ТЕОРИЯ СТАТИСТИКИ

МИНИСТЕРСТВО ВЫСШЕГО И СРЕДНЕГО СПЕЦИАЛЬНОГО ОБРАЗОВАНИЯ РЕСПУБЛИКИ УЗБЕКИСТАН

ТАШКЕНТСКИЙ ФИНАНСОВЫЙ ИНСТИТУТ

ТЕОРИЯ СТАТИСТИКИ

Рекомендовано Министерством высшего и среднего специального образования Республики Узбекистан в качестве учебника для студентов высших учебных заведений

ТАШКЕНТ
«IQTISOD-MOLIYA»
2021

УДК: 311(075.8)
ББК: 60.6я7
Т 73

Рецензенты: *д-р экон. наук, проф.* **Б.А. Бегалов;**
д-р экон. наук, проф. **Б.К. Гаибназаров**

Т 73

Теория статистики: Учебник. / Ё.А. Абдуллаев, У.У. Азизов, З.Х. Тошматов, М.М. Икрамов; – Т. «Iqtisod-Moliya», 2021. – 456 с.

В учебнике рассмотрен широкий круг вопросов статистической методологии: организация статистического наблюдения, обработка данных и их анализ. Особое внимание уделено статистическим методам анализа вариационных рядов и рядов динамики, выборочному наблюдению, изучению корреляционных связей, индексному методу.

Авторы постарались изложить учебный материал кратко и в доступной форме с использованием данных статистических ежегодников.

Для студентов, обучающихся по специальности «Статистика», а также для экономических вузов и факультетов.

УДК: 311(075.8)
ББК: 60.6я7

ISBN 978-9943-13-865-0

СОДЕРЖАНИЕ

ПРЕДИСЛОВИЕ

Теория статистики – одна из важных дисциплин в учебном плане экономических вузов, так как статистическая грамотность – неотъемлемая составляющая экономического образования. Работая с цифрами, каждый экономист должен знать, как получены те или иные данные, какова их природа, насколько они полны и достоверны. Кроме того, он должен уметь использовать различные статистические методы анализа массовых явлений. Всему этому учит курс «Теория статистики».

Данный учебник является первой частью учебника «Статистика», подготовленной этими же авторами и соответствует программе рассчитанной на 72 часа (для одного семестра). Изучающим статистику будет удобно получить в библиотеке вначале отдельно учебник «Теория статистики», а затем в следующем семестре получить учебник «Социально-экономическая статистика».

Раздельное издание учебника дает возможность студентам освоить некоторые основные приемы общей теории статистики, а потом перейти к более продвинутому этапу обучения – к системе показателей социально-экономической статистики соответствующей программе рассчитанной на 72 часа (для следующего семестра).

Этот учебник дает представление об основных статистических методах, их возможностях и границах применения. Если вы захотите более глубоко изучить соответствующий раздел статистики, то в конце каждой главы вы найдете список вопросов интеллектуального тренинга, список использованной и рекомендуемой литературы.

Кроме того к данному учебнику отдельно подготовлено и издано учебное пособие под названием «Приложение к учебнику «Теория статистики»», которое состоит из четырех разделов:

1) сборник тестов;
2) практикум;
3) сборник слайдов;
4) глоссарий.

Такой подход отличает учебник от других подобных изданий..

Глава I
ПРЕДМЕТ И МЕТОД СТАТИСТИКИ

Дорожная карта

1.1. Что такое статистика?

Слово «статистика» используется в нескольких значениях: прежде всего, как синоним слова *«данные»*. Именно в этом смысле можно сказать: «статистика рождаемости и смертности в Узбекистане или «статистика преступлений». Статистикой называется *отрасль знаний,* объединяющая принципы и методы работы с числовыми данными, характеризующими массовые явления. Статистикой называют также *отрасль практической деятельности,* направленной на сбор, обработку и анализ статистических данных.

Слово «статистика» происходит от латинского слова *status* – состояние, положение вещей. Первоначально оно употреблялось в значении «политическое состояние». Отсюда итальянское слово *stato* – государство и *statista* – знаток государства. В научный обиход слово «статистика» вошло в XVIII в. и первоначально употреблялось в значении «государствоведение». В настоящее время *статистика может быть определена как собирание, представление, анализ и интерпретация числовых данных.* Это особый метод, который используется в различных сферах деятельности, в решении разнообразных задач.

Исторически развитие статистики было связано с развитием государств, с потребностями государственного управления. Хозяйственные и военное нужды уже в древний период истории человечества требовали наличия данных о населении, его составе, имущественном положении. С целью налогообложения организовывались переписи населения, проводился учет земель и т.д. Первые работы такого рода отмечены даже в священных книгах разных народов. В античном мире был организован учет родившихся; молодые люди, достигшие 18 лет, вносились в списки военнообязанных, а по достижении 20 лет – в списки полноправных граждан. Составлялись земельные кадастры, в которые вносились, сведения о строениях, рабах, скоте, инвентаре, получаемых доходах. Появились описания государств. Большая заслуга в этом принадлежит греческому философу Аристотелю (384-322 г. до н. э.); он составил описание 157 городов и государств своего времени.

С образованием государств появилась необходимость в статистической практике, т.е. в сборе сведений о наличии земель, численности населения, о его имущественном положении. Несколько тысячелетий назад такой учет проводился в Китае, Древнем Риме и в Египте.

В период становления капитализма рост общественного производства, расширение торговых и международных отношений послужили стимулом развития учета и статистики. Наряду с простой бухгалтерией в Италии (приблизительно с начала XIV в.) появляется система двойной бухгалтерии, при которой операция фиксируется дважды – в дебете и кредите. Значительно возрастает потребность в анализе экономической конъюнктуры, поэтому объем статистической информации особенно резко увеличивается; требуются сведения о размерах и размещении промышленного и сельскохозяйственного производства, рынках сбыта товаров, рынках труда, сырьевых ресурсов и т.д.

Расширение практики учетно-статистических работ в различных странах способствовало формированию статистической науки. Статистика как наука стала развиваться с середины XVII в. по двум направлениям: описательному и математическому.

Важнейшими представителями *описательной школы* государствоведения были немецкие ученые Г. Конринг (1606-1681) и Г. Ахенваль (1719-1772).

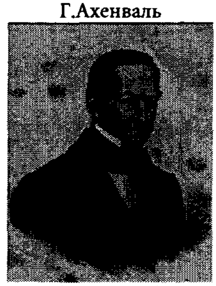
Аристотель

(384-322 г. до н.э.)

Г.Конринг

(1606-1681)

Г.Ахенваль



(1719-1772)

Первой отличительной чертой описательного направления было то, что задачей статистики его представители считали описание «государственных достопримечательностей». К их числу относили территорию государства, государственное устройство, население, религию, внешнюю политику и т.п.

Таким образом, в этом случае предмет статистики не ограничивался теми явлениями, которые имеют числовую характеристику. Более того, ранние представители описательной школы вообще избегали пользоваться числовыми данными и лишь позднее (в середине XVIII в.) числовые данные постепенно завоевали право быть включенными в работы описательной статистики. Вторая особенность описательного направления статистики заключалась в том, что в этих работах отсутствовал анализ закономерностей и взаимосвязей, присущих общественным процессам. Следовательно, то, что представители описательного направления называли статистикой было весьма далеко от действительной статистики в ее современном понимании.

Математическое направление зародилось в Англии. В отличие от описательной школы представители математического направления («политические арифметики») ставили своей задачей выявление закономерностей и взаимосвязей экономических явлений с помощью различных расчетов. Свои выводы они основывали на числовых: данных. Виднейшим представителем этого направления был У. Петти (1623-1687), которого К. Маркс назвал «в некотором роде изобретателем статистики». В дальнейшем это направление значительно развилось в работах Ф. Гальтона (1822-1911), К.Пирсона (1857-1936), Р. Фишера (1890-1962) и др.

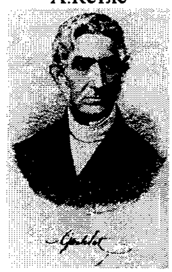
Средневековье оставило уникальный памятник – «Книгу страшного суда» (1061) – это свод материалов всеобщей переписи населения Англии и его имущества (включает данные о 240 тыс. дворов). Со временем собирание данных о массовых общественных явлениях приобрело регулярный характер; с середины XIX в. благодаря усилиям великого бельгийца - математика, астронома и статистика Адольфа Кетле (1796-1874) были выработаны правила переписей населения и регулярность их проведения в развитых странах. Для координации развития статистики по инициативе А. Кетле проводились международные статистические конгрессы, а в 1885 г. был основан Международный статистический институт.

Международной статистикой занимаются международные организации – ООН, ЮНЕСКО, МОТ, ЕС, Мировой банк и др. Международные организации и государственная статистика каждой отдельной страны занимаются сбором, представлением, сравнением, интерпретацией социально-экономических данных. Сложились методы работы, продолжающие традиции государствоведения.

У.Петти

(1623-1687)

А.Кетле



(1706-1874)

М.Ломоносов

(1711-1765)

На Руси уже в X-XII вв. собирались различного рода сведения, связанные с налогообложением. Петровские реформы, затронувшие все стороны общественной жизни, потребовали значительно большего числа точных статистических данных: вводится учет цен на хлеб, городов и городского населения, внешней торговли, регистрация новых фабрик и заводов. В этот же период зарождается текущий учет численности населения – осуществляемая церковью регистрация браков, рождений, смертей. По мере усложнения общественной жизни все более расширялся круг учитываемых явлений.

