значений, возникает вопрос об определении интервала группировки. *Интервал* – это разность между наибольшим и наименьшим значениями признака, т.е. промежуток колеблемости числового значения признака для каждой группы в пределах «от–до». Интервалы могут быть равными и неравными. Расчет равной величины интервала производится по формуле

$$i = \frac{X_{\max} - X_{\min}}{\text{Число групп}},$$

где $X_{\max}$, $X_{\min}$ – наибольшее и наименьшее значения признака.

Пример. Провести группировку рабочих вагонного депо по величине начисленной им за месяц заработной платы на основании следующих данных. Начислено заработной платы: по 70 у. е. – трём рабочим, по 80 у. е. – четырём, по 90 у. е. – трём, по 100 у. е. – шести, по 110 у. е. – пяти, по 120 у. е. – четырём, по 130 у. е. – одному, по 150 у. е. – восьми. Требуется распределить рабочих по размеру заработной платы, установив четыре группы с равными интервалами.

Решение. Находим разность между наибольшим и наименьшим значениями признака: 150 – 70 = 80 у. е.; определяем величину интервала: 80 / (число групп), т. е. (80 / 4) = 20 у. е. В результате получаем следующие группы (таблица 1.1).

Таблица 1.1 – **Группировка рабочих по размеру месячной заработной платы**

Группы рабочих по размеру месячной заработной платы, у. е.	Численность рабочих, чел.
От 70 до 90	7
„ 90 до 110	9
„ 110 до 130	9
„ 130 до 150	9
Итого	34

Изменим запись групп в таблице, например, первый интервал запишем до 90, последний – свыше 130. В этом случае первый и последний интервалы называются открытыми, оклад 90 у. е. служит верхней границей для первого интервала, оклад 130 у. е. – нижней границей последнего интервала; второй интервал (от 90 до 110 у. е.) будет закрытым. Таким образом, **открытые** интервалы имеют одну какую-нибудь обозначенную границу, верхнюю или нижнюю, и неопределенные границы, **закрытые** – и верхнюю, и нижнюю. При расчётах надо помнить, что длина открытых интервалов принимается равной длине рядом стоящих закрытых интервалов.

Виды рядов распределения. *Ряд распределения* – это упорядочение по определенному варьирующему признаку однородных групп единиц совокупности. В зависимости от признака, положенного в основание построения ряда распределения, различают атрибутивные и вариационные ряды распределения.

Атрибутивный ряд распределения – это ряд распределения, построенный по качественным признакам, не имеющим числового выражения и характеризующим свойство, качество изучаемого социально-экономического явления.

Вариационный ряд распределения строится по количественному признаку.

В зависимости от характера вариации признака различают дискретные и интервальные вариационные ряды распределения. *Дискретный вариационный ряд* распределения – это ряд, в котором группы составлены по признаку, изменяющемуся дискретно и принимающему целые значения. График дискретного вариационного ряда называется полигоном. *Интервальный вариационный ряд распределения* – это ряд, в котором группировочный признак, составляющий основание группировки, может принимать в определенном интервале любые значения. График интервального вариационного ряда распределения называется гистограммой.

Вариационные ряды распределения состоят из двух элементов: *вариант* и *частот*. Числовые значения количественного признака в вариационном ряду распределения называются *вариантами*. *Частоты* – это численности отдельных вариант или каждой группы вариационного ряда, т.е. это числа, показывающие, как часто встречаются те или иные варианты в ряду распределения. Сумма всех частот называется объёмом совокупности и определяет число элементов всей совокупности. Процесс накопления частот можно представить в виде графика, который называется *кумулятой*. Если при графическом изображении вариационного ряда в виде кумуляты оси поменять местами, то получим *огиву*.

Частости – это частоты, выраженные в виде относительных величин (долях единиц или процентах). Сумма частостей равна единице или 100 %.

1.4 Статистические таблицы, их элементы, виды и правила построения

1.4.1 Понятие о статистической таблице

Результаты статистического наблюдения, сводки и группировки обычно представляются в форме таблиц. Таблица может быть наглядным, кратким и последовательным изложением полученных цифровых данных. С помощью таблицы статистический материал излагается наиболее рационально. Основанием любой таблицы является сетка-скелет, в которой вертикальные столбцы называются графами, а горизонтальные − строками. Внешне таблицы представляют собой сетку из вертикальных и горизонтальных линий, в которой записываются числовые данные. Графы (сказуемое) и строки (подлежащее) образуют макет таблицы.

Статистическое подлежащее таблицы − это то, о чем говорится и что характеризуется в таблице.

Статистическое сказуемое таблицы показывает, какими признаками характеризуется подлежащее.

По построению подлежащего различают следующие виды таблиц: *простые, групповые, комбинационные*. В простой таблице подлежащее не делится на группы. Простые таблицы бывают перечневые (содержится перечень единиц, составляющих объект изучения), динамические (приводятся периоды времени), территориальные (даётся перечень территорий, стран, областей, городов). Часто они строятся в различном сочетании (перечневые и динамические, территориальные и динамические).

Групповыми называются таблицы, в которых подлежащее разделено на группы по какому-либо одному признаку (таблица 1.2).

Таблица 1.2 − **Структура работников дистанции гражданских сооружений**

Категория работников	Количество человек	В процентах к общей численности
Рабочие	387	73,43
Специалисты	51	9,68
Служащие	80	15,18
Руководители	9	1,71
Всего	447	100,00

Комбинационными называются такие таблицы, в которых подлежащее делится на группы не по одному, а по нескольким признакам, причем каждая группа, образованная по одному признаку, делится на подгруппы по другому признаку. Например, можно произвести группировку магазинов по размеру товарооборота, а затем каждую группу распределить по площади торгового зала. Примером группировки по двум признакам является таблица 1.3.

При построении таблиц нужно соблюдать следующие правила:

− таблица должна иметь небольшие размеры, чтобы ее удобно было читать и анализировать;

− название таблицы, заголовки подлежащего и сказуемого должны быть точными, краткими и ясными;

− в таблице должны быть точно обозначены единицы измерения, а также территория и период, к которым относятся приводимые данные;

− приводимые в подлежащем и сказуемом признаки должны быть расположены в логическом порядке с учётом необходимости их совместного рассмотрения. Обычный принцип размещения – от частного к общему, т.е. сначала показывают слагаемые, а в конце подводят итоги. Когда приводятся не все слагаемые, а лишь наиболее важные из них, применяется противоположный принцип: сначала показываются общие итоги, а затем выделяют наиболее важные части («в том числе», «из них»). Следует различать «Итого» и «Всего». «Итого» является итогом для определённой части совокупности, а «Всего» – итог для совокупности в целом;

− если число показателей подлежащего и сказуемого таблицы значительно, то строки и графы таблиц следует пронумеровать. Это значительно облегчает пользование таблицей, а в ряде случаев даёт возможность пояснить, какие действия надо произвести, чтобы получить те или иные показатели (например, графа 2 + графа 3);

− таблица может сопровождаться примечаниями, в которых указываются источники данных, более подробно раскрывается содержание показателей, даются другие пояснения, а также оговорки в случае, если таблица содержит данные, полученные в результате вычислений;

− при оформлении таблиц обычно применяются такие условные обозначения: при отсутствии данных следует ставить знак тире (−), а при отсутствии сведений − многоточие (…) или «нет сведений», *x* – если явление не имеет осмысленного содержания. Если сведения имеются, но числовое значение меньше принятой в таблице точности, оно выражается дробным числом (0,0);

− в таблице должны быть подсчитаны итоги;

− цифровой материал должен даваться с одинаковой степенью точности.

Таблица 1.3 − **Группировка магазинов организации рабочего снабжения (ОРС) по размеру товарооборота и по площади торгового зала**

Группы магазинов по размеру квартального товарооборота, млн руб.	Площадь торгового зала, кв. м	Количество розничных предприятий, ед.	Розничный товарооборот, млн руб.
До 10	До 30	1	1,2
	30−50	4	14,2
	50−100	2	9,3
	Свыше 100	3	28,4
От 11 до 20	До 30	−	
	30−50	1	12,8
	50−100	6	90,1
	Свыше 100	8	132,6

1.4.2 Статистические графики и правила их построения

Современный анализ социально-экономических явлений немыслим без применения графического метода представления данных. *Графический метод* есть метод условного изображения статистических данных при помощи геометрических фигур, линий, точек и разнообразных символических образов.

Главное достоинство статистических графиков – наглядность. Для построения графика необходимо точно определить, для каких целей он составляется, тщательно изучить исходный материал. Но самое главное условие – овладение методологией графических изображений. В статистическом графике различают следующие основные элементы: поле графика, графический образ, пространственные и масштабные ориентиры, экспликация графика.

Полем графика называют место, на котором он выполнятся. Это листы бумаги, географические карты, план местности и т. п. Поле графика характеризуется его форматом (размерами и пропорциями сторон). Размер поля графика зависит от его назначения. Стороны поля статистического графика обычно находятся в определенной пропорции. Принято считать, что наиболее близким к оптимальному для зрительного восприятия является график, выполненный на поле прямоугольной формы с соотношением сторон от 1:1,3 до 1:1,5. Этот вариант именуется правилом «золотого сечения». Иногда используется и поле графика с равными сторонами, т.е. имеющее форму квадрата.

Графический образ – это символические знаки, с помощью которых изображаются статистические данные: линии, точки, плоские геометрические фигуры (прямоугольники, квадраты, круги и т. д.).

Размещение графических образов на поле графика определяют пространственные ориентиры. Они задаются координатной сеткой или контурными линиями и делят поле графика на части, соответствующие значениям изучаемых показателей. В статистических графиках чаще всего применяется система прямоугольных (декартовых) координат. Для построения статистических графиков используется обычно только первый и изредка первый и четвертый квадраты. В практике графического изображения применяются также полярные координаты (рисунок 1.1).

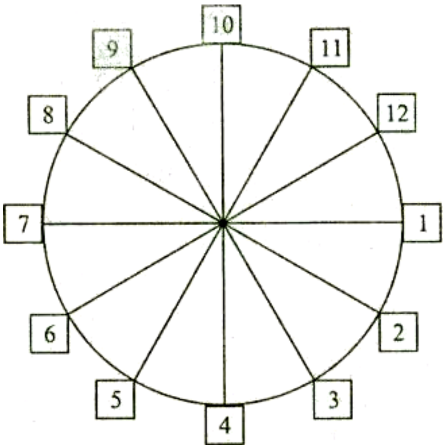

Рисунок 1.1 – Числовые интервалы в полярной системе координат

Они необходимы для наглядного изображения циклического движения во времени. В полярной системе координат один из лучей, обычно правый горизонтальный, принимается за ось координат, относительно которой определяется угол луча. Второй координатой считается ее расстояние от центра сетки, называемое *радиусом.* В радиальных графиках лучи обозначают моменты времени, а окружности – величины изучаемого явления.

На статистических картах пространственные ориентиры задаются контурной сеткой (контуры рек, береговая линия морей и океанов, границы государств) и определяют те территории, к которым относятся статистические величины.

Масштабные ориентиры статистического графика определяются масштабом и системой масштабных шкал. Масштаб статистического графика − это мера перевода числовой величины в графическую. *Масштабной шкалой* называется линия, отдельные точки которой могут быть прочитаны как определенные числа. Шкала имеет большое значение в графике и включает три элемента: линию (или носитель шкалы), определенное число помеченных черточками точек, которые расположены на носителе шкалы в определённом порядке, цифровое обозначение чисел, соответствующих отдельным помеченным точкам. Как правило, цифровым обозначением снабжаются не все помеченные точки, а лишь некоторые из них, расположенные в определенном интервале. По правилам числовое значение необходимо помещать строго против соответствующих точек, а не между ними.

Носитель шкалы может представлять собой как прямую, так и кривую линии. Поэтому различают шкалы *прямолинейные* (например, миллиметровая линейка) и криволинейные – дуговые и круговые (циферблат часов). Графические и числовые интервалы бывают равными и неравными. Если на всем протяжении шкалы равным графическим интервалам соответствуют равные числовые, то такая шкала называется *равномерной*. Когда же равным числовым интервалам соответствуют неравные графические интервалы и наоборот, то перед нами *неравномерная* шкала. *Масштабом равномерной шкалы* называется длина отрезка, принятого за единицу и измеренного в каких-либо мерах.

Последний элемент графика – *экспликация*. Каждый график должен иметь словесное описание. Оно включает его содержание, подписи вдоль масштабных шкал и пояснения к отдельным частям графика.

Диаграммы – наиболее распространенный способ графических изображений. Применяются диаграммы для наглядного сопоставления в различных аспектах независимых друг от друга совокупностей. *Статистические карты* – графики количественного распределения по конкретной территории.

Статистические графики можно классифицировать по разным признакам: назначению (содержанию), способу построения и характеру графического образа.

По назначению (содержанию) можно выделить графики сравнения, графики относительных величин (структуры, динамики). Особым видом графики являются диаграммы распределения величин, представленных вариационным рядом. Это гистограмма, полигон, огива, кумулята.

По способу построения графики можно разделить на диаграммы, картодиаграммы и картограммы.

По характеру графического образа различают графики точечные, линейные, полосовые, круговые, секторные, фигурные и объёмные.

1.5 Абсолютные и относительные статистические величины

1.5.1 Абсолютные величины, их виды, важность и возможность получения

Результаты статистического наблюдения регистрируются в виде абсолютных величин, которые отражают либо объём, либо уровень развития явления. Они имеют размерность (единицы измерения) и могут быть положительными и отрицательными. Абсолютные величины выражаются в различных единицах измерения − натуральных, стоимостных (денежных), условных, трудовых.

Натуральные единицы измерения характеризуют величину и размер изучаемых явлений. Они выражаются в определенных единицах измерения: килограммах, штуках, тоннах, метрах, километрах и т.д., то есть либо в объемных измерителях, либо в весовых, либо в метрах длины и площади. Все эти единицы измерения принято называть натуральными, а учет в них – учетом в натуральном выражении. Натуральные единицы можно суммировать только по однородным продуктам.

Стоимостные единицы применяются для оценки в стоимостном выражении многих статистических показателей: объема выпущенной продукции промышленности, размера розничного товарооборота и т.д. Показатели, выраженные в стоимостных измерителях, можно суммировать, получать по ним итоговые данные. При использовании стоимостных измерителей принимают во внимание изменение цен с течением времени. Этот недостаток стоимостных измерителей преодолевают применением «неизменных» или «сопоставимых» цен одного и того же периода.

Условные единицы измерения. Необходимость в применении условных единиц вызывается тем, что в ряде случаев не все виды даже однородной продукции можно суммировать. Так, нельзя суммировать мыло, топливо, консервы и т.п., так как мыло, например, имеет различный процент жирности, топливо − различную калорийность и т.п. Следовательно, условные единицы измерения применяются для учета однородной продукции различных разновидностей. Например, консервы выпускают в банках разной емкости. Поэтому для правильного определения объема производства консервов применяется пересчет этой продукции в тубах (в тысячах условных банок). За одну условную банку принят вес продукции нетто 400 г. Например, объём грузовых перевозок на железнодорожном транспорте определяется в тонно-километрах, пассажирских перевозок – в пассажиро-километрах, а общий объём работы – в приведенных тонно-километрах.

Трудовые единицы измерения — это человеко-часы, человеко-дни, человеко-месяцы и т.п. Трудовые измерители характеризуют использование трудовых ресурсов или затраты труда в производственной деятельности, сфере обслуживания и т.п.

По способу выражения абсолютные величины подразделяются на индивидуальные и суммарные.

Индивидуальные абсолютные величины − это показатели, выражающие размеры количественных признаков отдельных единиц исследуемых объектов, например размер посевной площади конкретного фермерского хозяйства, количество договоров, заключенных дилером на бирже за месяц, выработка определенного рабочего за указанный месяц и т.д. Они получаются непосредственно в процессе статистического наблюдения и играют значительную роль в статистическом исследовании.

Суммарные абсолютные величины характеризуют итоговое значение признака всех единиц изучаемой совокупности или отдельных ее групп и получаются в результате суммирования индивидуальных абсолютных величин.

Абсолютные статистические показатели подразделяются на показатели объема и показатели уровня.

Показатели объема позволяют характеризовать величину всей совокупности или ее частей. *Показатели уровня* характеризуют величину нагрузки единицы одной совокупности элементами другой совокупности.

Обобщающие абсолютные величины могут быть получены прямым подсчётом единиц изучаемой совокупности (например, численность локомотивов инвентарного парка); суммированием значений, характеризующих их количественные признаки (например, пробег локомотивов во главе поездов), а также в результате расчётов (например, наличный парк вагонов на железной дороге на конец отчётных суток определяется так: к величине наличного парка вагонов на дороге на начало суток прибавляется число вагонов, поступивших на дорогу, и вычитается число вагонов, убывших с дороги за сутки).

1.5.2 Относительные статистические величины

Кроме абсолютных величин, в анализе хозяйственной деятельности, в статистической и экономической работе широко используются относительные величины. *Относительная величина* – это обобщающий показатель, который получается в результате деления одной величины на другую и даёт числовую меру соотношения между ними.

Относительные величины – одно из важнейших средств анализа статистических данных. Они широко используются при изучении развития различных сфер деятельности, анализе работы предприятий или характеризуют территориальное размещение производства. Основное условие правильного расчёта относительных величин – сопоставимость сравниваемых показателей и наличие реальных связей между изучаемыми явлениями. Величина, с которой производится сравнение (знаменатель дроби), обычно называется *базой сравнения* или *основанием*. Величина, находящаяся в числителе, называется *текущей*.

В зависимости от результата сравнения для удобства чтения и восприятия относительной величины выделяется её целая часть. В качестве базы сравнения может быть выбрана 1, тогда в результате сравнения получится коэффициент. Если за базу сравнения принять 100 – получится процент (%), 1000 – промилле (‰) , 10000 – продецимилле(%$_{oo}$) .

Сопоставляемые величины могут быть как одноимёнными, так и разноимёнными (в последнем случае их наименования образуются от наименований сравниваемых величин, например, руб./чел.; т-км / км; руб./ м^2).

Относительные величины подразделяются на несколько видов: относительные величины динамики, планового задания, выполнения планового задания, относительные величины структуры, относительные величины координации, относительные величины сравнения и относительные величины интенсивности.

Относительная величина динамики ($i_{\text{д}}$) характеризует изменение (увеличение или снижение) показателей текущего периода по сравнению с прошлым периодом. Относительные величины динамики называются *темпами роста*:

$$i_{\text{д}} = \frac{\text{Фактические данные текущего периода}}{\text{Фактические данные прошлого периода}} \cdot 100.$$

Относительная величина динамики показывает развитие явлений во времени: рост грузооборота, пассажирооборота, изменение потребления различных видов ресурсов и т.д.

Существуют две схемы расчёта относительных величин динамики: цепная и базисная. *Относительные величины динамики цепные* характеризуют изменения каждого последующего уровня ряда динамики по сравнению с уровнем, ему предшествующим.

Относительные величины динамики базисные показывают темпы развития за каждый данный отрезок времени по сравнению с уровнем, принятым за базу сравнения, чаще всего это бывает начальный уровень ряда динамики. Они широко используются при характеристике развития явлений за определённый период времени.

Относительная величина планового задания ($i_{\text{пл.з}}$) рассчитывается как отношение уровня, запланированного на предстоящий период, к уровню, фактически сложившемуся в этом периоде.

Относительная величина выполнения планового задания ($i_{\text{вып.пл}}$) представляет собой отношение фактически достигнутого в данном периоде уровня к запланированному. Например, по плану в 2013 г. намечено произвести продукции на 3264,7 млн у. е., фактически произведено на 3330,0 млн у. е. Плановое задание выполнено на 3330,0 / 3264,7·100 = 102,0 %.

В плановом задании может устанавливаться величина прироста или снижения показателя. Например, планировалось снижение себестоимости единицы продукции на 24,2 у. е., а фактически снижение составило 27,5 у. е., плановое задание по снижению себестоимости выполнено на 27,5 : 24,2·100 = = 113,6 %.

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

$$i_{\text{д}} = i_{\text{пл.з}} i_{\text{вып.пл}}.$$

Относительная величина структуры характеризует состав изучаемой совокупности и определяется отношением отдельных частей к целому:

$$\text{Доля, \%} = \frac{\text{Часть совокупности}}{\text{Вся совокупность}} \cdot 100.$$

Относительная величина координации ($i_{\text{к}}$) характеризует соотношение между частями (элементами) одной совокупности:

$$i_{\text{к}} = \frac{\text{Одна часть совокупности}}{\text{Другая часть той же совокупности}} \cdot 100.$$

Относительные величины координации выражаются в виде коэффициентов. При исчислении относительных величин координации велико значение не только выбора базы сравнения, но и вообще выбора явлений, которые могут быть сравнимы между собой.

Относительная величина сравнения показывает соотношение одноименных величин, относящихся к разной территории или к разным объектам, за один и тот же период времени и применяется для сопоставления экономических показателей.

Рассчитывая относительные величины сравнения, следует обращать внимание на сопоставимость сравниваемых показателей с позиций методологии их исчисления, поскольку по целому ряду показателей методы их исчисления в разных странах или в разные периоды времени неодинаковы. Поэтому прежде чем рассчитывать относительные показатели сравнения, приходится решать задачу пересчёта сравниваемых показателей по единой методологии.

Относительная величина интенсивности ($i_{\text{и}}$) показывает степень распространенности данного явления в изучаемой среде и образуется в результате сравнения разноименных, но определенным образом связанных между собой абсолютных величин:

$$i_{\text{и}} = \frac{\text{Одна совокупность, характеризующая явление}}{\text{Другая совокупность, характеризующая среду}} \cdot 100.$$

Относительные величины интенсивности, в отличие от других видов относительных величин, всегда выражаются именованными числами, например, плотность населения 8,6 чел./км^2. Разновидностью относительных величин интенсивности являются относительные показатели уровня экономического развития, характеризующие уровни ВВП, ВНП, национального дохода и других показателей на душу населения и играющие важную роль в оценке развития экономики страны.

Пользуясь в анализе относительными величинами, необходимо показать, какие абсолютные показатели за ними скрываются. В противном случае можно прийти к неправильным выводам.

1.6 Средние величины и показатели вариации

1.6.1 Понятие о средних величинах

Наряду с абсолютными и относительными величинами в статистике большое применение находят средние величины.

В результате группировки единиц совокупности по величине варьирующего признака получают ряды распределения – первичную характеристику массовой статистической совокупности. Чтобы охарактеризовать такую совокупность в целом, часто пользуются средней величиной. *Средняя величина* – это обобщающая характеристика однородной совокупности явлений по определённому признаку.

Вычисление средней – один из распространённых приёмов обобщения. Средний показатель отражает то общее, что характерно (типично) для всех единиц изучаемой совокупности, в то же время он игнорирует различия отдельных единиц. В каждом явлении и его развитии имеет место сочетание случайности и необходимости. При исчислении средних в силу действия закона больших чисел случайности взаимопогашаются, уравновешиваются, поэтому можно абстрагироваться от несущественных особенностей явления, от количественных значений признака в каждом конкретном случае. В способности абстрагироваться от случайности отдельных значений, колебаний и заключена научная ценность средних как обобщающих характеристик совокупностей.

1.6.2 Виды средних величин и порядок их вычисления

В статистике используются различного вида средние величины: средняя арифметическая, средняя гармоническая, средняя геометрическая, средняя квадратическая, средняя хронологическая. Однако больше всего в экономической практике приходится употреблять среднюю арифметическую, которая может быть исчислена как средняя арифметическая простая и взвешенная. Чтобы определить *среднюю арифметическую простую*, нужно сумму всех значений данного признака Σx разделить на число единиц n, обладающих этим признаком. Она применяется в тех случаях, когда имеются несгруппированные индивидуальные значения признака.

$$\bar{X}_{\text{ар}} = \frac{\sum x}{n}.$$

Средняя арифметическая взвешенная есть частное от деления суммы произведений вариантов и соответствующих им частот Σxf на сумму всех частот Σf. Частоты (f), фигурирующие в формуле средней, принято называть *весами,* вследствие чего средняя арифметическая, вычисленная с учетом весов, и получила название взвешенной:

$$\bar{X}_{\text{ар}} = \frac{\sum xf}{\sum f}.$$

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

$$\bar{X}_{\text{ар}} = \frac{\sum xd}{\sum d},$$

где d – частость, т.е. доля каждой частоты в общей сумме всех частот.

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

$$\bar{X}_{\text{ар}} = \sum xd.$$

Часто приходится исчислять среднюю по групповым средним или по средним отдельных частей совокупности (частным средним), т. е. *среднюю из средних*. Так, например, средняя продолжительность жизни граждан страны представляет собой среднее из средних продолжительностей жизни по отдельным регионам данной страны.

Средние из средних рассчитываются так же, как и средние из первоначальных значений признака. При этом средние, которые служат для исчисления на их основе общей средней, принимаются в качестве вариантов.

Вычисление *средней арифметической взвешенной из групповых средних* $x_{\text{гр}}$ осуществляется по формуле

$$\bar{X}_{\text{ар}} = \frac{\sum x_{\text{гр}} f}{\sum f}.$$

Расчёт средней арифметической в рядах распределения. Если значения осредняемого признака заданы в виде интервалов («от – до»), т.е. интервальных рядов распределения, то при расчёте средней арифметической величины в качестве значения признаков в группах принимают середины этих интервалов, они рассчитываются как простая средняя между верхней и нижней границами каждого интервала. Если в ряду распределения есть открытые интервалы (первый и последний), их величину условно приравнивают к интервалам, примыкающим к ним. При таком исчислении средней допускается *некоторая неточность*, поскольку делается предположение о равномерности распределения признака внутри группы. Однако ошибка будет тем меньше, чем уже интервал и чем больше единиц в интервале.

Рассмотрим следующий пример (таблица 1.4). Рассчитаем среднемесячную заработную плату рабочих по формуле средней арифметической взвешенной. После того как найдены середины интервалов, варианты умножают на частоты (веса) и сумму произведений (тыс. д. е.) делят на сумму частот (весов):

$$\bar{X}_{\text{ар}} = \frac{\sum xf}{\sum f} = \frac{7290}{100} = 72{,}9 \text{ тыс. д. е.}$$

Таблица 4 − **Распределение рабочих по уровню ежемесячной оплаты труда**

Группы рабочих по оплате труда, тыс. д. е.	Число рабочих f чел.	Середина интервала, тыс. д. е. x	xf
До 50	5	45	225
50−60	15	55	825
60−70	20	65	1300
70−80	30	75	2250
80−90	16	85	1360
90 и более	14	95	1330
Итого	100	−	7290

При расчетах средней арифметической взвешенной использование частот позволяет упрощать расчеты, когда частота выражена большими, многозначными числами. В этих случаях используют некоторые свойства средней арифметической.

Средняя арифметическая и ее свойства. Рассмотрим основные свойства средней арифметической.

Первое свойство. Нулевое свойство средней величины заключается в том, что сумма отклонений вариант от их средней арифметической величины равна нулю.

Первое свойство средней может быть использовано, в частности, для контроля правильности вычислений арифметической средней: если средняя вычислена правильно, сумма отклонений должна равняться нулю (практически, с учетом округлений, допускаемых при вычислении средней, − очень близка к нулю):

$$\sum_{i=1}^{k}(x_i - \overline{x})m_i = 0.$$

*Второе свойство***.** Если все варианты уменьшить (или увеличить) на одно и то же постоянное число, то средняя арифметическая из этих вариант уменьшится (или увеличится) на то же самое число:

$$\frac{\sum_{i=1}^{k}(x_i \pm A)m_i}{\sum_{i=1}^{k}m_i} = \frac{\sum_{i=1}^{k}x_i m_i}{\sum_{i=1}^{k}m_i} \pm \frac{\sum_{i=1}^{k}Am_i}{\sum_{i=1}^{k}m_i} = \overline{x} \pm A.$$

Пример. Пусть заработная плата каждого работника фирмы увеличилась за некоторый период на 15000 д. е. Тогда средняя заработная плата всех работников фирмы увеличилась также на 15000 д. е.

*Третье свойство***.** Если все варианты одинаково увеличить (или уменьшить) в одно и то же число раз, то средняя арифметическая увеличится (или уменьшится) во столько же раз.

$$\frac{\sum_{i=1}^{k}\frac{x_i}{A}m_i}{\sum_{i=1}^{k}m_i} = \frac{\frac{1}{A}\sum_{i=1}^{k}x_i m_i}{\sum_{i=1}^{k}m_i} = \frac{1}{A}\overline{x}.$$

Пример. Так, если бы заработная плата каждого работника фирмы увеличилась на 10 %, то и средняя заработная плата всех работников фирмы увеличилась бы на 10 %.

*Четвертое свойство***.** Если же все веса средней одинаково увеличить (или уменьшить) в несколько раз, средняя арифметическая не изменится. Увеличение всех весов в несколько раз приводит к тому, что во столько же одновременно увеличится и числитель, и знаменатель дроби (средней арифметической), поэтому значение дроби не изменяется.

Для упрощения расчётов средней идут по пути уменьшения значений вариантов и частот. Наибольшее упрощение достигается, когда в качестве *A* выбирается значение одного из центральных вариантов, обладающего наибольшей частотой, в качестве *i* – величина интервала. Величина *A* называется началом отсчёта, поэтому такой метод вычисления средней называется «способом отсчёта от условного нуля» или «*способом моментов*»:

$$\overline{x} = m_1 i + A, \text{ где } m_1 = \frac{\sum\left(\frac{x-A}{i}\right)f}{\sum f}.$$

Средняя гармоническая применяется в тех случаях, когда частоты (веса) не приводятся непосредственно, а входят сомножителями в один из имеющихся показателей.

Пример. Автомобиль доставил товары в три магазина фирмы, которые удалены от головного предприятия на одинаковое расстояние. Так, до первого магазина, расположенного на шоссейной дороге, автомобиль прошел путь со скоростью 50 км/ч, до второго, по проселочной дороге, − 40 км/ч, а в третьем случае автомобилю пришлось полпути пройти через лесной массив, и скорость движения составила только 30 км/ч.

Требуется определить среднюю скорость движения автомобиля. На первый взгляд представляется, что средняя скорость движения может быть определена по формуле простой арифметической:

$$\bar{V}_a = \frac{50+40+30}{3} = \frac{120}{3} = 40 \text{ км/ч.}$$

Однако нетрудно убедиться, что средняя вычислена неправильно. В самом деле, производя расчет средней скорости по простой арифметической средней, исходим из того, что автомобиль во всех трех случаях прошел одинаковое расстояние, пройдя соответственно 50, 40 и 30 км, т.е. всего 120 км. Если бы условие этой задачи было сформулировано в такой форме, то средняя была бы рассчитана правильно и характеризовала бы пройденное автомобилем среднее расстояние.

В действительности же эта средняя рассчитана неверно, так как из условия задачи не следует, что автомобиль на преодоление расстояния до трех магазинов фирмы «Весна» проехал 120 км, так как скорость движения была различная. Следовательно, он прошел и разное расстояние.

В подобных случаях нужно применить формулу средней гармонической простой (не взвешенной):

$$\bar{X}_{\text{г}} = \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \ldots + \frac{1}{x_n}}.$$

или в сокращенном виде

$$\bar{X}_{\text{г}} = \frac{n}{\text{е} \frac{1}{x}},$$

где $X_{\text{г}}$ − средняя гармоническая;

$\frac{1}{x}; \frac{1}{x_1}; \frac{1}{x_2} \ldots \frac{1}{x_n}$ − числа, обратные заданным вариантам.

Иначе говоря, **простая гармоническая средняя есть отношение числа вариант к сумме обратных значений этих вариант.**

Для нашего примера будем иметь:

$$\overline{V_{г}} = \frac{1+1+1}{\frac{1}{50}+\frac{1}{40}+\frac{1}{30}} = \frac{3}{\frac{12+15+20}{600}} = \frac{3}{\frac{47}{600}} = \frac{1800}{47} = 38 \text{ км/ч}.$$

В нашем примере средняя арифметическая $(\bar{V}_a)$ оказалась больше средней гармонической $V_{г}$, при этом абсолютная ошибка завышения составляет 2 км/ч (38 – 40), а относительная $\left(\frac{2 \cdot 100}{40}\right)$ – 5 %.

Таким образом, неправильное использование арифметической средней привело бы к завышению средней скорости движения автомобиля, что может привести к неправильному определению объема перевозок.

Средняя геометрическая применяется в тех случаях, когда индивидуальные значения признака представляют собой, как правило, относительные величины динамики, полученные как произведение цепных коэффициентов роста. Средняя геометрическая рассчитывается как корень степени n из произведений вариантов признака x:

$$\bar{x} = \sqrt[n]{\prod_{i=1}^{n} x_i} = \sqrt[n]{x_1 x_2 x_3 \ldots x_n},$$

где n – число вариант; П – знак произведения.

Среднеквадратическая является разновидностью средней геометрической**.**

Средняя хронологическая − это средний уровень ряда динамики, т.е. средняя, исчисленная по совокупности значений показателя в разные моменты или периоды времени. В зависимости от вида ряда динамики применяются различные способы ее расчета, а именно расчет: средней хронологической интервального ряда, средней хронологической моментного ряда.

Средней хронологической интервального ряда является средняя величина из уровней интервального ряда динамики:

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

где $\bar{y}$ − средний уровень ряда;

y − уровень ряда динамики;

n − число членов ряда.

Средней хронологической моментного ряда является средняя величина из уровней моментного ряда динамики.

При равных промежутках времени между датами, на которые имеются данные, и равномерном изменении размера показателя между датами средняя хронологическая моментного ряда обычно исчисляется по формуле

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

Если периоды времени, отделяющие одну дату от другой, не равны между собой, то расчет средней хронологической моментного ряда производится по формуле средней взвешенной арифметической, в качестве весов которой принимаются отрезки времени между датами:

$$\bar{y} = \frac{\sum yT}{\sum T},$$

где T − время, в течение которого данный уровень ряда (y) оставался без изменения.

Пример. По состоянию на 01.01.2012 года численность работников локомотивного депо составила 1220 человек. 5 января на работу было принято 2 человека, 16 января уволено 3 человека, 29 января принято 4 человека. Среднесписочная численность работников за январь месяц составит 1221 человек.

$$\bar{y} = \frac{\sum yt}{\sum t} = \frac{1220 \cdot 4 + 1222 \cdot 11 + 1219 \cdot 13 + 1223 \cdot 3}{31} = 1221 \text{ чел.}$$

Структурные средние. Особым видом средних величин являются *структурные средние*. Они применяются для изучения внутреннего строения и структуры рядов распределения значений признака. К таким показателям относятся мода и медиана.

Мода и медиана. *Мода* (M_0) – это варианта, у которой частота (вес) в ряду распределения наибольшая.

Модальная величина в дискретном ряду – это величина признака, чаще всего встречающаяся в распределенной совокупности единиц.

Вычисление моды для интервальных рядов с равными интервалами производится по формуле

$$M_0 = M_0 = x_0 + r_i \frac{m_2 - m_1}{(m_2 - m_1) + (m_2 - m_3)},$$

где x_0 – начало модального интервала;

r_i – величина интервала;

m_2 – частота модального интервала;

m_1 – частота интервала, предшествующего модальному;

m_3 – частота интервала, следующего за модальным.

Медиана (M_e) – варианта, находящаяся в середине ранжированного (упорядоченного) ряда распределения. Для ее определения достаточно расположить в порядке возрастания или убывания все варианты. Серединная варианта и будет являться медианой. Медиана делит ряд на две равные части: со значениями признака меньше медианы и со значениями признака больше медианы. Если число значений в распределённом ряду чётное, то медиана равна средней из двух вариант, находящихся в середине ряда.

Медиана интервального ряда рассчитывается по формуле

$$M = x_{Me} + r_i \frac{\frac{\sum m}{2} - m_n}{m_{Me}},$$

где x_{Me} – начало (нижняя граница) медианного интервала;

r_i – величина интервала;

$\sum m$ – сумма накопленных частот ряда;

m_n – накопленная частота вариант, предшествующих медианному;

m_{Me} – частота медианного интервала.

Соотношение моды, медианы и средней арифметической указывает на характер распределения признака в совокупности, позволяет его асимметрию. Если $M_0 < M_e < \bar{x}$, то имеет место правосторонняя асимметрия, при $\bar{x} < M_e < M_0$ следует сделать вывод о левосторонней асимметрии ряда.

1.6.3 Показатели вариации

Вариация – это различие в значениях какого-либо признака у разных единиц данной совокупности в один и тот же период или момент времени. Измерение вариации, выяснение её причин, выявление влияния отдельных факторов даёт важную информацию для принятия научно обоснованных управленческих решений.

Средняя величина даёт обобщающую характеристику признака изучаемой совокупности, но она не раскрывает строения совокупности, которое весьма существенно для её познания. Средняя не показывает, как располагаются около неё варианты осредняемого признака, сосредоточены ли они вблизи средней или значительно отклоняются от неё. Поэтому возникает необходимость измерять вариацию признака в совокупностях. Для этой цели в статистике применяют ряд обобщающих показателей.

К показателям вариации относятся: размах вариации, среднее линейное отклонение, дисперсия, среднее квадратическое отклонение, коэффициент вариации.

Размах вариации R представляет собой разность между максимальным $x_{\max}$ и минимальным $x_{\min}$ значениями признака:

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

Среднее линейное отклонение:

− для не сгруппированных значений

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

где n – число членов ряда;

− для сгруппированных значений,

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

где Σf – сумма частот вариационного ряда.

Дисперсия признака представляет собой средний квадрат отклонений вариантов от их средней величины, она вычисляется по формулам:

− для несгруппированных значений $\sigma^2 = \frac{\sum (x_i - \bar{x})^2}{n}$;

− для сгруппированных значений $\sigma^2 = \frac{\sum (x_i - \bar{x})^2 f_i}{\sum f_i}$.