Становление статистической науки в России началось с развития описательного направления. Среди ярких представителей описательной школы следует назвать И.К. Кириллова (1689–1737), В.Н. Татищева (1686–1750), М.В. Ломоносова (1711–1765), К.Ф. Германа (1767–1838). Одним из первых систематизированных экономико-географических описаний России была работа И.К. Кириллова «Цветущее состояние

Всероссийского государства», написанная в 1727 г. по материалам Петровской ревизии. Много сделал в области статистики и экономической географии историк, географ, энциклопедист В.Н. Татищев. Им была разработана детальная программа для получения сведений, необходимых для составления географии России с полным экономическим описанием. Государствоведение нашло отражение и в работах М.В. Ломоносова. Так, в 1755 г. им была написана книга «Слово похвальное императору Петру Великому», где дана оценка Петровской ревизии. Особой заслугой М.В.Ломоносова является усовершенствование программы обследования, разработанной Татищевым для создания «Атласа Российского», целью которого являлась характеристика географии, населения и экономики страны в разрезе отраслей – сельского хозяйства, промышленности, торговли, транспорта. Бланки обследования, содержащие ее программу, были разосланы в города и уезды; материалы обследования поступали в академию в течение длительного времени и были обработаны уже после смерти М.В. Ломоносова. Несмотря на то, что М.В. Ломоносова относят к школе описательного направления, его работы не носили чисто описательного характера, а содержали и элементы анализа.

Впервые в русской статистической литературе проблемами теории статистики занимался К.Ф. Герман. Свои теоретические взгляды он изложил сначала в большой статье «Теория статистики» в «Статистическом журнале», а затем развил их в книге «Всеобщая теория статистики», изданной в 1809 г. Автор работ примыкал к описательной школе, однако не отрицал необходимости разработки теории статистики. При этом под теорией статистики он понимал не только рассуждения о содержании ее как науки, но и учение о той системе, которой нужно следовать, располагая материал при статистическом описании.

Описательное направление было господствующим в теоретических взглядах русских статистиков вплоть до 30-х годов XIX столетия, когда оно начало постепенно терять под собой почву. Начало критики этого направления было положено выходом в свет в 1838 г. небольшой работы В.С. Порошина (1809-1868) «Критическое исследование об основаниях статистики». Автор высказывает мысль, что наука не может состоять в простом описании фактов или в их систематизации, ее задача заключается в установлении и изучении взаимосвязей и закономерностей явлений. «Как летопись не наука, так описание не статистика», – пишет В.С. Порошин в своей работе. Однако не все русские статистики, писавшие в этот период о содержании своей науки, решительно порывают с описательным направлением. Влияние описательной школы оказалось слишком сильным, чтобы его можно было легко преодолеть. Идеи этой школы в известной мере продолжали сказываться на взглядах русских статистиков-теоретиков даже во второй половине 40-х годов XIX в.

Крупным событием в истории отечественной статистики было появление в 1846 г. работы Д.П. Журавского (1810–1856) «Об источниках и употреблении статистических сведений». В ней сформулированы специфические особенности статистики как науки «категорического исчисления», где массовое наблюдение – основа статистического исследования, а группировка – основной метод статистического анализа. Большое внимание в работе уделено критике источников статистических сведений, вопросам организации их получения, их достоверности. Талантливым свидетельством применения методов статистического анализа явилась другая работа Д.П. Журавского «Статистическое описание Киевской губернии», которую Н.Г. Чернышевский считал «одним из самых драгоценных приобретений, сделанных русской наукой», а ознакомление с ней – полезной для западноевропейских ученых.

Важный этап в развитии статистики и в преподавании этой дисциплины связан с именем профессора Московского университета А.И. Чупрова (1842–1908), издавшего «Курс статистики». Главное значение курса заключается в пробуждении интереса к этой науке, в пропаганде статистических знаний, в их популяризации. Особенно большой заслугой А.И. Чупрова перед отечественной наукой является его долголетняя руководящая роль в организации земской статистики.

Вслед за реформой 1861 г. в России: были **созданы земства – местные органы самоуправления**; их деятельность была ограничена решением хозяйственных нужд уезда или губернии. Для имущественного обложения и преодоления хозяйственных трудностей земства нуждались в статистических данных и поэтому в 70-х годах во многих губерниях европейской части России были созданы статистические бюро, в которых работала разночинная интеллигенция. Земская статистика не только собрала богатый материал о хозяйстве и быте русского народа, но и внесла большой вклад в развитие статистической науки, усовершенствовав методы статистического наблюдения, табличной сводки собранного материала и др. Но одним из существенных ее недостатков была несогласованность программ и методов статистического исследования, затруднявшая обобщение материалов земской статистики в масштабах всей страны.

Особое место в истории русской статистики занимает П.П. Семенов-Тян-Шанский (1827-1914, до 1906 – Семенов). Великий русский географ много сил отдал налаживанию практической статистики в стране.

С 1864 г. он возглавил ЦСК (Центральный статистический комитет) и руководил им в течение 33 лет. За это время была проведена большая работа по упорядочению исследования русского хозяйства: введены подворные обследования, налажена статистика урожаев, проведена первая всеобщая перепись населения 1897 г., перепись паровых двигателей, началось издание справочных материалов по фабрично-заводской статистике. П.П. Семенов-Тян-Шанский был автором многих ценных печатных работ в области статистики.

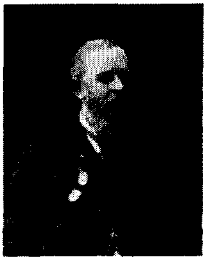
К концу 60-х годов прошлого столетия статистическая наука в России уже настолько окрепла, что в 1872 г. в Петербурге был проведен Международный статистический конгресс. П.П. Семенов сделал на нем доклад о принципах организации переписей населения. Основные положения, высказанные в этом докладе, легли в основу организации переписей многих стран. Участие представителей отечественной науки в международных статистических конгрессах способствовало изучению опыта зарубежной статистики.

Всемирной известностью пользовались работы представителей русской академической статистики – Ю.Э. Янсона (1835-1893) и А.А. Кауфмана (1864-1919).

П.П.Семенов-Тян-Шанский

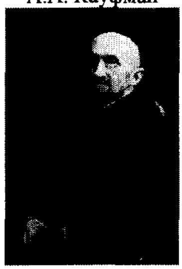
(1827-1914)

Ю.Э. Янсон



(1835-1893)

А.А. Кауфман



(1864-1919)

Профессор Петербургского университета Ю.Э. Янсон в своем учебнике «Теория статистики» обобщил результаты богатейшей русской и зарубежной статистической практики своего времени. Особый интерес представляла и другая его работа «Сравнительная статистика России и западноевропейских государств». В работах А.А. Кауфмана достаточно полно излагалась история статистической мысли в России.

К концу XIX в. Россия превратилась в один из признанных центров научной статистической мысли.

XX столетие характеризуется дальнейшим развитием практической и научной деятельности статистиков в России. Вызывалось это ростом производительных сил и потребностью в объективных данных для организации практической деятельности в области управления и планирования. В начале века отмечалось интенсивное развитие математической статистики и применение ее аппарата в практической деятельности.

Появляются специальные исследования о кривых распределения (А.В. Леонтович), о корреляционном анализе (Е.Е. Слуцкий, А.А. Чупров). Вопросы теории статистики получают глубокую трактовку в трудах ученых, работы которых широко известны во всем мире (В.С. Немчинов, С.Г. Струмилин, Б.С. Ястремский, А.Я. Боярский и др.).

В.С. Немчинов

(1835-1893)

С.Г. Струмилин

(1877–1974)

Историческое развитие статистики России обобщено в трудах В.И. Хотимского, В.С. Немчинова, В.Н. Старовского, М.В. Птухи и др. Совершенствование методологии статистического изучения социально-экономических явлений нашло отражение в работах видных российских, ученых: А.И. Ротштейна, Д.В. Савинского, А.И. Гозулова, П.П. Маслова, Н.М. Виноградовой, Т.В. Рябушкина, В.Е. Адамова, Е.Н.Фреймунда и др.

В.Н.Старовский

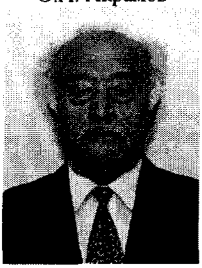
(1905–1975)

Т.В. Рябушкин

(1915-1986)

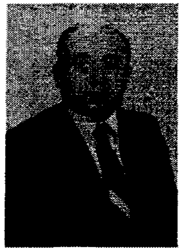
Развитие теории и практики Узбекистана обобщено в трудах В.Н. Крылова, Ф.Ф. Коррель, Э.А. Акрамова, Н.М. Сагатова.

Э.А. Акрамов



(1931)

Н.М. Сагатов



(1934-2013)

Совершенствование методологии экономико-статистического изучения социально-экономических явлений нашло отражение в работах профессоров С.Сирожиддинова, М.Хамраева, Ё.Абдуллаева, З.Ташматова, Р.Алимова, И.Эрматова, Х.Набиева, Х.Шодиева, доцентов А.Аюбжанова, Б.Усманова, А.Набиходжаева, Б.Казыбекова, Х.Хужакулова и др.