Расчёты по этим формулам можно упростить, если использовать **свойства дисперсии:**

первое – если все значения признака уменьшить или увеличить на одну и ту же величину А, то дисперсия от этого не изменится;

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

Используя второе свойство дисперсии, разделив все варианты на величину интервала, получим следующую формулу вычисления дисперсии в вариационных рядах с равными интервалами по способу моментов:

$$\sigma^2 = l_{ср} = i^2 (m_2 - m_1^2),$$

где i – величина интервала;

m_2 – момент второго порядка,

$$m_2 = \frac{\sum \left(\frac{\bar{x} - c}{\sum f_i} \right)^2}{\sum f_i};$$

m_1 – квадрат момента первого порядка,

$$m_1 = \frac{\sum\left(\frac{\overline{x}-c}{i}\right)^2}{\sum f_i}.$$

Среднее квадратическое отклонение:

– для несгруппированных значений $\sigma = \frac{\sum(x_i - \overline{x})^2}{n}$;

– для сгруппированных значений $\sigma = \frac{\sum(x_i - \overline{x})^2 f_i}{\sum f_i}$.

Среднее квадратическое отклонение – это обобщающая характеристика размеров вариации признака в совокупности; оно показывает, насколько в среднем отклоняются конкретные варианты от их среднего значения. Среднее квадратическое отклонение является абсолютной мерой колеблемости признака и выражается в тех же единицах, что и варианты.

Дисперсия альтернативного признака. Среди варьирующих признаков встречаются такие, вариация которых проявляется в том, что у одних единиц совокупности они встречаются, а у других нет. Обозначим: 1 – наличие интересующего нас признака; 0 – его отсутствие; p – доля единиц, обладающих данным признаком; q – доля единиц, не обладающих данным признаком: $p + q = 1$. Отсюда $q = 1 – p$.

Исчислим среднее значение альтернативного признака и его дисперсию:

– среднее значение альтернативного признака $\overline{x} = \frac{1p + 0q}{p+q} = p,$

– дисперсия альтернативного признака

$$\sigma_p^2 = \frac{(1-p)^2 p + (0-p)^2 q}{p+q} = q^2 p + p^2 q = pq(q+p) = pq.$$

Чем меньше значение дисперсии и среднего квадратического отклонения, тем однороднее совокупность и тем более типичной будет средняя величина.

В статистической практике часто возникает необходимость сравнения вариаций различных признаков, например, возраста рабочих и их квалификации. Для подобных сопоставлений показатели абсолютной колеблемости признаков непригодны. В этих случаях используют относительный показатель вариации – коэффициент вариации.

Коэффициент вариации представляет собой выраженное в процентах отношение среднего квадратического отклонения к средней арифметической:

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

Коэффициент вариации используют не только для сравнительной оценки вариации единиц совокупности, но и как характеристику однородности совокупности. Совокупность считается количественно однородной, если коэффициент вариации не превышает 33 %.

Правило сложения дисперсий. Вариация признака обусловлена различными факторами, некоторые из этих факторов можно выделить, если статистическую совокупность разбить на группы по какому-либо признаку. Тогда, наряду с изучением вариации признака по всей совокупности в целом, становится возможным изучить вариацию для каждой из составляющих её групп, а также и между этими группами. Если совокупность расчленена на группы по одному фактору, изучение вариации осуществляется посредством исчисления и анализа трёх видов дисперсий: *общей, межгрупповой и внутригрупповой.*

Общая дисперсия σ^2 измеряет вариацию признака по всей совокупности под влиянием всех факторов, обусловивших эту вариацию:

– для несгруппированных значений $\sigma^2 = \dfrac{\sum (x_i - \bar{x})^2}{n}$;

– для сгруппированных значений $\sigma^2 = \dfrac{\sum (x_i - \bar{x})^2 f_i}{\sum f_i}$.

Межгрупповая дисперсия δ^2 характеризует систематическую вариацию результативного порядка, обусловленную влиянием признака-фактора, положенного в основание – группировки. Она равна среднему квадрату отклонений групповых (частных) средних x_i от общей средней:

$$\delta^2 = \frac{\sum (\bar{x}_i - \bar{x})^2}{n}; \quad \delta^2 = \frac{\sum (\bar{x}_i - \bar{x})^2 f_i}{\sum f_i}.$$

Внутригрупповая (частная) дисперсия σ_i^2 отражает случайную вариацию, т. е. часть вариации, обусловленную влиянием неучтённых факторов и не зависящую от признака-фактора, положенного в основание группировки. Она равна среднему квадрату отклонений отдельных значений признака внутри группы x от средней арифметической этой группы и может быть исчислена как простая дисперсия или как взвешенная дисперсия по формулам:

$$\sigma_i^2 = \frac{\sum (x_i - \bar{x})^2}{n}; \quad \sigma_i^2 = \frac{\sum (x_i - \bar{x})^2 f_i}{\sum f_i}.$$

На величину частных, внутригрупповых дисперсий не влияет групповой признак. Поэтому, чтобы получить представление об общей вариации признака под влиянием случайных причин, следует рассчитать среднюю из внутригрупповых или частных дисперсий:

$$\bar{\sigma}^2 = \frac{\sum \sigma_i^2}{n}; \qquad \bar{\sigma}^2 = \frac{\sum \sigma_i^2 f_i}{\sum f_i}.$$

Согласно *правилу сложения дисперсий* общая дисперсия равна сумме средней из внутригрупповых и межгрупповых дисперсий:

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

1.7 Выборочное наблюдение

1.7.1 Основы выборочного наблюдения

Выборочное наблюдение является одним из видов несплошного статистического наблюдения, при котором наблюдению подвергается не вся совокупность единиц, а только часть их, отобранная на основе определенных научных принципов. При этом данные, полученные на основе отобранной части совокупности, распространяют на всю генеральную совокупность. Наблюдение организуется таким образом, что эта часть отобранных единиц представляет всю совокупность.

Выборочное наблюдение находит широкое применение во всех отраслях хозяйственной деятельности. Выборочным порядком выявляется покупательский спрос, проверяются нормы естественной убыли товаров и т.д. Применение выборочного метода часто является необходимым в тех случаях, когда изучение качества объекта ведет к его порче или полному уничтожению.

Вся изучаемая совокупность, из которой производится отбор некоторого числа единиц для выборочного наблюдения, называется **генеральной совокупностью.**

Часть генеральной совокупности, подлежащая выборочному обследованию, называется **выборочной совокупностью.**

Численность (объем) генеральной совокупности обозначим буквой N, а численность выборочной совокупности обозначим буквой n. При выборочном наблюдении обычно ставят две задачи: определение среднего размера изучаемого признака и определение доли изучаемого признака в данной совокупности. Основная задача выборочного наблюдения состоит в том, чтобы на основе характеристик выборочной совокупности (средней и доли) получить достоверные суждения о показателях средней и доли в генеральной совокупности. При этом следует иметь в виду, что при любых статистических исследованиях (сплошных и выборочных) возникают ошибки двух видов: регистрации и репрезентативности.

Ошибки регистрации могут иметь *случайный* (непреднамеренный) и *систематический* (тенденциозный) характер. *Случайные ошибки* обычно уравновешивают друг друга, поскольку не имеют преимущественного направления в сторону преувеличения или преуменьшения значения изучаемого показателя. Систематические ошибки направлены в одну сторону вследствие преднамеренного нарушения правил отбора. Их можно избежать при правильной организации и проведении наблюдения.

Ошибки репрезентативности присущи только выборочному наблюдению и возникают в силу того, что выборочная совокупность не полностью воспроизводит генеральную. Они представляют собой расхождение между значениями показателей, полученных при выборке, и значениями показателей этих же величин, которые были бы получены при проведённом с одинаковой степенью точности сплошном наблюдении, т.е. между величинами выборочных и соответствующих генеральных показателей.

Исчисленные обобщающие характеристики в генеральной совокупности называются генеральными: $\bar{x}$ – генеральная средняя, a – генеральное среднее квадратическое отклонение, P – генеральная доля, полученная как отношение числа M-единиц, обладающих данным признаком, ко всей численности N генеральной совокупности:

$$P = \frac{M}{N}.$$

Исчисленные обобщающие характеристики в выборочной совокупности называются выборочными: $\tilde{x}$ – выборочная средняя, σ – выборочное среднее квадратическое отклонение, w – выборочная доля (частость) – отношение числа m единиц выборочной совокупности, обладающих данным признаком, ко всей численности n выборочной совокупности, т.е.

$$w = \frac{m}{n}.$$

Пример. В таблице 1.5 приведены данные испытания крепости нити. Требуется определить среднее квадратическое отклонение для выборочной совокупности и оценить отклонение выборочной средней от генеральной средней. Для облегчения расчётов разделим величину отклонения на число, кратное величине интервала К.

Средняя выборочная (средняя крепость нити выборочной совокупности) равна 183,6 грамма:

$$\left(\tilde{x} = \frac{\sum xf}{\sum f} = \frac{18360}{100} = 183,6 \right).$$

Таблица 1.5 – **Данные испытания 100 одиночных нитей на крепость**

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

Среднее квадратическое отклонение равно 43,0 граммам:

$$\left(\sigma=\sqrt{\sigma_0^2-(\bar{x}-x_0)^2}=\sqrt{\frac{\left(\frac{x-x_0}{k}\right)^2 fk^2}{\sum f}-\left(\bar{x}-x_0\right)^2}=\sqrt{1888{,}0-40{,}96}=\sqrt{1847{,}04}=43{,}0\right).$$

Поскольку проведение сплошного наблюдения заменяется выборочным, а исчисление средней генеральной заменяется исчислением средней выборочной, важно установить, насколько полученная средняя выборочная является характерной для данной генеральной совокупности и представляет ли она среднюю генеральную.

Другими словами, необходимо установить, как велико отклонение Δ средней выборочной от средней генеральной $\bar{x}$, т.е. $\bar{x}=\tilde{x}\pm\Delta$.

Чем меньше величина отклонения Δ, тем точнее выборочная средняя воспроизводит генеральную среднюю. Величина этого отклонения и определяет степень точности выборочного наблюдения.

1.7.2 Ошибки выборки

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

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

Теория и практика применения выборочного метода показали, что данные выборочного наблюдения достаточно достоверны, так как выборочный метод базируется на применении закона больших чисел и теории вероятности. *Сущность закона больших чисел заключается в том, что чем больше будет взято единиц наблюдения, тем точнее средняя выборочная будет воспроизводить среднюю генеральную.* Теория выборочного метода дает формулу, по которой можно вычислить среднюю величину ошибки μ для выборочной совокупности, отобранной в случайном порядке, т.е. таким образом, что каждая единица генеральной совокупности имела бы равную возможность попасть в это число:

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

где μ – средняя ошибка выборки;

σ – среднее квадратическое отклонение;

n – численность выборочной совокупности.

Применяя эту формулу, получим следующую величину средней ошибки выборки для нашего примера:

$$\mu = \frac{\sigma}{\sqrt{n}} = \frac{43,0}{10} = \pm 4,30.$$

Величина средней ошибки выборки зависит, прежде всего, от показателей колеблемости значений признаков в выборочной совокупности. Степень колеблемости значений признаков определяется средним квадратическим отклонением σ.

Чем меньше величина среднего квадратического отклонения, тем меньше величина средней ошибки при той же численности выборки.

Кроме того, величина средней ошибки зависит от численности выборки. Увеличивая или уменьшая объем выборки n, можно регулировать величину ошибки μ. Чем больше единиц будет охвачено выборочным наблюдением, тем меньше будет величина ошибки, так как тем точнее будет представлена генеральная совокупность. Полученная величина ошибки μ характеризует среднее отклонение средней выборочной от средней генеральной.

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

Ошибка выборки, исчисленная с заданной степенью вероятности, представляет предельную ошибку выборки.

Если через Δ обозначим предельную ошибку, частное от деления Δ на μ приравняем к t, тогда можно записать $\frac{\Delta}{\mu} = t$, отсюда $\Delta = \mu t$, а так как $\mu = \frac{\sigma}{\sqrt{n}}$, то $\Delta = \frac{\sigma t}{\sqrt{n}}$.

Следовательно, величина предельной ошибки зависит от величины средней ошибки и коэффициента t. Коэффициент зависит от степени вероятности, с которой производится выборочное наблюдение. При вероятности 0,683 значение $t = 1$, при вероятности 0,954 значение $t = 2$, при вероятности 0,997 значение $t = 3$.

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

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

Указывая вероятные пределы случайной ошибки выборки, мы тем самым указываем и те пределы, за которые не выйдет характеристика генеральной совокупности.

Определим для нашего примера, в каких границах должна заключаться средняя крепость нити в генеральной совокупности, с вероятностью 0,997.

Средняя ошибка равна $\frac{\sigma}{\sqrt{n}}$, т.е. $\frac{43,0}{10} = \pm 4,30$ г.

Предельная ошибка при заданной степени вероятности А равна μ 3, т.е. $\pm 4,30 \cdot 3 = \pm 12,9$.

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

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

т.е. при сокращении ошибки вдвое численность выборки должна быть увеличена в четыре раза, при сокращении втрое объем выборки должен быть увеличен в девять раз и т.д. Следовательно, чтобы получить среднюю ошибку выборки для нашего примера, равную 2,15, нужно подвергнуть наблюдению не 100, а 400 нитей.

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

$$\mu = \sqrt{\frac{P(1-p)}{n}},$$

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

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

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

Допустим, что нужно установить для нашего примера долю нитей, имеющих крепость 190 граммов и больше. Частость (доля данного признака в выборочной совокупности) равна 0,56 $\left(\frac{56}{100}\right) = 0{,}56.$

Отсюда средняя ошибка для доли

$$\mu = \sqrt{\frac{0{,}56(1-0{,}56)}{100}} = \sqrt{0{,}002464} = \pm 0{,}04956.$$

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

$$\Delta p = t\sqrt{\frac{\omega(1-\omega)}{n}} = \pm 0{,}04956 \cdot 3 = \pm 0{,}14868.$$

Пределы генеральной доли определяем по формуле

$$P = \omega \pm \Delta p.$$

Отсюда P = 0,56 ±0,14868.

1.7.3 Виды отбора единиц в выборочную совокупность

Для каждого конкретного выборочного наблюдения значение ошибки репрезентативности может быть определено по соответствующим формулам, которые зависят от вида, метода и способа формирования выборочной совокупности.

По виду различают индивидуальный, групповой и комбинированный отбор. При *индивидуальном отборе* в выборочную совокупность отбираются отдельные единицы генеральной совокупности, при *групповом отборе* – качественно однородные группы или серии изучаемых единиц; *комбинированный отбор* предполагает сочетание первого и второго видов.

Случайный отбор. Теория выборочного наблюдения прежде всего указывает на необходимость осуществления случайного отбора. Принцип случайного отбора состоит в том, что единицы для наблюдения отбираются из всей совокупности. При этом каждая единица генеральной совокупности имеет равную возможность попасть в выборочную совокупность.

По *методу* отбора различают *повторную* и *бесповторную выборки*.

При *повторной выборке* общая численность единиц генеральной совокупности в процессе выборки остаётся неизменной. Ту или иную единицу, попавшую в выборку, после регистрации снова возвращают в генеральную совокупность, и она сохраняет равную возможность со всеми прочими единицами при повторном отборе единиц вновь попасть в выборку. Такая выборка встречается редко. Чаще пользуются бесповторной выборкой.

Бесповторным называется такой отбор, когда отобранная единица не возвращается обратно в генеральную совокупность. Следовательно, численность генеральной совокупности в каждой отобранной единице сокращается.

Способом отбора определяется конкретный механизм или процедура выборки единиц из генеральной совокупности. По степени охвата единиц совокупности различают *большие* и *малые* (менее 30 наблюдений) выборки.

В практике выборочных исследований наибольшее распространение получили следующие виды выборки: собственно-случайная, механическая, типическая, серийная, комбинированная.

Собственно-случайный отбор представляет собой отбор единиц из всей генеральной совокупности посредством жеребьёвки или какого-либо иного подобного способа, например, с помощью таблицы случайных чисел.

Механический отбор. При механическом отборе также применяется принцип случайного отбора. При этом из генеральной совокупности отбирается определенное число единиц через определенный интервал. При таком способе отбора генеральную совокупность механически разбивают на равные группы, число которых равно численности выборочной совокупности.

Если при случайном отборе возникает лишь возможность попадания в выборку представителей всех тех состояний, которыми характеризуется изучаемый признак общей совокупности, то механический отбор направлен на то, чтобы действительно обеспечить попадание в выборку таких представителей. При механическом отборе средняя ошибка выборки определяется по тем же формулам, как и при повторном случайном отборе.

Типический отбор используется в тех случаях, когда все единицы генеральной совокупности можно разбить на несколько типических групп. Типический отбор предполагает выборку единиц из каждой типической группы собственно-случайным или механическим способом.

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

Отобранное по каждой группе количество единиц является частной выборочной совокупностью n_i. Для каждой такой выборочной совокупности можно установить средний размер изучаемого признака $\overline{x_i}$ и среднее квадратическое отклонение $\overline{\sigma_i^2}$, которое характеризует внутригрупповую колеблемость признака в пределах своей группы. Этот показатель можно обобщить для всей совокупности в целом, т.е. найти показатель внутригрупповой колеблемости признака для всех групп совокупности, вместе взятых, $\overline{\sigma_i^2}$.

Чтобы получить общую выборочную среднюю для всех обследованных групп $\tilde{x}$, надо из частных выборочных средних $\bar{x}$ вывести среднюю арифметическую взвешенную, причем в качестве весов можно взять или общую численность каждой группы, или численность выборки в каждой группе. Результат будет одинаковым, так как количество обследуемых единиц распределяется по группам пропорционально их доле в общей совокупности.

Ошибка выборки при типическом отборе определяется по той же формуле, что и при случайном отборе, однако вместо общей дисперсии признака σ^2 в этой формуле участвует средняя дисперсия из внутригрупповых σ_i^2.

Серийная выборка. Иногда в практике выборочного наблюдения производят отбор целых групп единиц и внутри отобранных групп подвергают наблюдению все единицы без исключения. Для отбора серий применяют либо случайную выборку, либо механический отбор. Такая выборка называется *серийной.*

Серийный отбор имеет большое практическое значение, так как легче организовать отбор и изучение нескольких серий единиц, чем сотен отдельных единиц. Но серийный отбор оказывается менее точным в смысле репрезентативности изучаемых показателей, чем другие способы отбора.

Средняя ошибка серийной выборки исчисляется по формулам:

− при *повторном отборе* серий

для средней $\mu = \sqrt{\dfrac{{\sigma_S}^2}{S}}$; для доли $\mu = \sqrt{\dfrac{\omega_S(1-\omega_S)}{S}}$;

− при *бесповторном отборе* серий

для средней $\mu = \sqrt{\dfrac{\sigma_S^2}{S}\left(1-\dfrac{S}{R}\right)}$; для доли $\mu = \sqrt{\dfrac{\omega_S(1-\omega_S)}{S}\left(1-\dfrac{S}{R}\right)}$,

где ${\sigma_S}^2$ – межсерийный квадрат отклонений;

R – число серий в генеральной совокупности;

S – число отобранных серий;

ω_S – доля данного признака в среднем по всем обследованным сериям.

1.7.4 Определение необходимой численности выборки

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

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

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

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

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

$$\Delta_x^2 = \frac{t^2\sigma^2}{n}\text{, откуда } \Delta_x^2 n = t^2\sigma^2\text{, а } \ n = \frac{t^2\sigma^2}{\Delta_x^2}.$$

Объем выработки при исчислении доли определяется по этой же формуле, только вместо σ^2 берется $\omega(1-\omega)$, т.е. $n = \dfrac{t^2\omega(1-\omega)}{\Delta_x^2\omega}$.

При бесповторном отборе численность выборки

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

Возведя обе части уравнения в квадрат, получим:

$$\Delta_x^2 = \frac{t^2\sigma^2}{n}\cdot\frac{N-n}{N}.$$

После преобразований имеем:

$$\Delta_x^2 nN + t^2\sigma^2 n = t^2\sigma^2 N;\ \ n(\Delta_x^2 N + t^2\sigma^2) = t^2\sigma^2 N.$$

Отсюда $n = \dfrac{t^2\sigma^2 N}{\Delta_x^2 N + t^2\sigma^2}$.

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

Пример. Объём генеральной совокупности составляет 2500 единиц. Требуется определить необходимый объем собственно-случайной выборки для повторного и бесповторного отборов, для того чтобы с вероятностью 0,954 предельная ошибка выборки при определении средней не превышала 20 единиц при среднем квадратическом отклонении 300 единиц.

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

Для бесповторного отбора

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

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

Так же исчисляется объем выборки и при определении доли. Определим по данным нашего примера, сколько нужно отобрать единиц для выборочного наблюдения, чтобы ошибка доли с вероятностью 0,954 не превышала 3 % ($\Delta\omega$ = 0,03) при удельном весе рассматриваемых единиц в выборке, равном 80 % (ω = 0,8), N = 2500.

Для повторного отбора

$$n = \frac{t^2\omega(1-\omega)}{\Delta_x^2} = \frac{2^2 \cdot 0,8 \cdot 0,2}{0,03^2} = \frac{6400}{9} = 711 \text{ ед.}$$

Для бесповторного отбора

$$n = \frac{t^2\omega(1-\omega)N}{\Delta^2\omega N + t^2\omega(1-\omega)} = \frac{2^2 \cdot 0,8 \cdot 0,2 \cdot 2500}{0,03^2 \cdot 2500 + 2^2 \cdot 0,8 \cdot 0,2} = \frac{1600}{2,25 + 0,64} = 554 \text{ ед.}$$

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

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

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

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

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

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

Подставляем эти данные в формулу:

$$n = \frac{2^2 \cdot 0{,}08 \cdot 0{,}92 \cdot 5000}{0{,}02^2 \cdot 5000 + 2^2 \cdot 0{,}08 \cdot 0{,}92} = \frac{4 \cdot 0{,}0736 \cdot 5000}{0{,}0004 \cdot 5000 + 4 \cdot 0{,}0736} = \frac{1472}{22944} \approx 642 \text{ дет.}$$

Чтобы с вероятностью 0,954 можно было утверждать, что предельная ошибка доли брака не превышает 2 %, необходимо из 5000 деталей отобрать 642.

1.7.5 Способы распространения выборочных результатов на генеральную совокупность

Конечной целью выборочного наблюдения является характеристика генеральной совокупности на основе выборочных результатов. Распространение выборочных результатов на генеральную совокупность осуществляется разными способами. Обычно применяются *способ прямого пересчета и способ коэффициентов.*

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

Например, необходимо определить средний процент брака в партии деталей, состоящей из 10000 деталей. Для выборочного наблюдения в случайном порядке было отобрано 900 деталей. Анализ качества отобранных деталей показал, что средний процент брака в данной совокупности составил 1,5 %. Среднее квадратическое отклонение равно 0,3 %. Максимальная ошибка выборочного наблюдения с вероятностью 0,997 равна 0,3 %. Таким образом, средний процент брака в генеральной совокупности находится в пределах 1,5 ± 0,3 %, т.е. колеблется от 1,2 до 1,8 %.

Имея данные об общей величине партии, определяем общее количество бракованных деталей, которое будет колебаться в пределах 1,8–1,2 % от 10000, т.е. 180–120 единиц. Можно пределы не указывать, а пользоваться средней выборочной как генеральной средней. Тогда среднее количество бракованных банок в генеральной совокупности составит 1,5 % от 10000, т.е. 150 единиц.

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

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

1.8 Ряды динамики

1.8.1 Понятие рядов динамики

Одной из важнейших задач статистики является изучение изменений анализируемых показателей во времени, т.е. их динамика. Эта задача решается при помощи анализа рядов динамики (или временных рядов).

Рядом динамики называется ряд последовательно расположенных в хронологическом порядке статистических показателей, характеризующих изменение общественных явлений во времени.

Каждый ряд динамики состоит из даты времени, периода времени и статистических данных, которые называются *уровнями ряда динамики.* Выявление основной тенденции в изменении уровней, именуемой *трендом*, является одной из главных задач анализа рядов динамики.

Ряды динамики делят на ряды динамики абсолютных величин и ряды динамики производных величин. Ряды динамики абсолютных величин подразделяются на моментный и интервальный ряды динамики.

Моментный ряд динамики показывает состояние каких-либо явлений на определенный момент времени. Например, на начало, конец года, квартала, месяца.

Интервальный ряд динамики показывает статистические данные, т.е. цифровые данные, характеризующие размеры явлений за определенный промежуток времени (за ряд месяцев, лет и т.д.).

Особенностью интервальных рядов динамики является то, что итоги, полученные в результате суммирования составляющих их данных, имеют вполне реальное содержание. В отличие от моментных рядов динамики, интервальные ряды обладают следующим свойством: их *уровни можно складывать*. Уровни моментного ряда при своем сложении не дают новых уровней, т.е. их *суммировать нельзя*, так как явления, выраженные моментными рядами, получают не сплошную, а прерывистую характеристику.

Если уровни ряда представляют собой не непосредственно наблюдаемые значения, а производные величины: средние или относительные, то такие ряды называются *производными.*

Относительный ряд динамики – это ряд цифровых данных, характеризующих изменение относительных размеров либо экономических, либо социальных явлений (например, динамика доли городского и сельского населения в процентах).

Ряд динамики средних величин показывает изменение средних размеров признаков общественно-экономических явлений во времени, например, среднего уровня доходов населения.

По расстоянию между уровнями ряды динамики бывают *с равностоящими и неравностоящими уровнями во времени.*

Ряды динамики могут быть изображены графически. *Графическое изображение* позволяет наглядно представить развитие явления во времени и способствует проведению анализа изменения уровней рассматриваемого ряда. Наиболее распространённым видом графического изображения для аналитических целей является *линейная диаграмма*, которая строится в прямоугольной системе координат: на оси абсцисс откладывается время, а на оси ординат – уровни ряда.

Наряду с линейной диаграммой для графического изображения рядов динамики используется *столбиковая диаграмма*.

1.8.2 Правила построения рядов динамики

При построении динамических рядов необходимо соблюдать определённые правила: основным условием для получения правильных выводов при анализе рядов динамики и прогнозировании его уровней является сопоставимость уровней динамического ряда между собой.

Статистические данные должны быть сопоставимы по территории, кругу охватываемых объектов, единицам измерения, времени регистрации, ценам, методологии расчёта и др.

Сопоставимость по территории предполагает сравнение одноимённых показателей в границах одной и той же территории, например, динамика экономической мощи страны, изменение границ влияет на численность населения, объём производства продукции.

Сопоставимость по кругу охватываемых объектов означает сравнение совокупностей с равным числом элементов. При этом нужно иметь в виду, что сопоставляемые показатели динамического ряда должны быть однородны по экономическому содержанию и границам объекта. Например, при характеристике динамики численности студентов высших учебных заведений по годам нельзя в одни годы учитывать только численность студентов дневного обучения, а в другие – численность студентов всех видов обучения. Несопоставимость может возникнуть вследствие перехода ряда объектов из одного подчинения в другое.

Сопоставимость по времени регистрации для интервальных рядов обеспечивается равенством периодов времени, за которые приводятся данные. Нельзя, например, при изучении ритмичности работы предприятия сравнивать данные о доли продукции по декадам, так как число рабочих дней по отдельным декадам может быть разным, что приводит к различию в объёме выпуска продукции.

Для моментных рядов динамики показатели следует приводить на одну и ту же дату, например на 1-е число каждого месяца.

При приведении к сопоставимому виду продукции, измеренной в стоимостном выражении, необходимо учитывать непрерывное *изменение цен* и существование разных видов цен. Количество продукции, произведенное в разные периоды, оценивают в ценах одного и того же базисного периода, которые называют *неизменными* или *сопоставимыми*.

Рассмотренные примеры показывают, что часто приходится иметь дело с такими несопоставимыми данными, которые могут быть приведены к сопоставимому виду дополнительными расчётами. В ряде случаев несопоставимость может быть устранена путём обработки рядов динамики приёмом, который носит название *смыкание рядов динамики.* Если, например, имеются два ряда показателей, характеризующих динамику одного и того же явления в новых и старых границах по одному и тому же кругу объектов, то такие динамические ряды можно сомкнуть.

Чтобы показатели в рядах динамики (таблица 1.6) были сопоставимы, рассчитаем коэффициент соотношения уровней двух рядов по данным 2008 г.:

$$k = \frac{168}{140} = 1,20.$$

Таблица 1.6 – **Динамика объёма реализации продукции объединения в сопоставимых ценах (по годам)**

Объём реализации, млрд д. е.	2005 г.	2006 г.	2007 г.	2008 г.	2009 г.	2010 г.	2011 г.
Продукция 10 предприятий	120	125	130	140	–	–	–
Продукция 12 предприятий	–	–	–	168	180	195	215
Сопоставимый ряд	144	150	156	168	180	195	215

Умножая на этот коэффициент уровни первого ряда, получаем скорректированные данные за 2005–2007 гг. в новых границах, млрд д. е.:

$$Y_{2005} = 120 \cdot 1,2 = 144;$$
$$Y_{2006} = 125 \cdot 1,2 = 150;$$
$$Y_{2007} = 130 \cdot 1,2 = 156.$$

Смыкание рядов даёт возможность устранить несопоставимость уровней и получить представление о динамике за весь период.

1.8.3 Показатели ряда динамики

При изучении динамики общественных явлений возникает проблема описания интенсивности изменения и расчёта средних показателей динамики. Анализ интенсивности изменения во времени осуществляется с помощью показателей, получаемых при сравнении уровней. К таким показателям относят: *абсолютный прирост* (Δy)*, темп роста* ($Т_р$)*, темп прироста* ($Т_{пр}$), *абсолютное значение одного процента прироста* ($А_i$).

Система средних показателей включает *средний уровень ряда, средний абсолютный прирост, средний темп роста, средний темп прироста.*

Показатели рядов динамики могут рассчитываться двумя методами − базисным и цепным.

Абсолютный прирост ΔY_i показывает, на сколько единиц увеличился (или уменьшился) анализируемый уровень ряда Y_i относительно базисного уровня $Y_о$ (по базисной схеме) или уровень предшествующего года Y_{i-1} (по цепной схеме). Соответственно его определяют по формулам:

$$\Delta Y_i = Y_i - Y_о \quad \text{(по базисной схеме);}$$

$$\Delta Y_i = Y_i - Y_{i-1} \quad \text{(по цепной схеме).}$$

Пример. Предположим, объём погрузки на станции за I квартал (y_1) составил 2000 тыс. тонн, за II квартал (y_2) – 2200 тыс. тонн.

В нашем примере абсолютный прирост

$$\Delta y = y_2 - y_1 = 2200 - 2000 = 200 \text{ тыс. т.}$$

Цепные и базисные абсолютные приросты связаны между собой: *сумма последовательных цепных абсолютных приростов равна базисному, т.е. общему приросту за весь промежуток времени.*

Для характеристики интенсивности, т.е. относительного изменения уровня ряда за какой-либо период времени исчисляют коэффициенты роста (снижения).

При расчете *базисным* методом все уровни ряда относятся к уровню одного какого-либо периода, принятого за базу, то есть за 100 % или за единицу:

$$\frac{y_1}{y_0};\ \frac{y_2}{y_0};\ \frac{y_3}{y_0};\ \frac{y_4}{y_0};\ \dots;\ \frac{y_n}{y_0}.$$

При расчете *цепным* методом уровень каждого периода относится к уровню предыдущего периода:

$$\frac{y_1}{y_0};\ \frac{y_2}{y_1};\ \frac{y_3}{y_2};\ \frac{y_4}{y_3};\ \dots;\ \frac{y_n}{y_{n-1}}.$$

Цепные и базисные коэффициенты роста находятся во взаимосвязи. Если последовательно перемножить цепные коэффициенты роста, получим базисные коэффициенты роста.

Можно преобразовать и базисные коэффициенты роста в цепные. Для этого нужно каждый последующий коэффициент роста разделить на предыдущий базисный коэффициент роста.

Коэффициент роста ($К_р$) есть отношение данного уровня ряда к предыдущему (цепной) или к начальному (базисный). Для нашего примера

$$К_р = \frac{y_2}{y_1} = \frac{2200}{2000} = 1{,}1 \text{ раза.}$$

Коэффициентом прироста ($К_{пр}$) называется отношение абсолютного прироста к предыдущему (цепной) или начальному уровню (базисный). Коэффициент прироста рассчитывается:

цепной $К_{пр} = \frac{y_{i-1}}{y_{i-1}}$; базисный $К_{пр} = \frac{y_{i-1} - y_0}{y_0}$.

Коэффициент прироста может быть рассчитан как коэффициент роста минус единица:

$$К_{пр} = К_р - 1.$$

Для нашего примера $К_{пр}$ = 1,1 – 1,0 = 0,1 раза.

Между цепными и базисными коэффициентами роста существует взаимосвязь: *произведение последовательных цепных коэффициентов роста равно базисному коэффициенту роста за весь период, а частное от деления последующего базисного темпа роста на предыдущий равно соответствующему цепному темпу роста.*

Темпом роста $Т_р$ называется отношение данного уровня явления к предыдущему или начальному, выраженное в процентах:

$$Т_р = \frac{Y_i}{Y_0} \cdot 100 \quad \text{(по базисной схеме);}$$

$$Т_р = \frac{Y_i}{Y_{i-1}} \cdot 100 \quad \text{(по цепной схеме).}$$

Темп прироста $Т_{пр}$ показывает, на сколько процентов увеличился (или уменьшился) анализируемый уровень ряда по сравнению с базисным (по базисной схеме) или предшествующим уровнем ряда (по цепной схеме). Его определяют как отношение абсолютного прироста к уровню, принятому за базу сравнения по формулам:

$$\text{Т}_{\text{пр}} = \frac{\Delta Y_i}{Y_0} \cdot 100 \qquad \text{(по базисной схеме);}$$

$$\text{Т}_{\text{пр}} = \frac{\Delta Y_i}{Y_{i-1}} \cdot 100 \qquad \text{(по цепной схеме).}$$

Темпы роста и прироста связаны между собой, что видно из формул их расчета:

$$\text{Т}_{\text{пр}} = \frac{\Delta Y_i}{Y_{\text{о}}} \cdot 100 = \frac{Y_i - Y_{\text{о}}}{Y_{\text{о}}} \cdot 100 = \frac{Y_i}{Y_{\text{о}}} \cdot 100 - \frac{Y_{\text{о}}}{Y_{\text{о}}} \cdot 100 = \frac{Y_i}{Y_{\text{о}}} \cdot 100 - 100 = \text{Т}_{\text{р}} - 100.$$

Это дает основание определять темп прироста через темп роста:

$$\text{Т}_{\text{пр}} = \text{Т}_{\text{р}} - 100.$$

Темп роста для нашего примера

$$\text{Т}_{\text{р}} = \frac{y_2 \cdot 100}{y_1} = \frac{2200 \cdot 100}{2000} = 110\ \%.$$

Темп прироста равен темпу роста минус 100 %:

$$\text{Т}_{\text{пр}} = \text{Т}_{\text{р}} - 100 = 110 - 100 = 10\ \%.$$

Абсолютное значение одного процента прироста – это отношение абсолютного прироста к цепному темпу роста, выраженному в процентах. Оно определяется по формуле

$$\text{А}_i = \frac{\Delta Y_i}{\text{Т}_{\text{р}} \cdot 100} = \frac{Y_i - Y_{i-1}}{\frac{Y_i - Y_{i-1}}{Y_{i-1}}} = \frac{Y_{i-1}}{100}.$$

Абсолютное значение 1 % прироста так же равно абсолютному приросту, деленному на цепной темп прироста:

$$\text{А}_i = \frac{\Delta y}{\text{Т}_{\text{прц}}}.$$

Для нашего примера $\text{А}_i = \frac{200\ \text{тыс.т}}{10\ \%} = 20\,\text{тыс. т.}$

Этот показатель получается в тех же именованных числах, что и уровни ряда. В нашем примере он показывает, что каждый процент прироста составляет 20 тыс. т.

Для обобщающей характеристики динамики исследуемого явления определяют средние показатели: *средние уровни ряда* и *средние показатели изменения уровней ряда.*

Каждый ряд динамики состоит из ряда последовательных уровней. Могут быть начальные, средние и конечные уровни динамического ряда.

Рассмотрим, как рассчитываются средние уровни в моментном и интервальном рядах динамики, например, при выпуске изделий за период с 2006 по 2011 годы (таблица 1.7).

Таблица 1.7 – **Выпуск изделий на предприятии**

Показатель	Годы					
	2006	2007	2008	2009	2010	2011
Произведено изделий, тыс. ед.	2437,2	2657,0	2907,9	3144,4	3295,2	3477,9

Сколько же в среднем за год было выпущено изделий за период с 2006 по 2011 годы?

Средняя из этого интервального ряда вычисляется как средняя арифметическая простая из отдельных уровней:

$$\overline{y_{\text{а}}} = \frac{\sum y}{n} = \frac{2437,2 + 2657,0 + 2907,9 + 3144,4 + 3295,2 + 3477,9}{6} = 2986,6 \text{ тыс. изделий.}$$

Средний уровень для моментного ряда динамики, в котором промежутки между датами равны, вычисляются по формуле средней хронологической моментного ряда:

$$Y_{\text{хрон}} = \frac{\frac{1}{2}y_1 + y_2 + y_3 + ... + \frac{1}{2}y_n}{n-1},$$

где y_1 – начальный уровень ряда; y_n – конечный уровень ряда; n – число дат.

Пример. Произведем расчет среднего уровня моментного ряда. В таблице 1.8 показаны запасы материалов на складе на конец года в сопоставимых ценах.