1.2. Предмет статистической науки

Статистика как наука

Каждая наука обладает существенными специфическими особенностями, которые отличают ее от других наук и дают ей право на самостоятельное существование как особой отрасли знания. Главная особенность любой науки заключается в предмете познания, в принципах и методах его изучения, которые в совокупности образуют ее методологию.

Предметом исследования статистики являются массовые явления. социально-экономической жизни; она изучает количественную сторону этих явлений в неразрывной связи с их качественным содержанием в конкретных условиях места и времени.

Особенностями статистической науки являются то, что предметом ее изучения выступают совокупности – множества однокачественных, варьирующих явлений. В это определение входят три основные черты совокупности любых явлений: **во-первых**, - это множество явлений, т.е. предметом статистического изучения всегда выступают совокупности тех или явлений, включающие все множества проявлений исследуемой закономерности.

Во-вторых, это массовые явления, объединенные общим качеством, представляющие собой проявления одной и той же закономерности. То есть свойство статистических закономерностей проявляется лишь в массе явлений при обобщении данных по достаточно большому числу единиц; оно получило название **закона больших чисел**.

Сущность закона заключается в следующем: при суммировании данных по достаточно большому числу случаев (единиц статистической совокупности) различия отдельных единиц изучаемой массы случаев взаимопогашаются (взаимоуравновешиваются) и в общих средних числах выступают существенные, характерные черты и взаимосвязи явления в целом, т.е. совокупное действие большого числа случайных факторов приводит к результату, почти независящему от случая.

Следовательно, основное содержание закона больших чисел состоит во взаимном погашении индивидуальных отклонений от некоторого уровня, характерного для всей совокупности в целом. Именно в результате этого взаимопогашения и проявляется закономерность. Поэтому в основе статистического исследования всегда лежит массовое наблюдение фактов. Чем больше объем наблюдаемых единиц, тем ближе наблюдаемые средние величины воспроизводят закономерности изучаемого явления,

В-третьих, это множество *варьирующих* явлений, отличающихся по своим характеристикам. Именно последнее свойство выступает необходимостью изучения всего множества явлений одного вида. Если бы единицы совокупности были полностью тождественны друг другу, то не было бы потребности обращаться к множеству единиц, достаточно лишь изучить одну единицу, чтобы знать все о всех явлениях этого вида.

1.3. Методы статистической науки

Статистическая методология

Особенность предмета статистики заключается в изучении массовых варьирующих явлений, определяет специфику *статистического метода.* Очевидно, что нужно собрать данные о признаках всех единиц совокупности, обобщить их и выразить в сводных показателях. Отсюда статистический метод включает:

- *наблюдение* (сбор первичных данных);
- *обобщение собранных данных* (их группировка, расчет сводных показателей);
- *представление результатов обобщения в* форме статистических таблиц графиков с текстовыми пояснениями.

Только с помощью такого метода можно раскрыть закономерности развития изучаемых явлений.

Таким образом, специфический метод статистики основан на соединении анализа и синтеза. Сначала выделяются в составе изучаемого явления и раздельно изучаются части (группы и подгруппы), оценивается существенность или несущественность наблюдаемых различий в величине признака, выявляются причины различий, а затем дается характеристика явления в целом, во всей совокупности его сторон, тенденций и форм развития. Все стадии статистической работы тесно связаны друг с другом; недостатки, возникающие на одной из них, сказываются на всем исследовании в целом. Поэтому строгое соблюдение правил статистической науки обязательно на всех стадиях статистического исследования.

***Методология статистики** – совокупность приемов, применяемых в процессе статистического исследования.*

Методологической *основой* является *диалектический метод познания,* при котором объекты исследования изучаются в развитии и взаимосвязи. *При этом статистика опирается на такие диалектические категории, как:*

- ✓ количество и качество;
- ✓ необходимость и случайность;
- ✓ причинность;
- ✓ закономерность;
- ✓ единичное и массовое;
- ✓ индивидуальное и общее.

При изучении разных объектов в разных задачах, конечно же, используются различные методы. Тем не менее, существуют некоторые, общие принципы и методы статистической работы. В учебнике «Теория статистики» английских статистиков Дж. Э. Юла и М. Дж. Кендэла говорится: «Независимо от того, в какой отрасли знания получены числовые данные, они обладают определёнными свойствами, для выявления которых может потребоваться особого рода научный метод обработки. Последний известен как статистический метод или, короче, статистика».

Статистические методы включают как простые методы, которые могут быть понятны любому человеку, так и сложные математические процедуры, доступные специалистам высокого класса. В этом учебнике излагаются простые и часто используемые методы при изучении социальных и экономических явлений и процессов.

В процессе исследования своего предмета статистика может опираться и на другие общенаучные методы, такие, как:

- ✓ *аналогия* (перенесение свойств одного предмета на другой);
- ✓ *синтез* (объединение отдельных фактов);
- ✓ *выдвижение гипотезы* (научно обоснованное предположение о возможных причинных связях между явлениями) и др.

Все статистические методы используются комплексно, в неразрывной связи друг с другом.

1.4. Категории статистики

1. Основными *статистическими категориями* *т.е. понятиями, используемыми в статистической деятельности) являются:*

- ✓ статистическая закономерность;
- ✓ статистическая совокупность;
- ✓ признак;
- ✓ вариация;
- ✓ статистический показатель.

2. Статистическая закономерность – это *количественная закономерность изменения в пространстве и во времени* массовых явлений и процессов общественной жизни, состоящих из *множества элементов – единиц совокупности.*

Статистическая закономерность проявляется:

✓ только при достаточно большом количестве наблюдений;

✓ только в массе единиц (свойственна не отдельным единицам, а всей совокупности).

3. Статистическая совокупность (объект конкретного статистического исследования) – это *совокупность социально-экономических объектов или явлений общественной жизни (единиц совокупности):*

✓ *объединенных общей основой и системой связей;*

✓ но *отличающихся* друг от друга *отдельными признаками.*

Так, при определении объема розничного товарооборота все предприятия торговли, осуществляющие продажу товаров населению, рассматриваются как единая статистическая совокупность розничной торговли".

Примерами статистической совокупности могут быть:

➢ совокупность жителей Ташкентской области по состоянию на 1 января 2019 г.;

➢ совокупность студентов ТФИ в 2019/20 учебном году и др.

Единицы совокупности неделимые первичные элементы, выражающие основные признаки совокупности.

Пример: студенты – статистическая совокупность, объединенная одинаковым родом деятельности (обучение), но различающаяся разной успеваемостью, курсом, общественной деятельностью.

При проведении переписи торгового оборудования единицей наблюдения является торговое предприятие, а единицей совокупности – их оборудование (прилавки, холодильные агрегаты и т.д.).

Количество всех единиц совокупности составляют *объем статистической совокупности (n)*. Каждая единица совокупности имеет определенные признаки.

Основные свойства статистической совокупности:

- *неделимость* на более мелкие категории (например, население страны остается населением, несмотря на постоянно происходящие процессы естественного и механического движения, т.е. дальнейшее деление статистической совокупности не меняет ее качественных характеристик);
- *однородность* – объединение всех единиц совокупности общим свойством, основой. Например, множество предприятий, производящих однородную продукцию, относятся к одной отрасли, но различаются объемом производства, оборотными средствами, прибылью и т.п.

4. *Признак* – это *свойство, характерная черта или иная особенность единиц, объектов или явлений, которые могут быть наблюдаемы или изучены.*

Например, такая единица совокупности как студент имеет множество признаков:

- ✓ факультет (специальность);
- ✓ курс:
- ✓ успеваемость.

Значения признака у каждой единицы совокупности могут как совпадать, так и быть отличными друг от друга.

5. Вариация – это *измеряемость, многообразие величины признака у отдельных единиц совокупности, количественные изменения значений признака при переходе от одной единицы совокупности к другой,* складывающиеся под воздействием случайных причин или факторов.

Статистический показатель – это *количественная оценка свойства (единицы совокупности) изучаемого объекта или явления.* Здесь проявляется единство качественной и количественной сторон.

<u>*Статистические показатели делятся:*</u>

➢ на *первичные* (объемные, количественные, экстенсивные) – характеризуют либо общее число единиц совокупности, либо сумму значений какого-либо признака (общая численность студентов вуза, объем продаж за год и др.);

➢ *вторичные* (производные, качественные, интенсивные) – производные показатели (повышение производительности труда, конкурентоспособности предприятия, ликвидности предприятия и др.);

➢ *синтетические* – характеризуют сложный комплекс социально-экономических явлений и процессов (ВВП, НД и др.).

Так как все свойства статистической совокупности взаимосвязаны, то статистические показатели, характеризующие эти свойства, составляют *систему статистических показателей – совокупность взаимосвязанных показателей, объективно отражающая существующие между явлениями взаимосвязи.*

<u>*Связь между показателями в пределах одной системы может быть:*</u>

✓ *детерминированной,* т.е. *жестко установленной* (связь прибыли с объемом произведенной продукции);

✓ *стохастической,* т.е. *свободной, зависящей от множества различных факторов* (количество внесенных удобрений на одном участке напрямую не предопределяет урожайность ввиду действия других факторов, например, погодных).

1.5. Структура статистической науки

Статистика многоотраслевая наука

В ходе исторического развития статистической науки в ее составе обособился ряд самостоятельных статистических дисциплин; это объясняется наличием конкретного предмета исследования и особой системы статистических показателей для его характеристики.