Таблица 1.8 – **Запасы материалов на складе**

Показатель	Годы				
	2007	2008	2009	2010	2011
Запасы материалов, тыс. д. е.	26528	27567	29073	31561	35253

Среднегодовой запас за пятилетний период

$$Y_{\text{хрон}} = \frac{\frac{26528}{2} + 27567 + 29073 + 31561 + \frac{35253}{2}}{5-1} = 29772,9 \text{ тыс. д. е.}$$

Формула средней хронологической приемлема в тех случаях, если уровни ряда динамики равно отстоят друг от друга.

Если же в *моментном ряду динамики присутствуют неравные интервалы*, то применяется средняя хронологическая взвешенная:

$$y = \frac{\sum \bar{y}t}{\sum t} = \frac{\frac{y_1 + y_2}{2} t_1 + \frac{y_2 + y_3}{2} t_2 + ... + \frac{y_{n-1} + y_n}{2} t_n}{\sum t}.$$

Пример. Средняя численность работников транспортного предприятия характеризуется данными, приведенными в таблице 1.9.

Таблица 1.9 – **Средняя численность работников транспортного предприятия**

Годы	Работает на транспортном предприятии, тыс. чел.
1980	2203
1990	2802
2000	2768
2005	3136
2010	3109

Требуется определить средний уровень моментного ряда динамики средней численности работников предприятия:

$$\bar{y} = \frac{\frac{2203 + 2802}{2} \cdot 10 + \frac{2802 + 2768}{2} \cdot 10 + \frac{2768 + 3136}{2} \cdot 5 + \frac{3136 + 3109}{2} \cdot 5}{30} = 2603 \text{ тыс. чел.}$$

Средний абсолютный прирост (убыль) представляет собой обобщённую характеристику индивидуальных абсолютных приростов ряда динамики:

$$\Delta\bar{y}_{ц} = \frac{\sum y_{ц}}{n},$$

где n – число цепных абсолютных приростов ($\Delta y_{ц}$) в изучаемом периоде.

При анализе развития явления часто возникает потребность дать обобщающую характеристику интенсивности развития за длительный период времени, для этого исчисляют среднегодовые темпы роста.

Средний темп роста $\bar{Т}_{р}$ и *средний темп прироста* $\bar{Т}_{пр}$ характеризуют соответственно темпы роста и прироста за период в целом. Средний темп роста рассчитывается по данным ряда динамики по формуле средней геометрической:

$$\bar{T}_{р} = \sqrt[n-1]{\frac{Y_2}{Y_1} \cdot \frac{Y_3}{Y_2} \cdot \frac{Y_4}{Y_3} \; ... \; \frac{Y_{n-1}}{Y_{n-2}} \cdot \frac{Y_n}{Y_{n-1}}} = \sqrt[n-1]{\frac{Y_n}{Y_1}} \cdot 100,$$

где $n-1$ – количество цепных коэффициентов роста.

Исходя из соотношения темпов роста и прироста определяется средний темп прироста:

$$Т_{пр} = Т_р - 100.$$

Абсолютное значение одного процента прироста – это отношение абсолютного прироста к цепному темпу роста, выраженному в процентах. Оно определяется по формуле

$$А_i = \frac{\Delta Y_\text{i}}{Т_р \cdot 100} = \frac{Y_i - Y_{i-1}}{\frac{Y_i - Y_{i-1}}{Y_{i-1}} \cdot 100} = \frac{Y_{i-1}}{100}.$$

Если уровни ряда динамики снижаются, то средний темп роста будет меньше 100 %, а средний темп прироста – отрицательной величиной. Отрицательный темп прироста представляет собой средний темп сокращения уровня ряда динамики.

Сравнительные характеристики направления и интенсивности роста одновременно развивающихся во времени явлений определяются *приведением рядов динамики к общему (единому) основанию и расчётом коэффициентов опережения (отставания).*

По исходным уровням нескольких рядов динамики определяют *базисные темпы роста* или *прироста.* Принятый при этом за базу сравнения период времени (дата) выступает в качестве постоянной базы расчётов темпов роста для каждого из изучаемых рядов динамики. В зависимости от целей исследования базой может быть начальный, средний или конечный уровни ряда.

В таблице 1.10 приведены данные об объёмах производства продукции машиностроения в России и Беларуси. По данным таблицы 1.10 видно снижение объёмов производства продукции машиностроения как в России, так и в Беларуси (данные условные).

Таблица 1.10 − **Динамика объёмов производства продукции машиностроения**

Страна/ показатель	Годы					
	2007	2008	2009	2010	2011	2012
Россия, объём производства, млн у.е.	168 413	151 572	128 988	108 865	75 335	68 027
Базисный темп изменения,%	100	90	77	65	45	40
Беларусь, объём производства, млн у.е.	14 272	15 000	13 680	13 666	11 739	9 110
Базисный темп изменения,%	100	105	96	95,8	82	64

Однако непосредственно по ним нельзя определить, в какой стране это снижение идёт быстрее, так как различны значения абсолютных уровней этих рядов. Приведём абсолютные уровни рядов к одному основанию, приняв за базу сравнения уровни 2007 г., и получим сравнимые показатели –

базисные темпы изменения, которые показывают, что темпы снижения объёмов производства продукции машиностроения в России заметно превосходят соответствующие показатели Беларуси.

Сравнение интенсивности изменений уровней рядов во времени возможно с помощью *коэффициентов опережения (отставания)*, представляющих собой отношение базисных темпов роста (или прироста) двух рядов динамики за одинаковые отрезки времени. *Коэффициент опережения (отставания)* $К_{оп}$ показывает, во сколько раз быстрее растёт (отстаёт) уровень одного ряда динамики по сравнению с другим. При этом сравнении темпы должны характеризовать тенденцию одного направления.

Для нашего примера $К_{оп} = 64 / 40 = 1,6$. Это значит, что производство продукции машиностроения в России за рассматриваемый период сокращалось в 1,6 раза быстрее, чем в Беларуси.

1.8.4 Методы анализа основной тенденции развития в рядах динамики

Одной из важнейших задач статистики является определение в рядах динамики *общей тенденции развития явления.*

На развитие явления во времени оказывают влияние факторы, различные по характеру и силе воздействия. Одни из них оказывают практически постоянное воздействие и формируют в рядах динамики определённую *тенденцию* развития. Воздействие же других факторов может быть кратковременным или носить случайный характер. Поэтому при анализе динамики речь идёт не просто о тенденции развития, а об *основной тенденции*, достаточно стабильной на протяжении изученного этапа развития.

Основной тенденцией развития (трендом) называется плавное и устойчивое изменение уровня явления во времени, свободное от случайных колебаний. Задача состоит в том, чтобы выявить общую тенденцию в изменении уровней ряда, освобождённую от действия различных случайных факторов. С этой целью ряды динамики сглаживаются (выравниваются).

Выравнивание (сглаживание) производится:

1) методом укрупнения интервалов;
2) способом скользящей (подвижной) средней;
3) аналитическим способом.

Одним из наиболее простых методов изучения основной тенденции в рядах динамики является *укрупнение интервалов*. Он основан на укрупнении периодов времени, к которым относятся уровни ряда динамики. Например, ряд ежесуточного выпуска продукции заменяется рядом месячного выпуска продукции.

При выравнивании способом *скользящей средней* укрупняется интервал и вместо каждого уровня заданного ряда берутся средние из окружающих его уровней с той и другой стороны. Получается средняя, охватившая группу 3, 5, 7 уровней, в середине которых находится взятый рассчитанный средний уровень:

$$\frac{y_1+y_2+y_3}{3};\ \frac{y_2+y_3+y_4}{3};\ \frac{y_3+y_4+y_5}{3}.$$

Пример. Имеются следующие данные о выгрузке из вагонов картофеля (таблица 1.11). Произведем расчет подвижной средней путем сглаживания уровней ряда динамики.

Таблица 1.11 – **Расчет подвижной трехмесячной средней по выгрузке картофеля из вагонов**

Месяц	Отгрузка картофеля, т	Подвижная трехмесячная сумма	Подвижная трехчленная средняя
Январь	40,4	–	–
Февраль	36,8	40,4+36,8+40,6=117,8	117,8:3=39,27
Март	40,6	36,8+40,6+38,0=115,4	115,4:3=38,47
Апрель	38,0	40,6+38,0+42,2=120,8	120,8:3=40,3
Май	42,2	38,0+42,2+48,5=115,4	115,4:3 = 38,5
Июнь	48,5	–	–

Сглаженный ряд по трём месяцам короче фактического на один уровень в начале и в конце.

При подвижной пятичленной в выравненном ряду будут отсутствовать показатели первых двух начальных и конечных двух членов. Недостатком сглаживания ряда является «укорачивание» по сравнению с фактическим, а следовательно, потеря информации.

Для того чтобы создать количественную модель, выражающую основную тенденцию изменения уровней динамического ряда во времени, используется *аналитическое выравнивание*, при котором общая тенденция развития рассчитывается как функция времени: $\overline{y}_t = f(t)$, где $\hat{y}_t$ – уровни динамического ряда, вычисленные по соответствующему аналитическому уравнению на момент времени t.

Определение теоретических (расчётных) уровней $\overline{y}_t$ производится на основе так называемой *адекватной математической модели*, которая наилучшим образом отображает (аппроксимирует) основную тенденцию ряда динамики. Выбор типа модели зависит от цели исследования и должен быть основан на теоретическом анализе, выявляющем характер развития явления, а также на графическом изображении ряда динамики.

Подбор адекватной функции осуществляется методом наименьших квадратов, который предполагает, что отклонение суммы квадратов между теоретическими $\overline{y}_t$ и эмпирическими y_i уровнями должно быть минимальным:

$$\sum(\overline{y}_t - y_i) = \min.$$

Подбор математической функции зависит от типа развития рассматриваемого явления.

Равномерное развитие. Для него характерны постоянные абсолютные приросты, основная тенденция развития описывается линейной функцией

$$\overline{y_x} = a_0 + a_1 t,$$

где $\bar{y}_x$ – уровень, найденный по уравнению; a_0, a_1 – параметры уравнения, которые при применении способа наименьших квадратов находятся из решения системы нормальных уравнений; t – время или иной аргумент.

Равноускоренное (равнозамедленное) развитие. Уровни таких рядов динамики изменяются с постоянными темпами прироста. Основная тенденция развития в таких рядах динамики отображается функцией параболы второго порядка:

$$\bar{y}_x = a_0 + a_1 t + a_2 t^2 .$$

Развитие с переменным ускорением (замедлением). Для этого типа динамики основная тенденция развития выражается функцией параболы третьего порядка:

$$\bar{y}_x = a_0 + a_1 t + a_2 t^2 + a_3 t^3.$$

Развитие по экспоненте. Этот тип динамики характеризуют стабильные темпы роста. Основная тенденция в рядах динамики отображается показательной функцией

$$\bar{y}_x = a_0 a_1^t.$$

Развитие с замедлением роста в конце периода. Основная тенденция развития в этих рядах динамики выражается полулогарифмической функцией

$$\bar{y}_x = a_0 + a_1 \lg t.$$

При аналитическом выравнивании могут быть использованы и другие функции:

степенная $\bar{y}_x = a_0 t a 1$,

гиперболы $\bar{y}_x = a_0 + a_1 1/t$.

Аналитическое выравнивание динамического ряда по прямой. Выравнивание по прямой – это нахождение плавного изменения уровня ряда динамики согласно уравнению прямой:

$$\bar{y}_x = a_0 + a_1 t,$$

где $\bar{y}_x$ – теоретический уровень, найденный по уравнению;

a_0, a_1 – параметры уравнения;

t – время или иной аргумент.

Параметры a_0 и a_1 согласно методу наименьших квадратов, находятся из решения системы нормальных уравнений:

$$\begin{cases} na_0 + a_1 \sum t = \sum y; \\ a_0 \sum t + a_1 \sum t^2 = \sum yt. \end{cases}$$

Расчёт параметров значительно упрощается, если за начало отсчёта времени (t = 0) принять центральный интервал (момент). При чётном числе уровней значения t – условного обозначения времени – будут такими:

2008 г.	2009 г.	2010 г.	2011 г.
−3	−1	+1	+3

При нечётном числе уровней

2007 г.	2008 г.	2009 г.	2010 г.	2011 г.
−2	−1	0	+1	+2

В обоих случаях $\Sigma t = 0$, система уравнений принимает вид

$$\begin{cases} na_0 = \sum y \\ a_1 \sum t^2 = \sum yt. \end{cases}$$

Из первого уравнения $a_0 = \dfrac{\sum y}{n}$;

Из второго уравнения $a_1 = \dfrac{\sum yt}{\sum t^2}$,

n – число членов ряда.

Пример. Грузооборот по отделению дороги (ввоз), тыс. тарифных т·км, представлен следующим рядом динамики (таблица 1.12). Произведём его выравнивание по прямой.

Таблица 1.12 – **Грузооборот по отделению дороги (ввоз) и выравнивание динамического ряда по прямой**

Годы	Грузооборот (y)	t	t^2	yt	$\overline{y}_t$
2004	8,5	−7	49	−59,5	8,74
2005	8,7	−5	25	−48,5	9,10
2006	8,3	−3	9	24,9	9,46
2007	10,5	−1	1	10,5	9,82
2008	10,4	+1	1	10,4	10,18
2009	11,4	+3	9	34,2	10,54
2010	9,2	+5	25	46,0	10,90
2011	12,0	+7	49	84,0	11,26
Сумма	80,0	0	168	31,2	−

По приведенным выше формулам найдем a_0 = 10,00 и a_1 = 0,18. Уравнение прямой $\overline{y}_t$ = 10,00 + 0,18t. Таким образом, выравненный по прямой динамический ряд будет иметь следующие значения: 2004 год – 8,74 тыс. т·км; 2005 – 9,10; 2006 – 9,46; 2007 – 9,82; 2008 – 10,18 ; 2009 − 10,54; 2010 − 10,90; 2011 − 11,26 тыс. т·км.

Параболическое выравнивание динамического ряда – это нахождение плавного изменения уровня ряда в предположении о его развитии по параболе (кривой n-го порядка). Уравнение кривой 2-го порядка: $y = a + a_1t + a_2t^2$. Уравнение кривой 3-го порядка: $y = a_0 + a_1t + a_2t^2 + a_3t^3$. Уравнение кривой n-го порядка: $y = a_0 + a_1t + a_2t^2 + \ldots + a_nt^n$. Параболическое выравнивание сводится по существу к определению параметров кривой $a_0, a_1, a_2, \ldots, a_n$. Для этого при применении способа наименьших квадратов необходимо решить систему нормальных уравнений.

Так, например, для выравнивания по кривой 2-го порядка $y = a_0 + a_1t + a_2t^2$ система нормальных уравнений имеет вид

$$\begin{cases} na_0 + a_1\sum t + a_2\sum t^2 = \sum y; \\ a_0\sum t + a_1\sum t^2 + a_2\sum t^3 = \sum yt; \\ a_0\sum t^2 + a_1\sum t^3 + a_2\sum t^4 = \sum yt^2. \end{cases}$$

Число уравнений зависит от степени свободы (n) кривой. Так, для определения параметров уравнения параболы необходимо решить систему трех уравнений с тремя неизвестными, для определения параметров уравнения кривой 3-го порядка – систему четырех уравнений с четырьмя неизвестными и т.д. В решении вопроса о применимости параболического выравнивания по параболе того или иного порядка существенную помощь оказывает метод конечных разностей.

Конечные разности – это соотношение вида:

$\Delta f(x_n) = f(x_{n+1}) - f(x_n)$ − разность 1-го порядка;

$\Delta^2 f(x_n) = \Delta f(x_{n+1}) - \Delta f(x_n)$ − разность 2-го порядка;

$\Delta^k f(x_n) = \Delta^{k-1} f(x_{n+1}) - \Delta^{k-1} f(x_n)$ − разность k-го порядка;

где $x_n = x_0 + nh$; h – const; n – целое число.

Конечные разности исследуют функции при прерывном значении аргумента. Например, полагая, что разность двух последовательных значений x равна 1, и имея значения функции $f(x)$, получим следующие конечные разности (таблица 1.13)

Таблица 1.13 − **Определение конечных разностей**

X	$f(x)$	$\Delta f(x)$	$\Delta^2 f(x)$	$\Delta^3 f(x)$	$\Delta^4 f(x)$
0	1				
		2			
1	3		2		
		4		6	
2	7		8		0
		12		6	
3	19		14		0
		26		6	
4	45		20		
		46			
5	91				

Конечные разности можно применить к изучению ряда динамики, взяв за аргумент время, а за функцию − его уровни. В частности, используя свойство, по которому конечные разности кривой 1-го и более высокого порядков для функций n-й степени равны нулю, а конечные разности n-го порядка постоянны, с помощью анализа конечных разностей в известной мере можно судить о применимости тех или иных формул для определения основной тенденции развития динамических рядов. Так, линейная функция характеризуется постоянством первой разности и равенством нулю третьей разности. Кривая 3-го порядка характеризуется постоянством второй разности и равенством нулю третьей разности. Кривая 4-го порядка характеризуется постоянством третьей разности и равенством нулю четвертой разности.

Результаты аналитического выравнивания рядов динамики представляются в виде графиков фактических и расчётных (теоретических) значений.

1.8.5 Методы изучения сезонных колебаний

Сезонные колебания в рядах динамики являются результатом влияния природно-климатических условий, общих экономических факторов и других воздействий.

Суть сезонности заключается в разрыве между периодом производства и рабочим периодом: чем больше этот разрыв, тем выше показатель сезонности. Время производства − это время, необходимое для производства того или иного готового продукта. Время производства состоит как из рабочего периода, так из времени перерывов, иногда необходимых в процессе производства. Под *рабочим периодом понимается определенное число связанных между собой рабочих дней, необходимых в определенной отрасли производства для получения готового продукта*. Рабочий период может быть различным по своей продолжительности. В одних отраслях ежедневно изготовляется готовый продукт, а в других отраслях процесс создания готового продукта длится какое-то число дней, месяцев, а может быть, и лет, как, например, в производстве сложных машин, в выращивании скота и т.п. Кроме того, рабочий период может быть непрерывным, как в большинстве отраслей промышленности, непрерывным в собственном, прямом смысле слова, как, например, в горном деле, металлургии, транспорте, где процесс производства осуществляется непрерывно и более или менее равномерно. Обратный приток затраченных средств в этих отраслях также более или менее равномерен и происходит через определенные промежутки времени. В тех же отраслях, где рабочее время составляет лишь часть времени производства, оборотные средства затрачиваются неравномерно, а обратный их приток совершается разом, в момент, определяемый условиями производства. Особенно отчетливо это наблюдается в сельском хозяйстве, где рабочий период и период производства не совпадают и последний значительно продолжительнее.

Сезонность и сезонные колебания вызываются различными причинами. Но как в производстве, так и в обращении сезонные колебания отрицательно сказываются на развитии экономики страны, обусловливают неравномерность использования трудовых ресурсов и оборудования в течение года, а это, в свою очередь, приводит к понижению производительности труда и повышению себестоимости изготовляемой продукции. Сезонные колебания в одних отраслях экономики вызывают соответствующие колебания в других, иначе говоря, проблема сезонности является общей проблемой экономики.