Статистика состоит из следующих основных разделов:

✓*общая теория статистики* – наука о наиболее общих принципах, правилах и законах цифрового освещения социально-экономических явлений. Это *методологическая основа всех отраслевых статистик. По функциональному признаку статистическая наука делится на 2 больших раздела:*

• *социальная статистика* – наука, изучающая население, а также явления и процессы, характеризующие условия жизнедеятельности людей;

• *экономическая статистика* – наука, изучающая явления и процессы в области экономики. Она анализирует такие показатели, как национальное богатство (НБ), национальный доход (НД), валовой внутренний продукт (ВВП), валовой национальный продукт (ВНП) и др.;

✓*отраслевые статистики, т.е. статистика:*

• здравоохранения;
• труда;
• науки;
• культуры;
• уровня жизни;
• права;
• промышленности;
• строительства;
• сельского хозяйства;
• транспорта и т.д.

Каждая отраслевая статистика формируется на основе показателей социальной и экономической статистики, которые, в свою очередь, опираются на общую теорию статистики. Таким образом, статистика предстает как единая система, структуру которой можно представить следующим образом (рис. 1.1).

Составные части статистической науки

Общая теория статистики	**Социальная статистика**	**Экономическая статистика**	
		Макроэкономическая статистика	**Микроэкономическая статистика**
• История статистики • Математическая статистика • Теория статистики	• Статистика населения • Судебная статистика • Статистика труда • Статистика здравоохранения • Статистика образования и т.д.	•Международная статистика • Региональная статистика • Статистика рыночной экономики • Статистика финансов • Налоговая статистика и т.д.	• Статистика промышленности • Статистика сельского хозяйства • Статистика торговли • Статистика транспорта и связи • Статистика малого бизнеса и т.д.

Рис. 1.1. **Составные части статистической науки**

Отраслевые статистики формируются на базе показателей экономической или социальной статистики, те и другие основываются, в свою очередь, на категориях (показателях) и методах анализа, разработанных общей теорией статистики.

Общая теория статистики является той учебной дисциплиной, с изучения которой начинается формирование необходимых профессиональных знаний у экономистов, менеджеров, руководителей предприятий.

1.6. Взаимосвязь статистики с другими науками

Место статистики в системе наук

Статистика взаимосвязана со многими другими науками. В ней очень много расчетов поэтому она имеет тесные связи с математикой. Статистика опирается на экономическую теорию и сама «подпитывает» ее. Порой трудно найти границу между статистикой и анализом финансово-хозяйственной деятельности, который, в свою очередь, неразрывно связан с бухгалтерским учетом (рис. 1.2). Вообще все общественные науки в той или иной мере связаны со статистикой, используют ее методы и результаты статистических исследований.

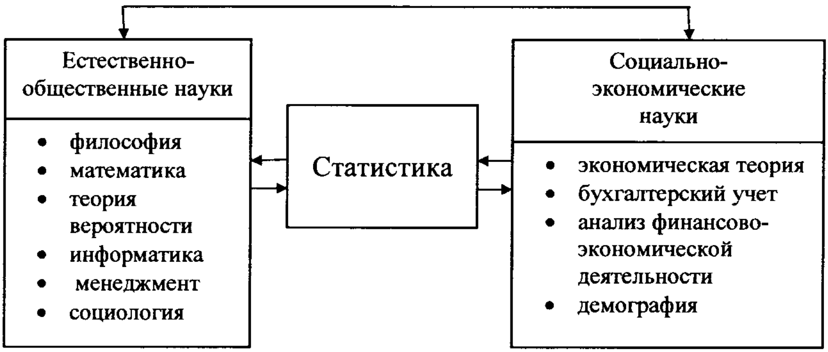


Рис. 1.2. **Укрупненная схема взаимосвязей статистики с другими науками**

Из специфики предмета статистики следует, что теоретической основой статистической науки являются положения исторического материализма и экономической теории, которые исследуют и формируют законы развития социально-экономических явлений, выясняют их природу и значение в жизни общества. Опираясь на знание положений экономической теории, статистика формирует статистические совокупности, устанавливает существенные признаки для выделения социально-экономических типов, осуществляет разработку адекватных методов их изучения.

Руководствуясь положениями экономической теории, статистика обогащает экономические науки фактами, полученными в статистическом исследовании, подтверждает или отрицает их теоретические догмы.

Экономическая теория, опираясь на статистику, формулирует законы развития социально-экономических явлений.

Статистика, характеризуя количественную сторону общественных явлений в конкретных исторических условиях, создает фундамент из точных и бесспорных фактов. Экономические науки используют статистическую информацию для проверки, обоснования или иллюстрации своих теоретических положений.

Статистика тесно связана с другими видами учета общественных явлений – бухгалтерским и оперативно-техническим учетом. Предметом бухгалтерского учета является производственный процесс отдельного предприятия. Бухгалтерия специфическими методами описывает кругооборот средств предприятия по источникам их образования и по направлению их использования. Бухгалтерский учет имеет целью описание экономических отношений между предприятием и всеми связанными с ним контрагентами. Для этого теория бухгалтерского учета выработала определенные научные принципы счетоводства, которые позволяют точно и объективно отобразить сложную структуру средств предприятия, их кругооборот и многообразные связи с другими предприятиями, учреждениями и лицами.

Все это бухгалтерия выражает в форме баланса предприятия. Статистика широко использует данные бухгалтерского учета в своих исследованиях. Так, на основании бухгалтерских данных статистика распределяет предприятия промышленности по величине основных фондов, величине прибыли и т.д.

Оперативно-технический учет также фиксирует на отдельных предприятиях единичные факты, относящиеся к экономике и технологии. Примером оперативно-технического учета может служить учет аварий оборудования или тех или иных неполадок в его работе. Данные оперативно-технического учета используются для оперативного воздействия определенного органа или работника предприятия на соответствующий участок производственной работы (ремонт станка, устранение причин его остановки и т.п.).

Характерным для оперативно-технического учета является то, что в центре его внимания всегда находятся единичные факты, учитываемые, как правило, в натуральном выражении.

В отличие от бухгалтерского и оперативно-технического учета, статистика изучает единичные факты или объекты как элементы массовых общественных явлений. Общей задачей статистики служит выявление закономерностей массовых общественных явлений.

1.7. Современная организация статистики Республики Узбекистан

Изучением экономического и социального развития страны, отдельных ее регионов, отраслей, объединений, фирм, предприятий занимаются специально созданные для этого органы, совокупность которых называется **статистической службой**.

Организация государственной статистики в стране, ее задачи видоизменялись в соответствии с изменением органов государственного управления, их функций, с учетом особенностей развития экономики и социальной жизни общества.

Первый государственный статистический орган России был создан в 1811 г. при департаменте полиции. Статистическое отделение сводило отчеты губернаторов и вело демографическую статистику. Органов для сбора первичной информации не существовало, отчеты губернаторов основывались на донесениях полицейских чиновников, церковных записях о рождениях, смертях, браках и т.п. В 1834 г. было образовано статистическое отделение при Министерстве внутренних дел; в 1852 г. оно было преобразовано в Статистический комитет, а спустя пять лет в 1857 г. – в Центральный статистический комитет (ЦСК) при Министерстве внутренних дел. Разработка методологии статистических исследований возлагалась на статистический совет при этом министерстве. В качестве местных органов правительственной статистики работали губернские статистические комитеты, а в 70-х годах были созданы земские статистические бюро. Земская статистика дала *много ценного для совершенствования статистической* практики и науки.

В XIX в. и вплоть до 1917 г. сбор и анализ статистических данных по стране стал одним из основных направлений деятельности Министерства внутренних дел царской России.

После революции 1917 г. (25 июля 1918 г.) правительством было принято положение о государственной статистике, в соответствии с которым был создан высший орган государственной статистики – Центральное статистическое управление (ЦСУ). На базе бывших статистических органов земств, губернских статистических комитетов на местах были созданы губернские, уездные и городские бюро и отделы. В основу создания органов государственной статистики этим положением был положен принцип централизованного руководства статистикой на базе единства ее организации и методологии.

В последующие годы организация статистики претерпела ряд изменений, так в 1930 г. ЦСУ было передано в ведение Госплана СССР и в 1931 г. переименовано в Центральное управление народнохозяйственного учета (ЦУНХУ) при Госплане СССР. Это слияние органов статистики и планирования объяснялось необходимостью укрепления планового начала в управлении хозяйством страны. В 1932 г. создается сеть районных и городских инспекций, ведающих учетом и статистикой на территории района, города. В 1941 г. ЦУНХУ было переименовано в Центральное статистическое управление Госплана СССР.

В 1948 г. была проведена следующая крупная реорганизация органов статистики, когда Центральное статистическое управление Госплана СССР было переименовано в Центральное статистическое управление при Совете Министров СССР, а в 1978 г. в Центральное статистическое управление СССР.

Особое значение государственная статистика приобрела в советскую эпоху. В условиях плановой экономики статистические данные использовались *при составлении пятилетних планов.* Органы статистики (ЦСУ и др.), наряду с Госпланом, играли в этом ключевую роль.

Использование статистических данных также было одним из излюбленных приемов советской *пропаганды*. Статистические данные наглядно демонстрировали достижения советской экономики. За точку отсчета, как правило, брались итоги предыдущего года, предыдущей пятилетки и особенно часто 1913 г. (последнего "мирного" года дореволюционной эпохи).

Статистические данные также активно использовались *в работе правоохранительных органов*.