В статистике существует ряд методов изучения и измерения сезонных колебаний. Самый простой заключается в построении специальных показателей, которые называются индексами сезонности I_s. Совокупность этих показателей отражает сезонную волну. *Индексами сезонности* являются отношения фактических (эмпирических) внутригрупповых уровней к теоретическим (расчётным) уровням, выступающим в качестве базы сравнения.

Методы анализа сезонности делятся на две группы:

1 *Метод простой средней* применяется для анализа сезонности явлений, уровни которых не имеют резко выраженной тенденции увеличения или уменьшения. Сущность этого метода заключается в определении сезонной волны (индекса сезонности) как процентного отношения средних квартальных уровней к общей средней.

Пример. Данные о сезонности отправления зерновых грузов (тыс. т) приведены в таблице 1.14. Вычислить сезонную волну методом простой средней арифметической.

Сначала определяем поквартальные средние уровни отправления как простые средние арифметические за каждый квартал на протяжении шести лет. Например, для I квартала средняя

$$\frac{51{,}87+47{,}99+43{,}02+46{,}29+40{,}03+36{,}45}{6}=\frac{265{,}55}{6}=44{,}26 \text{ тыс. т.}$$

Аналогичные расчёты производим для всех кварталов. Результаты расчётов заносим в таблицу 1.14. На основании полученных данных рассчитываем коэффициент сезонной волны. Из данной таблицы видно, что в среднем в I квартале отправлялось грузов меньше всего за рассматриваемый период времени.

Коэффициент сезонности в I квартале составил 92,58 % .

В среднем за шесть лет в I квартале отправлялось зерновых грузов на 92,58 – − 100,00 = −7,42 % меньше средней квартальной отправки, а в III квартале на 115,18 − – 100,00 = 15,8 % больше средней квартальной отправки.

Таблица 1.14 – **Анализ методов простой средней сезонности отправления зерновых грузов**

Год	Кварталы				Итого за год	Средние квартальные уровни
	I	II	III	IV		
1-й	51,87	54,65	62,31	54,12	223,25	55,81
2-й	47,99	48,73	48,03	46,61	191,36	47,84
3-й	43,02	49,62	58,44	48,00	199,08	49,77
4-й	46,29	47,99	57,17	45,42	196,87	49,22
5-й	40,03	40,15	47,34	35,29	162,81	40,70
6-й	36,35	39,11	57,27	47,71	174,44	43,61
Итого за 6 лет	265,55	280,55	330,56	271,15	1147,81	286,96
Средние уровни за шесть лет	44,26	46,76	55,09	45,19	191,30	47,83
Сезонная волна	92,58	97,76	115,18	94,48	400	100,00

Применение метода простой средней для расчета сезонной волны дает возможность нейтрализовать случайные колебания показателей исследуемого ряда динамики и определить сезонные колебания в среднем за весь период. Правильность полученной сезонной волны зависит как от числа уровней ряда, привлекаемых для анализа, так и от характера их изменения: чем продолжительнее период анализа, чем больше число лет привлекается к расчетам, тем устойчивее будут полученные данные.

При наличии маловыраженной общей тенденции подъема уровней ряда динамики, ее влияние на сезонную волну можно уменьшить с помощью некоторого преобразования уровней ряда. Для этого исчисляются процентные отношения уровней ряда к их среднеквадратичному показателю за каждый год, а затем из полученных отношений определяется средняя арифметическая для каждого квартала за весь анализируемый период – *индекс сезонности*.

2 *Метод относительных чисел.* Этот метод применяется для анализа сезонности тех рядов динамики, развитие общей тенденции которых происходит равномерно.

Пример. Провести анализ сезонного отправления лесных грузов методом относительных чисел, использовав при этом данные о его отправлении (тыс. т) поквартально за шесть лет. Исходные данные для расчёта приведены в таблице 1.15.

Таблица 1.15 – **Поквартальное отправление лесных грузов в течение шести лет**

Год	Отправление лесных грузов по кварталам				Итого за год
	I	II	III	IV	
1-й	44,7	43,2	44,7	54,6	187,2
2-й	55,3	44,5	43,4	51,5	194,7
3-й	51,9	40,1	41,5	55,9	189,4
4-й	54,3	46,7	43,8	59,8	204,4
5-й	57,9	48,7	44,9	60,0	211,6
6-й	60,7	51,0	51,7	69,0	232,4

Цепные отношения вычисляются как процентные отношения объемов отправления грузов за каждый квартал к объему отправления предшествующего квартала, в результате получается система относительных чисел, связанных в цепь. Например, отношение объёма отправления в I квартале второго года к первому году 55,3 / 44,7 = = 1,2371 или 123,71 %, третьего ко второму 51,9 / 55,3 = 0,9385 или 93,8 %. Далее из относительных чисел вычисляется простая средняя величина для каждого квартала за все шесть лет. Затем средняя за I квартал приравнивается к единице (или 100), а средние за остальные кварталы определяются по методу цепных произведений. Таким образом, если средний уровень I квартала будет 100, то во II квартале он будет равен 84,75, в III – 83,60, в IV – 108,56.

При отсутствии общей тенденции подъема или снижения произведение преобразованной средней за IV квартал на среднюю из цепных отношений I квартала дает первоначальный уровень преобразования средней, т.е. 100 %; оно будет более 100, если наблюдается общая тенденция увеличения и, напротив, менее 100, если наблюдается общая тенденция уменьшения. Расхождение между произведением преобразованной средней за IV квартал на среднюю из цепных отношений I квартала и 100 % характеризует погрешность, которая возникла в результате повышающейся или понижающейся общей тенденции, эту погрешность необходимо устранить. Наиболее простой способ устранения погрешности состоит в равномерном распределении ее на все кварталы.

Для анализа внутригодовой динамики явлений могут применяться гармоники ряда Фурье. При аналитическом выражении изменении уровней ряда динамики используется формула $y_t = a_0 + (a_k \cos Rt + b_k \sin Rt)$.

В формуле R определяет номер гармоники. При анализе ряда внутригодовой динамики по месяцам значение R равно 12. При изучении сезонных колебаний получают систему нормальных уравнений, параметры которых вычисляются по формулам:

$$a_0 = \frac{\sum y_i}{n}; \qquad a_k = \frac{2}{n}\sum y_i \cos rt_i; \qquad b_k = \frac{2}{n}\sum y_i \sin rt_i.$$

По полученным параметрам синтезируется математическая модель. На основе модели определяются для каждого месяца расчетные уровни y_{ti}.

1.9 Индексы

1.9.1 Понятие об индексах и их значения

Индексы − показатели особого рода. Прежде всего, это относительные величины, характеризующие динамику явления (выполнение плана или сравнение регионов по тем или иным экономическим показателям). От обычных относительных величин индексы отличает то, что они характеризуют отношение сложных явлений, складывающихся под влиянием различных причин. Индексы, как правило, не ограничиваются простым показом отношения, а выявляют роль и значение отдельных условий и составных частей данного сложного явления. Например, индекс цен показывает, как изменились цены на все товары или отдельную группу товаров, как отразилось это изменение на соотношении количества и цен отдельных товаров, как само изменение цен отразилось на товарообороте, покупательной способности рубля, степени удовлетворения покупательского спроса.

Под индексом в статистике понимают относительный показатель, характеризующий изменение величины какого-либо явления (простого или сложного, состоящего из соизмеримых или несоизмеримых элементов) во времени, пространстве или по сравнению с любым эталоном (нормативом, планом, прогнозом и т.д.).

Индекс применяется также для изучения роли факторов, оказывающих влияние на изменение данного явления. Так, с помощью взаимосвязи индексов можно определить, в какой мере от увеличения объема продукции зависит от роста производительности труда и в какой мере от увеличения численности.

Таким образом, индекс характеризует изменение величины сложного экономического явления, состоящего из элементов, которые непосредственно нельзя суммировать, поэтому он является более сложным и многосторонним показателем, чем относительные или средние величины. Например, можно ли определить все изменения товарооборота в натуральном выражении? Нет, нельзя, так как реализуемые товары имеют различные натуральные единицы измерения (крупа в килограммах, растительное масло в литрах, обувь в парах, ткани в метрах). Следовательно, складывать объемы разнородных товаров для определения динамики товарооборота нельзя. Суммирование будет возможным только в тех случаях, когда все товары будут приведены к сопоставимому виду, что достигается путем индексных расчетов.

Элементами любого индекса являются: а) индексируемая величина; б) тип (форма) индекса; в) веса индекса; г) сроки исчисления.

Основным элементом индексного отношения является *индексируемая величина*. *По содержанию индексируемых величин* индексы разделяются на индексы количественных (объёмных) и индексы качественных величин. Индексы количественных показателей − это индекс физического (натурального) объема продукции. Индексы качественных показателей − это индекс цен, индексы производительности труда и т.д.

По степени охвата единиц совокупности индексы делятся на индивидуальные и общие. *Индивидуальные индексы* служат для характеристики изменения отдельных элементов сложного явления (например, изменение объёма производства конкретного вида продукции).

Общий индекс отражает изменение всех элементов сложного явления. При этом под сложным явлением понимают такую статистическую совокупность, отдельные элементы которой непосредственно не подлежат суммированию (физический объём продукции, включающий разные наименования товаров).

Если индексы охватывают не все элементы сложного явления, а лишь их часть, то их называют *групповыми индексами или субиндексами* (например, индексы продукции по отдельным отраслям промышленности).

В зависимости от способа построения различают индексы *агрегатные* и индексы *средние*. Средние индексы — это индексы средние арифметические, индексы средние гармонические, индексы средние геометрические.

В зависимости от весов различают индексы *простые* (невзвешенные) и индексы *взвешенные*, а среди последних — индексы с постоянными (неизменными) весами и индексы с переменными весами (в меру необходимости с течением времени пересматриваемыми).

В зависимости от сроков исчисления рассматривают индексы базисные (с постоянной, неизменной во времени базой) и индексы цепные (если числовые значения индексируемой величины в каждый данный «текущий» срок сопоставляются с их значениями в предшествующий срок).

1.9.2 Формы индексов

Индексы могут быть индивидуальными и сводными (общими).

Индивидуальными индексами называются относительные числа, характеризующие соотношение отдельных величин экономических явлений: цены одного товара, себестоимости изделия, количества какого-либо одного реализованного продукта, обозначаются буквой *i* и снабжаются подстрочным знаком индексируемого показателя.

При расчете индексов особое внимание следует уделять базе сравнения. В индексах, характеризующих изменение явления в динамике, различают два периода: базисный и текущий (отчетный). *Базисный* − это начальный период, т.е. период, с которым производится сравнение, *текущий* (отчетный) − это период, уровень которого сравнивается с базисным.

Индивидуальный индекс как относительное число получается в результате сравнения двух абсолютных уровней изучаемого явления.

Индивидуальный индекс цен

$$i_{\text{p}} = \frac{p_1}{p_0},$$

где p_1 − цена за единицу продукта в текущем или отчетном периоде;

p_0 − цена за единицу продукта в базисном периоде.

Для того чтобы показать изменение количества выпуска продукции (объёма работы), употребляется индивидуальный индекс количества или физического объема (i_q):

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

где q_1 − количество продукции (объёма работы) в текущем периоде;

q_0 − количество продукции (объёма работы) в базисном периоде.

В экономических расчётах для измерения динамики сложного явления чаще всего используются общие (сводные) индексы.

Сводная форма индексов. Методика расчёта общих индексов сложнее, чем индивидуальных, и зависит от характера индексируемых показателей, наличия исходных данных и целей исследования.

Сводными индексам называются относительные числа, характеризующие соотношения между такими совокупностями величин экономических явлений, которые непосредственно в своей натуральной форме несоизмеримы. Сводные индексы могут быть построены как *агрегатные* и как *средние из индивидуальных*. Последние подразделяются на средние гармонические и среднеарифметические.

Основной формой сводных индексов является агрегатная. Одной из важнейших проблем, возникающих при построении сводных индексов, является определение соизмерителей, то есть весов индексов, при помощи которых несоизмеримые элементы индексов приводятся к сопоставимому виду.

Каждый сводный индекс состоит из двух элементов: *индексируемой величины*, то есть величины, которая изучается в данном индексе, и *весов индекса*, при помощи которых несоизмеримые показатели индекса приводятся в сопоставимый вид. Иначе говоря, веса − это одинаковые величины в числителе и знаменателе индекса. При расчете агрегатного индекса для разнородной совокупности находят такой общий показатель, в котором можно объединить все ее элементы. Если мы сравним доход железной дороги по *n* видам деятельности в текущем периоде с его величиной в базисном периоде, то получим сводный индекс доходов:

$$I_q = \frac{\sum_{i=1}^{n} p_{i_1} q_{i_1}}{\sum_{i=1}^{n} p_{i_0} q_{i_0}},$$

где p_{i_1} и p_{i_0} − цена, а q_{i_1} и q_{i_0} – объем *i*-го вида деятельности соответственно в текущем и базисном периодах.

Числитель данного индекса представляет собой доходы текущего периода (сумма цен на различные измерители работы, умноженных на их объемы), знаменатель – доходы предшествующего периода.

На величину индекса доходов оказывают влияние как изменение цен на измерители работы, так и изменение их объемов.

Для того чтобы оценить изменение только цен (индексируемой величины), необходимо объём работы (веса индекса) зафиксировать на каком-либо постоянном уровне. При исследовании динамики таких показателей, как цена, себестоимость, производительность труда, количественный показатель обычно фиксируют на уровне текущего периода. Таким способом получают сводный индекс цен (индекс цен Пааше):

$$I_q = \frac{\sum_{i=1}^{n} q_{i_1} p_{i_0}}{\sum_{i=1}^{n} q_{i_0} p_{i_0}},$$

Индекс цен Пааше показывает, насколько товары в текущем периоде подорожали (подешевели) по сравнению с базисным периодом, а индекс цен Ласпейерса показывает, во сколько раз товары базисного периода дороже (дешевле) в результате изменения цен в отчетном периоде. Как правило, индекс цен, рассчитанный по формуле Пааше, несколько занижает, а по формуле Ласпейерса – завышает темпы инфляции.

Третьим индексом в данной индексной системе является сводный индекс физического объема. Он характеризует изменение объёма работы не в денежных, а в физических единицах измерения:

$$I_q = \frac{\sum_{i=1}^{n} q_{i_1} p_{i_0}}{\sum_{i=1}^{n} q_{i_0} p_{i_0}}.$$

Весами в данном индексе выступают цены, которые фиксируется на базисном уровне.

Между рассчитанными индексами также существует взаимосвязь:

$$I_p = I_q = I_{pq}.$$

Рассмотрим применение индексного метода при анализе изменения затрат на производство и себестоимость продукции. Для определения общего изменения уровня себестоимости нескольких видов продукции, выпускаемых предприятием, рассчитывается сводный индекс себестоимости. При этом себестоимость взвешивается по объему производства отдельных видов продукции текущего периода:

$$I_z = \frac{\sum_{i=1}^{n} z_{i_1} q_{i_1}}{\sum_{i=1}^{n} z_{i_0} q_{i_0}},$$

где z_{i_1} и z_{i_0} – себестоимость i-го вида продукции соответственно в текущем и базисном периодах.

Числитель этого индекса отражает затраты на производство текущего периода, а знаменатель – условную величину затрат при сохранении себестоимости на базисном уровне. Разность числителя и знаменателя показывает сумму экономии или потерь предприятия от изменения себестоимости:

$$E = \sum_{i=1}^{n} z_{i_1} q_{i_1} - \sum_{i=1}^{n} z_{i_0} q_{i_1}.$$

Сводный индекс физического объема продукции, взвешенный по себестоимости, имеет следующий вид:

$$I_q = \frac{\sum_{i=1}^{n} q_{i_1} z_{i_0}}{\sum_{i=1}^{n} q_{i_0} z_{i_0}}.$$

Третьим показателем в данной индексной системе является *сводный индекс затрат на производство:*

$$I_{zq} = \frac{\sum_{i=1}^{n} z_{i_1} q_{i_1}}{\sum_{i=1}^{n} z_{i_0} q_{i_0}}.$$

Все три индекса взаимосвязаны между собой соотношением $I_z \cdot I_q = Izq$.

Индивидуальные индексы производительности труда, основанные на этих показателях, имеют следующий вид:

$$i_w = \frac{w_1}{w_0} = \frac{q_1}{T_1} : \frac{q_0}{T_0};$$

$$i_t = \frac{t_0}{t_1} = \frac{T_0}{q_0} : \frac{T_1}{q_1},$$

где T – суммарные затраты времени на выпуск данной продукции, чел·ч, чел·дн. или чел·мес.

Трудоемкость является показателем, обратным показателю производительности труда, поэтому снижение трудоемкости в текущем периоде по сравнению с базисным свидетельствует о росте производительности труда.

Располагая данными о трудоемкости различных видов продукции (I = 1, 2, …, n) и объемах их производства, можно рассчитать сводный индекс производительности труда (по трудоемкости):

$$I_t = \frac{\sum_{i=1}^{n} t_{i_0} q_{i_1}}{\sum_{i=1}^{n} t_{i_1} q_{i_1}}.$$

Знаменатель этого индекса отражает реально имевшие место общие затраты времени на выпуск всей продукции в текущем периоде (T_1). Числитель представляет собой условную величину, показывающую, какими были бы затраты времени на выпуск этой продукции, если бы трудоемкость не изменилась.

Пример. По данным о выпуске продукции на предприятии и её трудоёмкости (таблица 1.16) определить индекс производительности труда.

Таблица 1.16 – **Выпуск продукции на предприятии и её трудоёмкость**

Вид продукции (i)	Трудоёмкость, чел·ч		Объём выпуска, шт.		Расчётные графы	
	Январь (t_{i_0})	Февраль (t_{i_1})	Январь (q_{i_0})	Февраль (q_{i_1})	$t_{i_0} q_{i_1}$	$t_{i_1} q_{i_1}$
Изделие (1)	1,0	0,9	458	450	450,0	405,0
Изделие (2)	1,2	1,0	311	324	388,8	324,0
Изделие (3)	0.9	0.8	765	752	676,8	601,6
И т о г о					1515,6	1330,6

Рассчитаем сводный индекс производительности труда по данным о её трудоемкости:

$$I_t = \frac{\sum_{i=1}^{n} t_{i_0} q_{i_1}}{\sum_{i=1}^{n} t_{i_1} q_{i_1}} = \frac{1515,6}{1330,6} = 1,139 \text{ или } 113,9\ \%.$$

Прирост производительности труда по предприятию составил 13,9 %.