Однако главную роль статистика играла именно в экономике. Обширные данные статистики были одним из главных критериев, по которым оценивалось экономическое развитие отраслей, регионов, экономики в целом, составлялись планы дальнейшего экономического развития;

После распада СССР в 1991 г. роль статистики стала иной.

✓ статистика *в меньшей степени влияет на экономическое планирование* (в условиях рыночной экономики);

✓ но она *остается важнейшим критерием оценки состояния дел* в самых разных сферах жизни общества (экономика, борьба с преступностью, развитие регионов, демография и др.).

Важнейшая задача современных статистических; органов – регулярное информирование общественности и органов государственного управления о происходящих изменениях в социально-экономическом развитии страны. В связи с этим возникает необходимость достоверной и своевременной статистической информации.

Для современной государственной статистики характерно:

- ✓ централизованное управление;
- ✓ единое организационное строение и методология;
- ✓ неразрывная связь с органами государственной власти.

В настоящее время система органов государственной статистики РУз *состоит из 3 уровней. Ее можно представить следующим образом (рис 1.3):*

Рис. 1.3. **Органиазция статистики в РУз**

Как видно, организация статистики в республике представляет собой иерархическую систему, верхним звеном которой является Государственный комитет по статистике Республики Узбекистан (РУз) и Республики Каракалпакстан (РКК) – методологический и организационный центр по сбору, обработке и анализу статистических данных на государственном уровне.

Промежуточные звенья данной системы областные статистические управления статистики г. Ташкента. Низовыми звеньями статистики являются районные отделы статистики.

Данные, полученные и обобщенные Госкомитетом РУз и РКК, представляются органам государственной власти для реализации целей контроля и управления, а также публикации для широкого использования.

1.8. Ведомственная статистика

Наряду с общегосударственной статистикой существует ведомственная статистика, ведущаяся на предприятиях, в объединениях, ведомствах, министерствах. Ведомственная статистика выиолняет работы, связанные с получением, обработкой и анализом статистической информации, необходимой для руководства и планирования их деятельности. Для ведения статистики на предприятиях, в объединениях, концерна, ассоциациях, министерствах созданы те или иные статистические органы (ячейки). На отдельных предприятиях статистическую работу может вести один человек, даже по должности не статистик; в крупных объединениях, министерствах имеются специальные отделы, управления.

Значение ведомственной статистики в настоящее время значительно возросло в силу того, что развитие рыночной экономики, самостоятельность предприятий и полная ответственность за результаты производственно-хозяйственной деятельности требуют более глубокого анализа экономических процессов, происходящих на предприятиях. Такой анализ нуждается в обширной статистической информации, которая может быть получена не только на основе первичного учета, ведущегося на предприятиях, но и дополнительно путем специальных обследований, использования нормативных и информационных материалов, в частности, информации ЕГРПО (Единый государственный реестр предприятий).

Цель создания ЕГРПО – это обеспечение единого государственного учета предприятий и организаций, формирование информационного фонда.

Информационный фонд состоит из четырех разделов:

1) **идентификационный** – регистрационный код объекта, являющийся уникальным для всего информационного пространства России;

2) **классификационный** – сведения об отраслевой, территориальной принадлежности субъекта, его подчиненности, виде собственности, организационной форме;

3) **справочный** – фамилия руководителя, адрес объекта, номера телефонов, факсов и т.д.; сведения об учредителях;

4) **экономический** – показатели, характеризующие субъект (среднесписочная численность работников, стоимость основных средств, уставный фонд, балансовая прибыль и др.).

Первые три раздела реестра заполняются в процессе государственной регистрации; источником формирования четвертого раздела является квартальная и годовая отчетность, представляемая в региональные органы статистики.

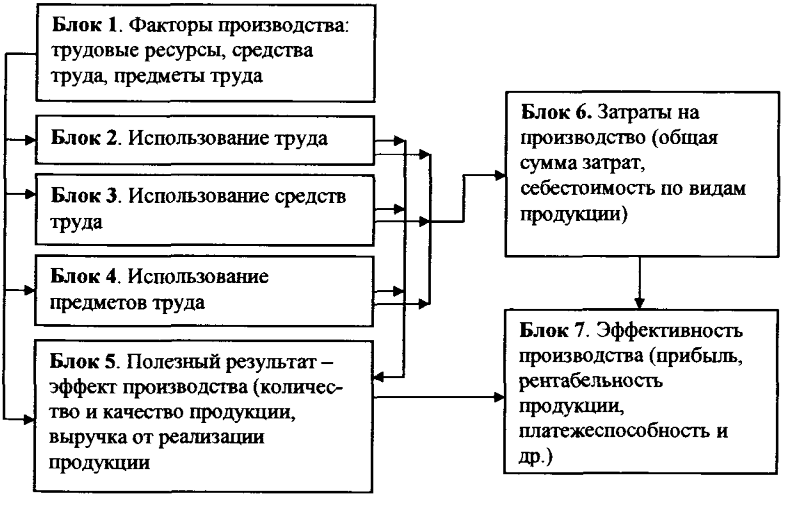


Рис. 1.4. **Блок-схема формирования показателей производственной деятельности предприятия**

Главная задача ведомственной статистики заключается в обеспечении информацией, характеризующей выполнение внутрифирменных (внутрипроизводственных) планов, наличие внутрипроизводственных резервов увеличения выпуска продукции, улучшения использования производственного потенциала. На рис. 1.4 в общем виде представлено формирование показателей производственной деятельности предприятия (объединения, концерна) любой формы собственности.

Благодаря статистике управляющие органы получают всестороннюю характеристику управляемого объекта – отрасли экономики, предприятия или его отдельных подразделений. Статистика системой своих показателей выражает результаты их работы за истекший период, осуществляет контроль за выполнением плана, что требует единства методологии показателей в статистике и планировании.

1.9. Задачи статистики

Статистика как самостоятельная наука ставит перед собой следующие основные задачи:

✓ **общие:**

• *обобщение и прогнозирование тенденций развития,* как отдельных сфер, так и всего народного хозяйства;

• *разработка и внедрение современных методов исследований* экономических и социальных процессов, происходящих в обществе;

• *определение и выявление имеющихся резервов эффективности* как отдельных сфер деятельности, так и всего общественного производства;

• *постоянное обеспечение органов государственной власти и местного самоуправления достоверной информацией,* необходимой для принятия решений;

✓ **специальные:**

• изучение; уровня и структуры массовых социально-экономических явлений;

• рассмотрение *взаимосвязи* массовых социально-экономических явлений и процессов;

• анализ *динамики* массовых социально-экономических явлений.

Основными задачами государственной статистики согласно Указу Президента Республики Узбекистан от 24.12.

2002 года «О государственной статистике Республики Узбекистан» являются:

➡ сбор, обработка, накопление, хранение, обобщение, анализ и публикация статистической информации о социально-экономических явлениях, процессах и их результатах;

➡ обеспечение единой статистической методологии, соответствующей международным стандартам;

➡ обеспечение государственных органов и органов самоуправления граждан, юридических лиц, государственных учреждений и международных организаций и общественности статистической информацией в установленном порядке;

➡ ведение системы экономико-статистических классификаторов, необходимых для организации статистических работ и Единого государственного регистра предприятий и организаций;

➡ обеспечение выполнения Программы государственных статистических работ;

➡ обеспечение совместимости информационной системы государственной статистики и её взаимодействие с другими государственными информационными системами в едином информационном пространстве Республики Узбекистан;

➡ организация государственных статистических наблюдений за ходом выполнения государственных программ социально-экономического развития страны, её регионов, отраслей и секторов экономики;

➡ организация статистических обследований домашних хозяйств, проведение переписей, одноразовых учётов, опросов, выборочный монографический и иных обследований;

➡ проведение международных и межрегиональных статистических сопоставлений;

➡ соблюдение 10 основных принципов государственной статистики: достоверность, объективность, беспристрастность, актуальность, сопоставимость, стабильность, доступность, прозрачность, открытость, конфиденциальность.

Целью нового Закона «О государственной статистике» и «О реорганизации государственной статистики» является создание правовой основы функционирования статистики на качественно новом этапе развития экономики и интеграции в мировое сообщество.

Согласно Указу Президента о «Реорганизации статистики», главными органами статистики являются Госкомстат Республики Узбекистан и статистический совет, который является коллегиальным совещательным органом по проблемам развития, функционирования и координации государственной статистики при уполномоченном органе государственной статистики. Этот совет образован в соответствии с новым Законом «О государственной статистике» от 12.12.2002 года и Постановлением Кабинета Министров от 08.01.2003 года «Об реорганизации деятельности Госкомстата Республики Узбекистан». Статистический совет может назначать экспертные комиссии для рассмотрения отдельных вопросов государственной статистики. На заседания Статистического совета и экспертных комиссий могут приглашаться эксперты и заслушиваться представители министерств, государственных комитетов и ведомств по вопросам государственной статистики. Состав Статистического совета утверждается Президентом Республики Узбекистан.

Важнейшая *задача* статистических органов – регулярное информирование общественности и органов управления о происходящих изменениях в социально-экономическом развитии страны. В связи с этим кардинальный вопрос работы системы органов Госкомстата РУз это – обеспечение достоверности статистических данных, их надежности, подход к использованию языка статистической информации как средства международного общения.

Интеллектуальный тренинг

1. Что означает термин «Статистика»?

2. Когда употребляют слово «статистика» с тем или иным эпитетом – красноречивая, удручающая и обнадеживающая и т.д., – имеют в виду те или иные статистические данные, способные вызвать определенные эмоции. В этом смысле упот-

ребил слово «статистика» американский экономист Митчелл: «Статистика – это, солома, которую я, как и всякий другой экономист, должен спрессовать, чтобы получить брикеты». Когда итальянский статистик К. Джини пишет, что статистика – это царица не только полной, но и неполной индукции, он имеет в виду не статистическую информацию, а статистический метод умозаключений. Мы читаем у Гегеля, что в статистике *числа, которыми она занимается, также интересны лишь* их качественные результаты. Голые числовые изыскания без указанной здесь руководящей точки зрения справедливо считаются предметом пустого любопытства, который не может удовлетворить ни теоретического, ни практического интереса, и нам ясно, что здесь речь идет о конкретном статистическом исследовании той или иной проблемы. Не так ли?

3. Чем обусловлено возникновение и развитие статистической практики и науки?

4. Какие ученые являлись важнейшими представителями описательной школы государствоведения?

5. Какие ученые являлись важнейшими представителями школы математического направления?

6. Нередко мы слышим и читаем: «... эти цифры говорят о...», «...как говорят нам статистические цифры...». В чем смысл подобных оборотов речи? Ведь на самом деле, конечно, говорят не цифры, вернее, не числа, состоящие из этих цифр, а сам читатель, лектор, диктор.

Когда же мы употребляем выражение «цифры рассказывают», то имеем в виду вовсе не математические свойства самих чисел. Цифры рассказывают нам о жизни нашей страны, развитии народного хозяйства, людях, природе – о мире, в котором мы живем и трудимся. Значит, имеется ввиду не само по себе какое-нибудь число, а содержание выраженного в этом числе явления, процесса. Не так ли?

7. Что является предметом исследования статистической науки? Приведите примеры явлений общественной жизни, изучаемых статистикой.

8. Дайте определение категориям статистики, перечислите их виды.

9. В чем заключается сущность статистической методологии?

10. Перечислите стадии статистического исследования, раскройте их основное содержание.

11. Какова роль и значение экономической теории, философии и математики в статистическом исследовании?

12. Перечислите части (разделы) статистической науки и объясните, чем вызвано выделение самостоятельных статистических дисциплин.

13. Каковы принципы организации статистики в Республике Узбекистан в настоящее время?

14. Каковы задачи государственной статистики в условиях перехода к рыночной экономике?

15. Каковы задачи ведомственной статистики и чем объясняется возрастание ее роли в современных условиях?

Использованная и рекомендуемая специальная литература

1. Герман К. Всеобщая теория статистики. СПб., 1809.

2. Дружинин Н.К. Развитие основных идей статистической науки. М.: Статистика, 1979.

3. Кимбл Г. Как правильно пользоваться статистикой. М.: Финансы и статистика, 1982.

4. Льесс А. Статистика. Ее трудности, приемы и результаты. СПб., 1903.

5. Петти В. Экономические и статистические работы. М.: Соцэгиз, 1940.

6. Янсон Ю. Теория статистики. СПб., 1913.

Глава II
СТАТИСТИЧЕСКОЕ НАБЛЮДЕНИЕ

Дорожная карта

2.1. Формирование информационной базы статистического исследования

Как уже отмечалось, статистическая работа состоит в том, чтобы собрать числовые данные о массовых явлениях, обработать их, представить в форме, удобной для анализа, проанализировать и интерпретировать полученные результаты.

Собирание данных лежит в основе всего исследования. От качества используемых данных, от их достоверности и точности зависит достоверность результатов анализа. Люди по-разному относятся к статистической информации: одни не воспринимают ее, другие безоговорочно верят, третьи согласны с мнением английского политика Б. Дизраэли (1804-1881): «Есть ложь, есть наглая ложь, а есть статистика». Однако ему же принадлежит следующее утверждение: «В жизни, как правило, преуспевает больше тот, кто располагает лучшей информацией». На основе статистической информации правительство разрабатывает свою экономическую и социальную политику, оценивает ее результаты, составляет экономические прогнозы. Статистическая информация обеспечивает подготовку двухсторонних и многосторонних экономических соглашений между государствами.

Статистика дает информацию для решения региональных задач, для предпринимательской деятельности – об уровне цен на товары в разных регионах, объемах реализации товаров, условиях кредитования, уровне и темпах инфляции, занятости и т.д.; наконец, в той или иной степени статистика нужна каждому из нас для принятия решений по выбору. стратегии поведения. Для этих целей специальный статистический аппарат занимается систематическим сбором данных, их обработкой и представлением результатов в виде статистической информации государственным и другим органам, коммерческим пользователям. Она формируется в результате статистического наблюдения, которое является начальной стадией экономико-статистического исследования.

2.2. Понятие и формы статистического наблюдения

Статистическим наблюдением называется планомерный, научно организованный сбор данных о явлениях и процессах общественной жизни путем регистрации характеризующих их признаков.

В задачи статистического наблюдения входят: получение достоверной исходной информации, обеспечение полноты информации, проведение статистического наблюдения в возможно короткие сроки.

Статистическое наблюдение осуществляется в трех формах: путем предоставления отчетности, проведения специально организованных статистических обследований и составление регистров (рис. 2.1).

Рис. 2.1. **Формы статистического наблюдения**

Отчетность Это форма статистического наблюдения, при которой сведения поступают в статистические органы в виде обязательных отчетов о деятельности предприятий, учреждений и организаций по заранее установленным программам и в определенные сроки.

Бланки (формуляры) статистической отчетности называются *формами статистической отчетности.*

Статистическая отчетность имеет обязательный характер, все предприятия должны представлять ее в указанные сроки. Опоздание на одни сутки считается нарушением сроков представления государственной статистической отчетности, а опоздание на двое и более суток рассматривается как ее непредставление. Статистическое наблюдение в форме отчетности использует только один источник данных – документы (в первую очередь – документы бухгалтерского учета), и сама статистическая отчетность является документом, скрепленным подписями лиц, ответственных за представление и точность информации, и имеет юридическую силу. Формы статистической отчетности разрабатывает и утверждает Государственный комитет по статистике, включающие финансовые показатели, утверждаются, кроме того, Минфином РУз.

Статистическая отчетность чаще всего базируется на данных бухгалтерского учета. Она служит важным источником статистической информации для отдельных потребителей и экономики в целом, а также является основным видом статистического наблюдения, который используется для регулирования рыика.

Первичный учет ➢ Это регистрация различных фактов (событий, процессов и т.д.), производимая по мере их свершения и, как правило, в первичном учетном документе. В функции первичного учета входят операции наблюдения, т.е. регистрация данных и подсчет итогов. Примером может служить свидетельство о рождении ребенка. Другой пример – документы первичного учета в торговле (накладная, иные документы).

Каждое предприятие или учреждение представляет в органы статистики установленные формы статистической отчетности, характеризующие различные стороны их деятельности. Все формы статистической, отчетности утверждают органы государственной статистики. Руководство статистической отчетностью и ее организация также осуществляются органами государственной статистики.

Реквизиты (признаки) статистической отчетности:

- номер (индекс) формы;
- название формы;
- отчетный период;
- срок представления отчета;
- наименование и адрес предприятия;
- подписи ответственных лиц и печать.

Классификация форм отчетности:

✓ *по содержанию*:

• *общая* – это отчетность, содержащая одни и те же данные для определенной отрасли экономики и для предприятий (учреждений) всего отраслей экономики;

• *специализированная* – специфические показатели отдельных отраслей промышленности, сельского хозяйства и т.п.;

✓ *периоду времени, за который предоставляется отчетность*: *годовая* – сведения представляются за год;

• *текущая* – отчетность за другие периоды в течение года: квартальная, месячная и др.;

✓ *способу представления:*

• по телеграфу (срочная);

• по электронной почте (срочная);

✓ *масштабу:*

• *внешняя* – устанавливается государственными органами, министерствами и ведомствами;

• *внутренняя* – формируется в соответствии с учетной политикой предприятия.

Специально организованные статистические наблюдения обычно охватывают те явления и стороны общественной жизни, в которых в данный момент существует потребность, – переписи, единовременный учет и тематические статистические обследования.

Переписью называется специально организованное статистическое наблюдение, характеризующее определенное массовое явление или процесс на определенный момент или период времени. Например, проводятся переписи населения, различного рода учреждений и т.п.

Проведению переписей предшествует большая подготовительная работа, которая заключается в:

• составлении списков для охвата всех без исключения единиц совокупности, подлежащих наблюдению;

• разбивке административных районов на переписные участки и распределении объектов наблюдения между регистраторами;

• подготовке кадров;

• печатании и рассылке на места проведения переписи статистического инструментария;

• проведении пробных переписей.

Характерные особенности переписи населения:

1. Одновременность проведения ее на всей обследуемой территории.

2. Регистрация всех единиц наблюдения по состоянию на один и тот же критический момент переписи.

3. Программа наблюдения излагается в переписном листе, в котором каждый переписываемый указывается поименно.

4. Запись сведений при переписи населения всегда проводится путем опроса населения без предъявления документов, подтверждающих правильность ответа, по месту нахождения и по желанию переписываемого.

5. Организационно-подготовительные работы включают:

а) составление предварительного списка единиц наблюдения;

б) подбор и инструктаж кадров, выдача им документов;

в) подготовка инструментария (различного рода бланков, инструкций, портфелей и др.);

г) пропаганда проводимых статистических работ через средства массовой информации.