При расчете сводного индекса производительности труда в стоимостном выражении (по выработке) необходимо количество продукции, произведенной за каждый период, взвесить по каким-либо ценам, принятым за сопоставимые. В качестве сопоставимых могут выступать цены текущего или базисного периода, какого-либо другого периода или средние цены. Индекс в этом варианте рассчитывается по формуле

$$I_\omega = \frac{\sum_{i=1}^{n} q_{i_1} p_i}{\sum_{i=1}^{n} T_{i_1}} : \frac{\sum_{i=1}^{n} q_{i_0} p_i}{\sum_{i=1}^{n} T_{i_0}}.$$

Первая часть этой формулы представляет собой среднюю выработку в отчетном периоде, вторая часть – в базисном.

Сводные индексы в среднеарифметической и среднегармонической формах. В ряде случаев вместо индексов в агрегатной форме удобнее использовать среднегармонические индексы. Любой сводный индекс можно представить как среднюю взвешенную из индивидуальных индексов. Однако при этом форму средней нужно выбрать таким образом, чтобы полученный средний индекс был тождествен исходному агрегатному индексу.

Предположим, мы располагаем данными о стоимости проданной продукции в текущем периоде ($p_1\ q_1$) и индивидуальными индексами цен ($i_p = \frac{p_1}{p_0}$), полученными, скажем, в результате выборочного наблюдения.

Тогда в знаменателе сводного индекса цен $I_p = \frac{\sum_{i=1}^{n} p_{i_1} q_{i_1}}{\sum_{i=1}^{n} p_{i_0} q_{i_1}}$ можно использовать следующую замену: $p_{i_0} = \frac{1}{i_{ip}} p_{i_1}$.

Таким образом, сводный индекс цен будет выражен как средний гармонический индекс

$$I_p = \frac{\sum_{i=1}^{n} p_{i_1} q_{i_1}}{\sum_{i=1}^{n} \frac{1}{i_{i_p}} p_{i_1} q_{i_1}}.$$

Пример. По данным таблицы 1.17 требуется получить сводную оценку изменения грузовых тарифов. Для этого необходимо рассчитать среднегармонический индекс тарифов, на основе данных об общем размере доходов в текущем периоде и изменении тарифов по сравнению с базисным периодом. Результаты расчётов сведём в таблицу 1.17.

Таблица 1.17 – **Объём доходов и изменение тарифов**

Грузы (i)	Доходы, тыс. д. е. ($p_{i_1} q_{i_1}$)	Изменение тарифов в текущем периоде по сравнению с базисным, % ($i_{i_p} \cdot 100\% - 100\ \%$)	Расчетные графы	
			i_{i_p}	$\frac{p_{i_1} q_{i_1}}{i_{i_p}}$
1	23000	+4,0	1,040	22115
2	21000	+2,3	1,023	20528
3	29000	−0,8	0,992	29234
Итого	73000			71877

Вычислим средний гармонический индекс тарифов по формуле

$$I_p = \frac{\sum_{i=1}^{n} p_{i_1} q_{i_1}}{\sum_{i=1}^{n} \frac{1}{i_{i_p}} p_{i_1} q_{i_1}} = \frac{73000}{71877} = 1{,}016 \text{ или } 101{,}6\ \%.$$

Тарифы по данным видам грузов в текущем периоде по сравнению с базисным в среднем возросли на 1,6 %.

При расчете *сводного индекса физического объема работы* можно использовать формулу

$$I_p = \frac{\sum_{i=1}^{n} q_{i_1} p_{i_0}}{\sum_{i=1}^{n} \frac{1}{i_{i_p}} q_{i_0} p_{i_0}}.$$

При этом в числителе производится замена:

$$qi_1 = i_{iq} q_{i0}.$$

Тогда индекс примет вид

$$I_q = \frac{\sum_{i=1}^{n} i_{i_q} q_{i_0} p_{i_0}}{\sum_{i=1}^{n} \frac{1}{i_{i_p}} q_{i_0} p_{i_0}}.$$

Пример. По данным таблицы 1.18 о величине доходов от перевозок и изменении их объёмов по разным видам грузов рассчитать индекс физического объема перевозок.

Таблица 1.18 − **Доходы и изменение объёма перевозок**

Грузы (i)	Доходы, тыс. д. е. ($q_{i_0} p_{i_0}$)	Изменение объема перевозок в текущем периоде по сравнению с базисным, % ($i_{i_q} \cdot 100\ \% - 100\ \%$)	Расчетные графы	
			i_{i_q}	$i_{i_q} q_{i_0} p_{i_0}$
1 2 3	46000 27000 51000	−6,4 −8,2 +1,3	0,936 0,918 1,013	43056 24786 51663
Итого	124000			119505

Рассчитаем средний арифметический индекс объёма перевозок согласно формуле

$$I_q = \frac{\sum_{i=1}^{n} t_{i_q} q_{i_0} p_{i_0}}{\sum_{i=1}^{n} \frac{1}{t_{i_p}} q_{i_0} p_{i_0}} = \frac{119505}{124000} = 0{,}964 \text{ или } 96{,}4\ \%.$$

Объем перевозок в среднем снизился на 3,6 %.

В среднеарифметической форме также может рассчитываться и индекс производительности труда по трудоемкости, известный как индекс

$$I_\omega = \frac{\sum_{i=1}^{n} i_1 T_{i_1}}{\sum_{i=1}^{n} T_{i_1}} : \frac{\sum_{i=1}^{n} \left(\frac{T_{i_0}}{q_{i_0}} : \frac{T_{i_1}}{q_{i_1}} \right) T_{i_1}}{\sum_{i=1}^{n} T_{i_1}}.$$

1.9.3 Индексы постоянного и переменного состава

Индексы широко используются в анализе для выявления меры влияния факторных показателей на средний уровень определяемого или результативного показателя. Индекс цен переменного состава представляет собой отношение полученных средних значений:

$$I_p^{\text{п.с}} = \frac{\sum_{i=1}^{n} p_1 q_{i_1}}{\sum_{i=1}^{n} q_{i_1}} : \frac{\sum_{i=1}^{n} p_{i_0} q_{i_0}}{\sum_{i=1}^{n} q_{i_0}}.$$

Данный индекс характеризует не только изменение индивидуальных цен в местах продажи, но и изменение структуры реализации по предприятиям розничной и оптовой торговли, рынка, городам и регионам. Для оценки воздействия второго фактора рассчитывается *индекс структурных сдвигов*:

$$I^{\text{стр}} = \frac{\sum_{i=1}^{n} p_{i_0} q_{i_1}}{\sum_{i=1}^{n} q_{i_1}} : \frac{\sum_{i=1}^{n} p_{i_0} q_{i_0}}{\sum_{i=1}^{n} q_{i_0}}.$$

Последним в данной системе является рассмотренный выше *индекс цен фиксированного состава,* который не учитывает изменение структуры:

$$I_{\text{р}}^{\text{ф.с.}} = \frac{\sum_{i=1}^{n} p_{i_1} q_{i_1}}{\sum_{i=1}^{n} p_{i_0} q_{i_1}}.$$

Между данными индексами существует следующая взаимосвязь:

$$I_p^{\text{ф.с}} . I^{\text{стр}} = I_p^{\text{п.с}}.$$

Индексы широко используются в факторном анализе для выявления меры влияния факторных показателей на средний уровень определяемого или результативного показателя. Например, необходимо определить, на сколько процентов изменение среднего уровня себестоимости $\bar{Z}$ перевозок обусловлено изменением самой себестоимости как таковой $I_{\bar{z}(z)}$ и на сколько процентов – изменением структуры перевозок $I_{\bar{z}(q)}$. Известны объем перевозок каждого рода груза и их себестоимость в текущем и базисном периодах (данные условные).

Таблица 1.19 – **Динамика объема и себестоимости перевозки грузов**

Род груза	Объем перевозок, млн т·км		Себестоимость перевозок, руб. /10 т·км	
	Базисный	Отчетный	Базисный	Отчетный
Каменный уголь	14400	17500	4,0	5,0
Руда	2000	2500	3,0	3,5
Строительные материалы	600	1000	1,5	2,0

Изменение среднего уровня себестоимости определяем как отношение среднего уровня себестоимости перевозок по всем грузам в отчетном и базисном периодах:

$$I_{\bar{z}} = \frac{\bar{Z}_1}{\bar{Z}_0}.$$

Среднюю себестоимость, в свою очередь, определяют как отношение общих затрат на производство и объем продукции:

$$\bar{Z}_1 = \frac{\Sigma \bar{Z}_1 q_1}{\Sigma q_1}, \qquad \bar{Z}_0 = \frac{\Sigma \bar{Z}_0 q_0}{\Sigma q_0}.$$

Сопоставляя средние уровни себестоимости отчетного и планового периодов, мы наблюдаем изменение среднего уровня себестоимости перевозок всех грузов под влиянием изменения двух факторов себестоимости Z и объема перевозок q:

$$I_{\bar{z}(zq)} = \frac{\bar{Z}_1}{\bar{Z}_0} : \frac{\Sigma \bar{Z}_1 q_1}{\Sigma q_1} : \frac{\Sigma \bar{Z}_0 q_0}{\Sigma q_0};$$

$$I_{\bar{z}(zq)} = \frac{5,0 + 17500 + 3,5 \cdot 2500 + 2,0 \cdot 1000}{17500 + 2500 + 1000} : \frac{4,0 \cdot 14400 + 3,0 \cdot 2000 + 1,5 \cdot 600}{14400 + 2000 + 600} = 1,2331.$$

Средняя себестоимость перевозок всех грузов под влиянием роста себестоимости и объема перевозок выросла на 23,31 %. Этот индекс называют индексом переменного состава. Чтобы определить влияние изменения себестоимости перевозок отдельных грузов на среднюю себестоимость перевозок всех грузов, надо исключить влияние структуры перевозок на ее величину. Для этого объемы перевозок берут на одном уровне. Поскольку это объемный показатель, то берем их на уровне отчета:

$$I_{\bar{z}(z)} = \frac{\Sigma \bar{Z}_1 q_1}{\Sigma q_1} : \frac{\Sigma \bar{Z}_0 q_1}{\Sigma q_1};$$

$$I_{\bar{z}(z)} = \frac{5,0 + 17500 + 3,5 \cdot 2500 + 2,0 \cdot 1000}{17500 + 2500 + 1000} :$$

$$: \frac{4,0 \cdot 17500 + 3,0 \cdot 2500 + 1,5 \cdot 100}{17500 + 2500 + 1000} = 1,2438.$$

Как показывают расчеты, за счет роста себестоимости перевозок отдельных грузов в среднем себестоимость выросла на 24,38 %.

Этот индекс называют индексом постоянного состава, он отражает влияние только индексируемого показателя. По существу это тот же сводный индекс себестоимости

$$I_{\overline{z}(z)} = \frac{\Sigma \bar{Z}_1 q_1}{\Sigma q_1} : \frac{\Sigma \bar{Z}_0 q_1}{\Sigma q_1} = \frac{\Sigma \bar{Z}_1 q_1}{\Sigma z_0 q_1} : \frac{\Sigma q_1}{\Sigma q_1}.$$

Для оценки влияния изменения объема перевозок по отдельным грузам, т.е. влияния структуры перевозок на средний уровень себестоимости грузов, необходимо нивелировать влияние изменения себестоимости перевозки отдельных грузов на ее средний уровень. С этой целью себестоимость перевозки отдельных грузов берем на одном уровне – плановом, поскольку это качественный показатель:

$$I_{\overline{z}(q)} = \frac{\Sigma \bar{Z}_0 q_1}{\Sigma q_1} : \frac{\Sigma \bar{z}_0 q_0}{\Sigma q_0};$$

$$I_{\overline{z}(q)} = \frac{4,0 + 17500 + 3,0 \cdot 2500 + 1,5 \cdot 1000}{17500 + 2500 + 1000} :$$

$$: \frac{4,0 + 14400 + 3,0 \cdot 2000 + 1,5 \cdot 600}{14400 + 2000 + 600} = 0,9915.$$

Как показывают расчеты, за счет изменения структуры средняя себестоимость перевозок всех грузов снизилась на 0,85 %.

Этот индекс называют индексом структурных сдвигов, он отражает влияние структуры объема работ на средний уровень индексируемого показателя.

Правильность выполненных расчетов можно проверить через взаимосвязь индексов:

$$I_{\overline{z}(z,q)} = I_{\overline{z}(z)} I_{\overline{z}(q)} = 1,2438 \cdot 0,9915.$$

1.9.4 Цепные и базисные индексы

Индексы с постоянной базой сравнения называются базисными, индексы с переменной базой сравнения − цепными индексами. Цепные и базисные индексы могут быть рассчитаны для простых и сложных явлений. При построении цепных индексов цены каждого периода сравниваются с ценами предшествующего периода. В базисных индексах цены каждого периода сравниваются с ценами (как правило, первого) периода.

Индексы также могут иметь постоянные или переменные веса. В первом случае при переходе от индекса к индексу веса остаются неизменными, во втором случае – каждый раз используются новые веса. Сочетания этих подходов позволяют получить четыре основных варианта построения индексной системы в динамике. Рассмотрим их на примере сводного индекса цен, рассчитываемого за *m* периодов:

1) Цепные индексы цен:

− с переменными весами

$$I_{p_0}^{1}=\frac{\sum_{i=1}^{n}p_{i_1}q_{i_1}}{\sum_{i=1}^{n}p_{i_0}q_{i_1}};\quad I_{p_0}^{2}=\frac{\sum_{i=1}^{n}p_{i_2}q_{i_2}}{\sum_{i=1}^{n}p_{i_1}q_{i_2}};\quad I_{p_{m-1}}^{m}=\frac{\sum_{i=1}^{n}p_{i_m}q_{i_m}}{\sum_{i=1}^{n}p_{i_{m-1}}q_{i_m}};$$

− с постоянными весами

$$I_{p_0}^{1}=\frac{\sum_{i=1}^{n}p_{i_1}q_{i_0}}{\sum_{i=1}^{n}p_{i_0}q_{i_0}};\quad I_{p_0}^{2}=\frac{\sum_{i=1}^{n}p_{i_2}q_{i_0}}{\sum_{i=1}^{n}p_{i_1}q_{i_0}};\quad I_{p_{m-1}}^{m}=\frac{\sum_{i=1}^{n}p_{i_m}q_{i_0}}{\sum_{i=1}^{n}p_{i_{m-1}}q_{i_0}};$$

2) базисные индексы цен:

− с переменными весами:

$$I_{p_0}^{1}=\frac{\sum_{i=1}^{n}p_{i_1}q_{i_1}}{\sum_{i=1}^{n}p_{i_0}q_{i_1}};\quad I_{p_0}^{2}=\frac{\sum_{i=1}^{n}p_{i_2}q_{i_2}}{\sum_{i=1}^{n}p_{i_0}q_{i_2}};\quad I_{p_0}^{m}=\frac{\sum_{i=1}^{n}p_{i_m}q_{i_m}}{\sum_{i=1}^{n}p_{i_0}q_{i_m}};$$

− постоянными весами:

$$I_{p_0}^{1}=\frac{\sum_{i=1}^{n}p_{i_1}q_{i_0}}{\sum_{i=1}^{n}p_{i_0}q_{i_0}};\quad I_{p_0}^{2}=\frac{\sum_{i=1}^{n}p_{i_2}q_{i_0}}{\sum_{i=1}^{n}p_{i_0}q_{i_0}};\quad I_{p_0}^{m}=\frac{\sum_{i=1}^{n}p_{i_m}q_{i_0}}{\sum_{i=1}^{n}p_{i_0}q_{i_0}}.$$

Индексы аналитические. Это один из основных типов индексных показателей. В отличие от синтетических индексов, дающих сравнительную характеристику уровней экономических явлений, индексы аналитические позволяют оценить степень изменения сложного явления воздействием изменения каждого из связанных с ним простых явлений. Система индексов аналитических состоит из полного индекса, характеризующего изменение рассматриваемого сложного явления под воздействием всех определяющих его факторов, и частных индексов, каждый из которых отражает изменение сложного явления под воздействием изменения того или иного из определяющих его явлений − факторов.

Важнейшей предпосылкой построения системы индексов аналитических является установление формы связи между сложным явлением и определяющими его явлениями − факторами. Для построения системы индексов аналитических необходимо: а) исходя из установленной формы связи между сложными явлениями и его факторами построить полный индекс; б) последовательно элиминируя (исключая) влияние изменения всех факторов, кроме того, влияние которого на изменение сложного явления изучается, построить частные индексы всех рассматриваемых факторов.

Наибольшие трудности возникают при построении системы аналитических индексов для формы связи типа $w = \ldots \Sigma xyz$. В этом случае полный индекс имеет вид

$$I_w = \frac{\sum x_1 y_1 z_{1\ldots}}{\sum x_0 y_0 z_{0\ldots}}.$$

Совокупность же частных индексов может быть построена разными путями в зависимости от принятого метода элиминирования (исключения). Различают цепной метод построения частных индексов (метод цепных подстановок) и метод выявления изолированного влияния отдельных факторов. В первом случае частный индекс каждого фактора строится при элиминировании всех ранее исследованных факторов (частные индексы которых уже построены) на уровне текущего периода, а факторов, влияние которых предстоит исследовать (частные индексы которых еще не построены), на уровне базисного периода. Этот метод приводит к множеству возможных вариантов построения частных индексов, дающих неоднозначные, а порой и противоречивые результаты. Метод выявления изолированного влияния отдельных факторов, в отличие от цепного, приводит к однозначному разложению полного индекса на частные. В этом случае частные индексы всех факторов строятся путем элиминирования изменения всех остальных факторов на уровне базисного периода. Однако здесь совокупность частных индексов, помимо индексов, отражающих влияние изолированного изменения каждого из факторов на изменение сложного явления, содержит еще индексы, отражающие результат взаимосвязанного изменения отдельных групп факторов на изменение сложного явления.

1.10 Методологические основы анализа статистической информации

Анализ статистической информации представляет собой социально-экономическое исследование. Его можно расчленить на следующие этапы: формулировка задачи и подбор источников информации; констатация фактов и их оценка; анализ динамики и структуры; выявление взаимосвязей и анализ влияния показателей на результаты деятельности предприятия; формулировка выводов и выработка предложений по повышению эффективности деятельности предприятия.

Формулировка задачи и подбор источников информации. Изучаемые статистикой явления общественной жизни обладают множеством свойств и взаимосвязей, поэтому целесообразно четко сформулировать задачу исследования. Например, оценить выполнение плана перевозок грузов железной дорогой. В этом случае мы рассматриваем один из видов деятельности и используем конкретный показатель – процент выполнения плана. Если нам надо оценить результаты деятельности железной дороги за определенный период времени, анализ будет носить комплексный характер и для его осуществления мы должны использовать целую систему взаимосвязанных показателей.

Точная формулировка задач анализа необходима не только для придания ему целенаправленности, но и для подбора необходимых источников информации.

Источником информации для анализа могут служить как опубликованные статистические данные, так и материалы отдельных обследований, отчетов и т.д. Но во всех случаях, прежде чем приступить к анализу, необходимо оценить их с точки зрения достоверности содержащихся в них сведений и пригодности их для решения поставленных задач.