Особое значение переписи населения приобрели после Второй мировой войны в связи с распадом колониальной системы и развитием освободившихся стран. Сегодня в мире нет государства, в котором ни разу не проводилась бы перепись населения.

К подготовительной работе относится также пропаганда значения и порядка проведения переписи.

Единовременный учет ставит своей задачей определение численности и размещения изучаемого объекта или его составных частей по определенной территории на определенный момент или период времени (например, учет численности студентов по факультетам, учет установленного оборудования). Специальные статистические обследования, как правило, носят выборочный характер. Сейчас статистические наблюдения являются основной формой сбора статистической информации в Республике Узбекистан.

Единовременное наблюдение проводится один раз для решения какой-либо задачи или повторяется через неопределенные промежутки времени по мере надобности, например, перепись жилого фонда, школьная перепись и т.д.

Регистровая форма наблюдения

Это форма непрерывного статистического наблюдения за долговременными процессами. Примером статистического регистра является Единый государственный регистр предприятий и организаций всех форм собственности (*ЕГРПО*) – данные обо всех предприятиях, организациях и учреждениях и объединениях независимо от формы собственности (также индивидуальных предпринимателей и предприятий с иностранными инвестициями). *ЕГРПО содержит информацию по каждому предприятию*:

✓ регистровый код;

✓ отраслевая принадлежность и форма собственности;

✓ справочные сведения об учредителях;

✓ экономические показатели.

В практике статистики различают также:

1. Регистр предприятий – поименованный перечень предприятий.

Информационный фонд регистра предприятий содержит:

а) название и регистровый код (ИНН) хозяйствующего субъекта;

б) сведения о территориальной и отраслевой принадлежности;

в) справочные сведения (фамилии руководителей, номера телефонов и т.п.);

г) экономические показатели (уставный фонд, среднесписочная численность, объем выпускаемой продукции, финансовые результаты и т.п.).

2. Регистр населения – поименованный перечень жителей страны.

Информационный фонд регистра населения содержит:

а) Ф.И.О.;

б) пол;

в) дата и место рождения;

г) образование;

д) дата вступления в брак;

е) количество детей и т.д.

По учитываемым признакам регистры населения очень схожи с переписями, но регистры дают представление о юридическом населении, а переписи – о наличном и постоянном. Поэтому полностью отказаться от переписи населения, опираясь только на данные регистров, демографам не представляется возможным. Однако имея развитую электронную регистрацию населения, государства могут проводить переписи по сокращенной программе.

2.3. Программно-методологические вопросы наблюдения

Процесс проведения статистического наблюдения включает три этапа (рис. 2.2).

Этапы статистического наблюдения

1	2	3
Подготовка к наблюдению	Непосредственно сбор данных	Контроль, проверка собранных данных

Рис. 2.2. **Этапы статистического наблюдения**

1 этап – подготовительный.

Здесь решаются программно-методологические и организационные вопросы проведения наблюдения.

Программно-методологические вопросы наблюдения

1. Составление программы наблюдения
2. Определение объекта наблюдения, включая ограничение объекта по времени и территории
3. Определение единицы наблюдения
4. Определение формуляра наблюдения
5. Разработка инструкции наблюдения

Программа наблюдения Важнейшей частью статистического обследования является разработка программы наблюдения. Она представляет собой перечень вопросов, на которые необходимо получить ответ в процессе наблюдения.

Вопросы программы статистического наблюдения составляются с учетом требований, сформулированных еще в прошлом веке известным бельгийским статистиком Адольфом Кетле (1796–1874), который считается одним из основоположников современной научной постановки статистического наблюдения. Первое правило Кетле гласит, что программы статистического наблюдения должны включать только те вопросы, которые необходимы для решения поставленной цели. Второе правило Кетле – в программу наблюдения не следует включать вопросы, на которые невозможно получить ответ удовлетворительного качества. Третье правило Кетле гласит, что в программу наблюдения не должны включаться вопросы, которые могут быть расценены опрашиваемыми как вмешательство в сферу их личных интересов.

Следовательно, программа статистического наблюдения должна отвечать следующим параметрам:

- ✓ рассматривать только существенные признаки;

- включать только точные, легкие для понимания, а также сформулированные в логической последовательности вопросы.

С разработкой программы наблюдения связано определение:

- цели;
- объекта;
- единицы наблюдения;
- статистического формуляра;
- инструкции.

Целью наблюдения является сбор информации о социально- экономических процессах. Цель наблюдения формулируется в документе, на основании которого проводится наблюдение, – распоряжении Госкомстата Ре***спублики Узбекистан.***

Объект статистического наблюдения

Это совокупность единиц изучаемого явления, подлежащая статистическому изучению (предприятия, фермы, определенные группы населения и т.д.). Установить объект наблюдения – это значит точно определить состав и границы совокупности. Например, объектом переписи населения является совокупность всех живущих в данной стране лиц, объектом наблюдения при изучении промышленности – совокупность промышленных предприятий, при изучении сельского хозяйства – совокупность сельскохозяйственных предприятий и т.д.

Виды объектов наблюдения

Различают следующие виды объектов наблюдения:

- физические лица (население, работники предприятий, студенты);
- физические единицы (машины, оборудование, общественный транспорт);
- юридические лица (фирмы, акционерные общества, фермерские хозяйства);
- иные объекты.

Выделение объекта наблюдения представляет собой, как правило, сложную и ответственную задачу. Массовые явления и процессы общественной жизни обладают многими свойствами, они тесно связаны между собой, взаимно переплетаются. Поэтому недостаточно одного указания на объект исследования, нужно вместе с тем дать научное его определение, которое могло бы служить основанием для отграничения данного объекта от смежных с ним объектов, представляющих предмет самостоятельного исследования.

Определение объекта статистического наблюдения должно содержать точные указания главных его черт и свойств. Недостаточно, например, указать, что наблюдению подлежит совокупность промышленных предприятий. Признак *«промышленные предприятия»* сам по себе не может служить основанием для отграничения промышленных предприятий от сельскохозяйственных, транспортных и др. Выделение промышленных предприятий в качестве объектов статистического наблюдения требует точного установления системы признаков промышленного предприятия.

Точное и правильное отграничение объекта статистического исследования необходимо для получения сопоставимых данных, т.е. таких, которые в первую очередь можно было бы сравнивать с данными за прошлые периоды, а также во избежание при наблюдении случаев пропуска какой-либо категории его элементов или случаев двойного учета.

В каждом конкретном исследовании важное значение имеет точная формулировка *задач наблюдения.* Задачи статистического наблюдения могут быть правильно сформулированы только на базе глубокого и всестороннего знания объекта исследования и связанных с ним практических и научных нужд.

Единицу наблюдения следует отличать от единицы совокупности (отчетной единицы):

✓единица совокупности – это то, что подвергается обследованию;

✓единица наблюдения – это источник получаемых сведений.

Например, при проведении переписи оборудования в промышленности единицей наблюдения является промышленное предприятие, а единицей совокупности – промышленное оборудование. Иногда, единица наблюдения и отчетная единица совпадают. Например, при определении освоенных за год капитальных вложений предприятие-застройщик является одновременно и единицей совокупности, и отчетной единицей.

Программно-методологические вопросы наблюдения включают формуляры и инструкции по их заполнению.

Статистический формуляр – документ, в виде которого оформляется программа наблюдения для облегчения единообразия получаемых сведений. Может быть в виде:

✓карточки;

✓переписного листа;

✓опросного бланка;

✓анкеты и т.п.

Эго документ единого для всех отчетного образца, который содержит программу и результаты наблюдений и *состоит из обязательных элементов:*

✓титульная часть – информация о наименовании статистического наблюдения и органа, проводящего наблюдение, а так же о том, кто и когда утвердил этот формуляр;

✓адресная часть – информация отчетной единицы.

Иногда в формуляре после вопроса сразу же даются некоторые варианты ответа на него. Это называется *статистическим подсказом.*

По количеству единиц наблюдения статистический формуляр бывает 2 видов:

✓ индивидуальный – вопросы только об одной единице наблюдения;

✓ списочный – вопросы о нескольких единицах наблюдения.

Индивидуальный формуляр содержит сведения об одной единице совокупности (например, формы статистических отчетов о товарообороте № 1-торг и 3-торг заполняются каждой торговой организацией в отдельности).

Списочный формуляр

В списочном формуляре содержатся данные по нескольким единицам совокупности. Например, при переписи населения члены каждой семьи записываются в один переписной лист. Списочная форма носителя информации более удобна для машинной обработки, при которой с меньшими затратами производятся такие трудоемкие операции, как шифровка, перфорация и др.

Индивидуальные формуляры легче обрабатывать вручную. К статистическим формулярам составляется инструкция.

Как бы тщательно ни была составлена программа и разработан формуляр, для обеспечения единообразия его заполнения, толкования вопросов программы наблюдения все же необходима инструкция.

Инструкция

Это документ определяющий проведение наблюдения (его порядок) и заполнение формуляров. Это совокупность разъяснений и указаний по программе наблюдения. Другими словами этот документ содержит объяснения вопросов программы с конкретными примерами, указания по взаимосвязи вопросов. Инструкция издается либо в виде отдельной брошюры, либо дается в подсказках, либо на самом формуляре наблюдения (обычно на оборотной стороне).

2.4. Организационные вопросы статистического наблюдения

Статистическое наблюдение проводится по плану и программе, составленным заранее.