Корреляционный анализ. Корреляционный анализ призван, прежде всего, ответить на вопрос, как выбрать с учетом специфики и природы анализируемых переменных подходящих измеритель статистической связи (коэффициент корреляции, корреляционное отношение, ранговый коэффициент корреляции и т.д.). Корреляционный анализ позволяет найти методы проверки того, что полученное числовое значение анализируемого измерителя связи действительно свидетельствует о наличии статистической связи. Наконец, он помогает определить структуру связей между исследуемыми k признаками $x_1, x_2, \ldots, x_k$, сопоставив каждой паре признаков ответ («связь есть» или «связи нет»).

Корреляционный анализ количественных признаков. Одним из основных показателей взаимозависимости двух случайных величин является *парный коэффициент корреляции*, служащий мерой линейной статистической зависимости между двумя величинами. Этот показатель соответствует своему прямому назначению, когда статистическая связь между соответствующими признаками в генеральной совокупности линейна. То же самое относится к *частным и множественным коэффициентам* корреляции.

Парный коэффициент корреляции, характеризующий тесноту связи между случайными величинами x и y, определяется по формуле

$$r = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sigma_x \sigma_y},$$

σ_y

где σ_x и σ_y – среднеквадратические отклонения.

Парный коэффициент корреляции изменяется в пределах от −1 до +1, то есть $-1 \le p \le +1$. При этом между величинами x и y связь функциональная (прямая при $p = +1$ и обратная при $p = -1$). Если же $p = 0$, то между величинами x и y линейная связь отсутствует и они называются *некоррелированными.*

Выборочный парный коэффициент корреляции, найденный по выборке объемом n, где (x_i, y_i) – результат i-го наблюдения, $i = 1, 2, \ldots, n$, определяется по формуле

$$r_{xy} = r = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{S_x S_y},$$

где $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$; $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$; $S_x = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2}$; $S_y = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2}$.

Формула симметрична, т.е. $r_{xy} = r_{xy} = r$. Парный коэффициент корреляции может быть рассчитан по формуле

$$r = \frac{\overline{xy} - \overline{x}\overline{y}}{S_x S_y},$$

где $\overline{xy}$ – средняя арифметическая двух величин, т.е.

$$\overline{xy} = \frac{1}{n}\sum_{i=1}^{n} x_i y_i .$$

Пример. На основании выборочных данных таблицы 1.20 о деятельности $n = 6$ структурных подразделений железной дороги оценить тесноту связи между прибылью (y, млн руб.) и затратами на 1 руб. произведенной продукции (x).

Таблица 1.20 – **Исходные и расчетные данные для определения *r***

Номер структурного подразделения (i)	x_i	y_i	$x_i y_i$	x_i^2	y_i^2
1	96	0,22	21,12	9216	0,049
2	78	1,07	83,46	6084	1,145
3	77	1,00	77,00	5929	1,000
4	89	0,61	54,29	7921	0,372
5	81	0,78	63,18	6561	0,608
6	82	0,79	64,78	6724	0,624

Сумма	503	4,47	363,83	42435	3,798
Средняя	83,833	0,745	60,638	7072,5	0,633

Используя формулу $r = \dfrac{\overline{xy} - \overline{x}\overline{y}}{S_x S_y}$, прежде всего определим S_x и S_y:

$$S_x = \sqrt{\overline{x}^2} - (\overline{x})^2 = \sqrt{7072{,}5 - (83{,}833)^2} = 6{,}673;$$

$$S_y = \sqrt{\overline{y}^2 - (\overline{y})^2} = \sqrt{0{,}633 - (0{,}745)^2 = 0{,}279.}$$

Тогда $r = \dfrac{60{,}638 - 83{,}833 \cdot 0{,}745}{6{,}673 \cdot 0{,}279} = -0{,}976.$

Следовательно, между прибылью (y) и затратами на 1 руб. произведенной продукции (x) существует достаточно тесная обратная зависимость, т.е. фирмы, имеющие большую прибыль, имеют, как правило, меньшие затраты на 1 руб. произведенной продукции.

Рассмотрим на примере трехмерной генеральной совокупности (x_1, x_2, x_3) понятия и правила вычисления частных и множественных коэффициентов корреляции. Пусть каждый экономический объект, элемент генеральной совокупности характеризуется тремя показателями: x_1, x_2 и x_3. Требуется по данным выборки объемом n из генеральной совокупности исследовать взаимосвязь между этими показателями.

В этом случае выборка объемом n будет представлять собой матрицу наблюдений X:

$$X = \begin{pmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ \dots & \dots & \dots \\ x_{i_1} & x_{i_1} & x_{i_3} \\ \dots & \dots & \dots \\ \dots & \dots & \dots \\ x_{n_1} & x_{n_2} & x_{n_3} \end{pmatrix}.$$

В ней каждая i-я строка ($i = 1, 2, \dots, n$) характеризует i-й экономический объект, а столбец, например первый, содержит значение для 1-го показателя для всех n объектов.

По данным первого столбца матрицы X можно определить среднее значение $\overline{x}_1$ и выборочную дисперсию s_1^2 первого показателя:

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_{i_1}; \quad s_1^2 = \frac{1}{n}\sum_{i=1}^{n}(x_{i_1} - \overline{x}_i)^2.$$

Аналогичным образом определяются выборочные характеристики $\bar{x}_2, \bar{x}_3$ и s_2^2, s_3^2.

Рассчитаем выборочные парные коэффициенты корреляции r_{12}, r_{13}, r_{23}.

Частный коэффициент корреляции $p_{12/3}$ характеризует степень линейной зависимости между двумя величинами, например x_1 и x_2, при исключенном влиянии остальных величин, включенных в модель (в нашем случае – это x_3).

Выборочный частный коэффициент корреляции, как выборочный анализ $p_{12/3}$, определяется по формуле

$$r_{12/13} = r\ (x_{1,}\, x_2\ /\ x_3) = \frac{r_{12} - r_{13}r_{23}}{\sqrt{(1-r_{13}^2)(1-r_{23}^2)}}.$$

где r_{12}, r_{13}, r_{23} – выборочные парные коэффициенты корреляции.

В трехмерной модели имеются еще два частных коэффициента корреляции $r_{12/3}$ и $r_{23/1}$, которые рассчитываются аналогично.

Мы имеем два коэффициента корреляции: парный r_{12} и частный $r_{12/3}$, которые характеризуют степень линейной зависимости между величинами x_1 и x_2. Однако если парный коэффициент r_{12} оценивает степень зависимости на фоне влияния x_3, то частный коэффициент корреляции $r_{12/3}$ – при исключенном влиянии x_3.

Таким образом, частный коэффициент корреляции более точно характеризует степень линейной зависимости.

Частный коэффициент корреляции обладает всеми свойствами парного, т.е. изменяется в пределах от –1 до +1. Если частный коэффициент корреляции равен ±1, то связь между двумя величинами функциональная, а равенство его нулю свидетельствует о линейной независимости этих величин.

Множественный коэффициент корреляции, например $p_{12/3}$, характеризует степень линейной зависимости между величиной x_1 и остальными переменными (x_2, x_3), входящими в модель. Он изменяется в пределах от 0 до 1. Равенство его единице свидетельствует о функциональной зависимости между, например, x_1 и остальными переменными (x_2, x_3), входящими в модель, а равенство его 0 свидетельствует об отсутствии линейной зависимости между x_1 и переменными (x_2, x_3).

Выборочный множественный коэффициент корреляции, выборочный аналог генерального коэффициента $p_{1/23}$, можно выразить через парные коэффициенты:

$$r_{1/2,3} = \sqrt{\frac{r_{12}^2 + r_{13}^2 - 2r_{12}r_{13}r_{23}}{1-r}}.$$

В трехмерной модели имеются еще два множественных коэффициента корреляции $r_{2/13}$ и $r_{3/12}$, которые рассчитываются аналогично.

Квадрат коэффициента корреляции называют *коэффициентом детерминации.* При этом *множественный* коэффициент детерминации, например $r^2_{1/23}$, характеризует долю дисперсии x_1, объясняемую влиянием показателей x_2 *и* x_3. Например, если r^2_{1-23} = 0,85, то это свидетельствует, что 85 % дисперсии x_1 объясняется влиянием показателей x_2 и x_3, а 15 % дисперсии x_1 объясняется влиянием факторов, которые не вошли в модель.

Таким образом, коэффициент детерминации r^2_{xy} характеризует долю дисперсии одной величины, например y, объясняемой влиянием фактора x.

Пример. Деятельность структурных подразделений железной дороги (n = 6) характеризуется тремя показателями: x_1 – прибыль (млн руб.), x_2 – затраты на 1 руб. произведенной продукции (коп./руб.) и x_3 – стоимость основных фондов (млн руб.). По данным таблицы 1.21 требуется определить частный $r_{12/3}$ и множественный $r_{1/23}$ коэффициенты корреляции.

Таблица 1.21 – **Исходные и расчетные данные**

Номер структурных подразделений железной дороги (i)	x_{i_1}, млн руб.	x_{i_2}, коп. /руб.	x_{i_3}, млн руб.	$x^2_{i_3}$	$x_{i_1}x_{i_2}$	$x_{i_1}x_{i_3}$	$x_{i_2}x_{i_3}$
1	0,22	96	4,3	18,49	21,12	0,946	412,8
2	1,07	78	5,9	34,81	83,46	6,313	460,2
3	1,00	77	5,9	34,81	77,00	5,900	454,3
4	0,61	89	3,9	15,21	54,29	2,379	347,1
5	0,78	81	4,9	24,01	63,18	3,822	396,9
6	0,79	82	4,3	18,49	64,78	3,397	352,6
Сумма	4,47	503	29,2	145,82	363,83	22,757	2423,9
Средняя	0,745	83,833	4,867	24,303	60,638	3,793	403,983

Воспользовавшись результатами решения, $s_1 = 0{,}279$; $s_2 = 6{,}673$ и $r_{12} = -0{,}976$.

Найдем $s_3 = \sqrt{\overline{x}_3^2 - (\overline{x}_3)^2} = \sqrt{24{,}303 - (4{,}867)^2} = 0{,}784.$

Определим $r_{13} = \dfrac{\overline{x_1x_3} - \overline{x_1}\,\overline{x_3}}{s_1 \cdot s_3} = \dfrac{3{,}793 - 0{,}745 \cdot 4{,}867}{0{,}279 \cdot 0{,}784} = 0{,}764;$

$$r_{23} = \frac{\overline{x_2x_3} - \overline{x_2}\,\overline{x_3}}{s_2 \cdot s_3} = \frac{403{,}983 - 83{,}833 \cdot 4{,}867}{6{,}673 \cdot 0{,}784} = -0{,}771.$$

Частный коэффициент корреляции

$$r_{12/3} = \frac{r_{12} - r_{13}r_{23}}{\sqrt{(1 - r_{13}^2)(1 - r_{23}^2)}} = \frac{-0{,}976 - 0{,}764 \cdot (-0{,}771)}{\sqrt{(1 - 0{,}764^2)(1 - 0{,}774^2)}} = -0{,}948.$$

Сравнивая значения парного r_{12}= −0,976 и $r_{12/3}$= −0,948 коэффициентов корреляции, можно утверждать, что фактор x_3 слабо влияет на степень зависимости между величинами x_1 и x_2.

Определим теперь множественный коэффициент корреляции:

$$r_{1/23} = \sqrt{\frac{r_{12}^2 + r_{13}^2 - 2r_{12}r_{13}r_{23}}{1 - r_{23}^2}} = \sqrt{\frac{(-0,976)^2 + 0,764^2 - 2 \cdot 0,976 \cdot 0,764 \cdot 0,771}{1 - (-0,771)^2}}$$

и окончательно $r_{12/3}$ = 0,976.

Корреляционный анализ порядковых переменных: ранговая корреляция. Порядковая переменная позволяет упорядочивать статистически исследованные объекты по степени появления в них анализируемого свойства. К порядковым переменным обращаются в ситуациях, когда количественно измерить данную степень проявления свойства невозможно или когда измерения рассматриваются как вспомогательное средство для последующего ранжирования рассматриваемых объектов.

Под ранговой корреляцией понимается статистическая связь между порядковыми переменными. Речь идет об измерении статистической связи между двумя или несколькими ранжировками одного и того же конечного множества объектов O_1, O_2, …, O_n.

Ранжировкой называют расположение объектов в порядке убывания степени проявления в них k-го изучаемого свойства. В этом случае $x_i^{(k)}$ называют рангом i-го объекта по k-му признаку. Ранг характеризует порядковое место, которое занимает объект O_i в ряду n объектов.

В случаях неразличимости рангов используют «объединенные» (или «связные») ранги. Всем «связным» рангам присваивается один и тот же ранг, равный средней арифметической от рангов, входящих в данную группу. Например, если в ранжировке объекты, находящиеся на 3–6-м местах неразличимы по данному признаку, то каждому из них присваивается ранг, равный $\frac{3+4+5+6}{4} = 4,5,$ т.е. мы получим последовательность: 4,5; 4,5; 4,5; 4,5.

Пример. Два эксперта проранжировали 10 предложенных проектов реорганизации с точки зрения их эффективности. Ранжировка 1-го эксперта: 1; 2; 3; 4; 5; 6; 7; 8; 9; 10. Ранжировка 2-го эксперта: 2; 3; 1; 4; 6; 5; 9; 7; 8; 10. Вычисления дают результат:
$r_{12}^{(s)} = 1 - \frac{6}{1000 - 10}(1 + 1 + 2^2 + 0 + 1 + 1 + 2^2 + 1 + 1 + 0) = 1 - \frac{6}{990} \cdot 14 = 0,915$, что свидетельствует о положительной ранговой связи между переменными.

Метод наименьших квадратов. Согласно ему минимизируется квадрат отклонения наблюдаемых значений результативного показателя y_i (i = = 1, 2, …, n) от значений $\tilde{y}_i$ = f(x_i), где x_i – значение вектора аргументов в i-м наблюдении: $\sum_{i=1}^{n} y_i - f(x_i)^2 \to \min$. Получаемая регрессия называется *среднеквадратической.*

Метод наименьших модулей. Согласно ему минимизируется сумма абсолютных отклонений наблюдаемых значений результативного показателя от модульных значений $\tilde{y}_i = f(x_i)$. Получаем *среднеабсолютную медианную регрессию* $\sum_{i=1}^{n} |y_i - f(x_i)| \to \min$.

Регриссионным анализом называется метод статистического анализа зависимости случайной величины y от переменных x_j (j = 1, 2, … k), рассматриваемых в регрессионном анализе как неслучайные величины, независимо от истинного закона распределения x_j.

Двумерное линейное уравнение регрессии. Пусть на основании анализа исследуемого явления предполагается, что в «среднем» y есть линейная функция от x, т.е. имеет уравнение регрессии

$$\tilde{y} = \beta_0 + \beta_1 x,$$

где β_0 и β_1 – неизвестные параметры генеральной совокупности, которые надлежит оценить по результатам выборочных наблюдений.

Согласно методу наименьших квадратов в качестве оценок неизвестных параметров β_0 и β_1 следует брать такие значения выборочных характеристик b_0 и b_1, которые минимизируют сумму квадратов отклонений значений результативного признака y_i, от средней.

Пример. По данным годовых отчетов десяти (n = 10) вагоностроительных предприятий провести регрессионный анализ зависимости производительности труда y (тыс. д.е. на 1 чел.) от объема производства x (млн д. е.). Предполагается, что уравнение регрессии линейно и имеет вид $\tilde{y} = \beta_0 + \beta_1 x$,

Исходные данные для анализа представлены в таблице 1.22.

Таблица 1.22 – **Расчет параметров уравнения регрессии**

Номер предприятия (i)	y_i	x_i	x_1^2	$(x_i - x_i)^2$	$\tilde{y}_i$	$e_i = y_i - \tilde{y}_i$	$\overline{\delta_i}$
1	2,1	3	9	20,25	2,77	−0,67	−31,9
2	2,8	4	16	12,25	3,52	−0,72	−25,7
3	3,2	5	25	6,25	4,27	−1,07	−33,4
4	4,5	5	25	6,25	4,27	0,23	5,1
5	4,8	5	25	6,25	4,27	0,53	11,0
6	4,9	5	25	6,25	4,27	0,63	12,9
7	5,5	6	36	2,25	5,02	0,48	8,7
8	6,5	7	49	0,25	5,77	0,73	11,2
9	12,1	15	225	56,25	11,75	0,35	2,9
10	15,1	20	400	156,25	15,50	−0,4	−-2,6
Сумма	61,5	75	835	272,5	–	–	–
Средняя	6,15	7,5	83,5	–	–	–	–

Учитывая, что $\sum_{i=1}^{10} x_i y_i = 666{,}5$, получим:

$$b_1 = \frac{666{,}5 - \frac{1}{10}\cdot 75\cdot 61{,}5}{835 - \frac{1}{10}\ 75^{\ 2}} = \frac{205{,}25}{272{,}5} = 0{,}753;\ b_0 = 6{,}15 - 0{,}753\cdot 7{,}5 = 0{,}502.$$

Таким образом, оценка уравнения регрессии будет иметь вид

$$\hat{y} = b_0 + b_1 x\ .$$

После подстановки окончательно получим:

$$\hat{y} = 0{,}502 + 0{,}753x.$$

Из уравнения регрессии следует, что при увеличении объема производства на единицу его измерения производительность труда в среднем увеличивается на 0,753 тыс. руб.

Для интерполяции модели также можно воспользоваться коэффициентом эластичности, значение которого $e_i = b_i \frac{\bar{x}}{\bar{y}} = 0{,}753\frac{7{,}5}{6{,}15}$ = 0,918 показывает, что при увеличении объема производства x на 1 % производительность труда y в среднем увеличится на 0,918 %.

Перейдем к статистическому анализу полученного уравнения регрессии и рассчитаем исправленную выборочную дисперсию $\hat{s}^2$, абсолютные $e_i = y_i - \hat{y}_i$ и относительные $\delta_i = \frac{e_i}{y_i}100$ % ошибки аппроксимации. Выборочная дисперсия

$$\hat{s}^2 = \frac{1}{n-2}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = 0{,}486.$$

Теперь среднюю относительную ошибку аппроксимации определим по формуле

$$\bar{\delta} = \frac{1}{n}\sum_{i=1}^{n}\left|\delta_i\right| = \frac{1}{10}145{,}4 = 14{,}54\ \%,$$

где $\left|\delta_i\right|$ – абсолютное значение относительной ошибки аппроксимации. Среднее значение относительной ошибки 14,54 % говорит о том, что наша модель достаточно хорошо согласуется с исходными данными.

Самую низкую производительность труда, как следует из таблицы 1.22, имеет третье предприятие. У этого предприятия производительность труда y_3 = 3,2 тыс. руб. на 1 человека, что на 33,4 % ниже того, что имело бы «среднее» предприятие с объемом производства x_3 = 5,0 млн руб.

Лучшим по критерию производительности труда является шестое предприятие, у которого этот показатель на 12,9 % выше среднего значения по совокупности рассматриваемых предприятий с объёмом производства 5 млн д. е.