Организационный план статистического наблюдения это *документ, в котором содержится перечень подготовитель-*

ных работ и проведения статистического наблюдения с указанием конкретных сроков их проведения.

В организационную часть плана статистического наблюдения включаются следующие вопросы:

Организационные вопросы наблюдения

- органы наблюдения
- место проведения наблюдения
- время и сроки наблюдения
- подбор и подготовка кадров
- порядок приема и сдачи материалов наблюдения

Здесь прежде всего указываются органы, выполняющие статистическое наблюдение. Это могут быть органы государственной статистики РУз или конкретные службы. Четко определяются и разграничиваются права и обязанности конкретных учреждений и организаций, участвующих в наблюдении.

Помимо постоянных органов, осуществляющих статистическое наблюдение, иногда создаются временные органы для проведения, как правило, крупных обследований.

Решение вопроса о месте наблюдения особенно важно при изучении перемещающихся единиц. Так, за продукцией можно наблюдать на предприятиях, где она производится, или в торговле, куда она поступает для продажи. Или работника можно наблюдать как на работе, так и дома.

При проведении статистического наблюдения большое значение имеет решение вопроса о времени наблюдения.

Время наблюдения – это то время, к которому относятся собираемые сведения. В годовой отчетности это – год, в месячной - месяц и т.п., причем должны быть точно отграничены начало и конец каждого периода. Так, например, в годовую или месячную отчетность включаются сведения о продукции, произведенной не позднее 12 час. ночи последнего дня отчетного периода (и соответственно не ранее 0 час. первого дня отчетного периода).

При проведении переписей большое значение имеет правильное установление даты и сроков проведения наблюдения. Размеры некоторых явлений и характеризующих их признаков подвержены в течение года значительным колебаниям. Так, численность населения, перемещающегося из деревни в город и обратно, а также из одной местности в другую, далеко не одинакова в различные времена года. Осенью и зимой она обычно бывает значительно меньше, чем весной и летом. Вполне естественно поэтому, что картина территориального распределения населения может значительно изменяться в зависимости от того, в какое время года осуществляется перепись.

Тот момент времени или дата, на которую фиксируется состояние объекта, называется *критическим моментом наблюдения.*

Критический момент *Это конкретный день и час, по состоянию на который должна быть проведена регистрация признаков по каждой единице наблюдения.* Например, перепись населения проводится по состоянию на 12 часов ночи (0 ч 00 мин) определенного числа (например, если переписчик пришел к респонденту в 16.00, а в 14.30 в семье родился ребенок, то он не учитывается в данной переписи населения, так как данные собираются на фиксированный момент – начало данных суток (00 ч 00 мин).

Например, при переписи населения 1989 г. таким критическим моментом было 12 часов ночи с 12 на 13 января. В соответствии с этим все лица, умершие до 12 часов ночи, регистрации подлежат, а умершие после 12 часов ночи – не подлежат, родившиеся до 12 часов регистрации подлежат, а родившиеся после 12 часов – не подлежат.

Правильный выбор критического момента имеет большое значение. Так, например, выбор критического момента переписи населения в январе имеет тот смысл, что в это время население находится в состоянии наибольшего покоя, менее, нежели в другое время года, передвигается по стране.

Время проведения наблюдения Это время, в течение которого оно осуществляется, может отличаться от времени наблюдения. Так, например, при месячной отчетности время наблюдения отчетный месяц, а время проведения наблюдения – месяц и еще несколько дней (в отчетности, например, промышленных предприятий о продукции – месяц и три дня, поскольку ежемесячная отчетность о продукции должна быть выслана не позднее третьего числа следующего за отчетным месяца).

Решение вопроса о том, выбрать ли в качестве времени наблюдения период времени или дату (момент), определяется природой изучаемого явления.

Многие статистические показатели по своему характеру могут быть выражены только за определенные отрезки времени, например, выработка продукции, фонд заработной платы и др. Для таких показателей устанавливаются периоды времени за которые должны быть получены сведения при проведении статистического наблюдения. При статистическом наблюдении в форме отчетности в большинстве случаев показатели составляются за определенный период, например, декаду, месяц, квартал и т.д. Наряду с этим в отчетности содержатся показатели, относящиеся к определенным датам, например, численность персонала показывается обычно на конец месяца.

Срок наблюдения Это время, отведенное на массовый сбор данных. Он определяется из следующих параметров:

• объем исследования (число исследуемых признаков и единиц статистической совокупности);

• численность персонала, занятого проведением исследования.

Важнейшее место в системе подготовительных работ имеют **подбор и подготовка кадров**, а также инструктаж аппарата учетно-экономических служб, привлеченных для сбора необходимой информации.

Проведение массовых работ требует участия множества исполнителей (в переписях населения участвуют тысячи счетчиков). Все они должны пройти специальное обучение – ***инструктаж*** и провести пробное заполнение тех формуляров, которые предполагается использовать в статистическом наблюдении.

Для них организуется инструктаж по разъяснению вопросов анкеты (или другого формуляра наблюдения) и пользованию инструкцией. Объясняется, например, что при наличии подсказов счетчик обязан ознакомить респондента со всеми вариантами ответов, не выделяя те из них, которые он сам считает наиболее вероятными. Затем проводится пробное заполнение анкет, итоги которых коллективно обсуждаются.

В целях успешного осуществления статистического наблюдения немаловажное значение имеют подготовка статистического инструментария (различного рода бланков, инструкций и т.п.), его размножение и своевременное снабжение им персонала, проводящего наблюдение.

Должна быть составлена смета на проведение специального обследования, в которой предусматриваются размножение материалов наблюдения (бланков, инструкций), оплата услуг средств связи, транспорта, работы инструкторов, счетчиков и др.

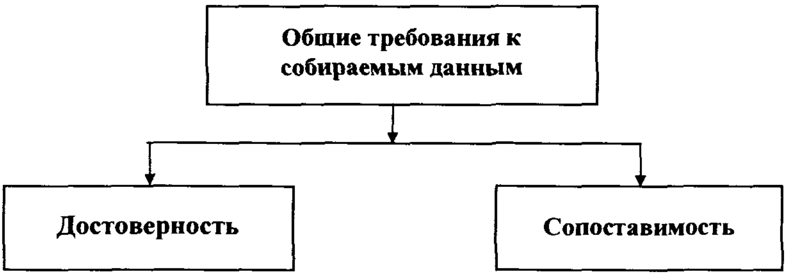
Наконец, к числу важнейших подготовительных мероприятий относится пропаганда проводимых статистических работ средствами печати, радио, телевидения (разъяснение задач и целей обследования).

Все это способствует более успешному их проведению. Следует отметить, что статистическое обследование – дорогостоящая и трудоемкая процедура.

Проведение обследований должно быть обосновано и подкреплено финансовыми, материальными и трудовыми ресурсами.

2.5. Требования, предъявляемые к собираемым данным

Результаты исследования будут ценны лишь в том случае, если они базируются на фактическом материале. Даже теоретический анализ, основанный на закономерностях развития явлений и позволяющий углубить наше понимание существа процессов, как правило, базируется на выводах, вытекающих из конкретных фактов, т.е. связан с необходимостью сбора исходных данных. Всякие ли данные, факты, собранные в процессе статистического наблюдения, могут быть использованы для дальнейшего исследования? Нет, не всякие. Они должны отвечать определенным требованиям.



Достоверность

Это соответствие данных тому, что есть на самом деле. Вся методика, организация и техника проведения статистического наблюдения должны быть нацелены на обеспечение достоверных данных. Чтобы понять характер задач возникающих при этом, представим статистическое наблюдение в виде взаимодействующих компонентов (рис 2.3).

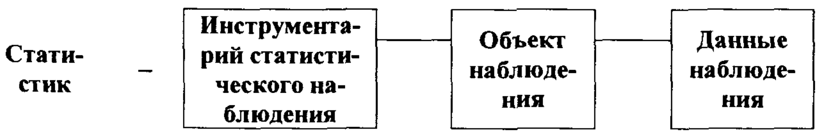


Рис. 2.3. **Составляющие статистического наблюдения**

Очевидно, что достоверность данных зависит как от характеристик самого статистика – его профессиональной подготовки, коммуникабельности, организационных навыков и т.д., так и от качества используемого инструментария – программы наблюдения, бланков, анкет, инструкций по их заполнению. Они, в конечном счете, тоже зависят от статистика (непрофессионализм или заведомая ложь).

Заведомая ложь (так называемая подтасовка фактов) может обнаружиться в таких социально острых вопросах, как детская смертность, преступность, наркомания, загрязнение окружающей среды, гибель солдат во время службы в армии, проституция, бедственное положение беспризорных детей, отсутствие надзора за беспомощными стариками и т.п.

На достоверность данных влияет и подготовленность объекта к статистическому обследованию. Это может быть сделано в форме предварительного извещения населения о предстоящем обследовании – в газетах, по радио, телевидению. Влияет на достоверность и упорядочение названия улиц и нумерации домов, квартир и т.д.

Общими условиями обеспечения достоверности полного охвата наблюдаемого объекта являются; полнота и точность регистрации данных по каждой единице наблюдения. Другими словами: непосредственный сбор массовых данных должен обеспечить полноту фактов, относящихся к рассматриваемому вопросу, так как явления находятся в постоянном изменении, развитии. В том случае, если отсутствуют полные данные, анализ и выводы могут быть ошибочными. Полнота обеспечивается, **во-первых**, охватом единиц исследуемой совокупности.

Например, менеджер должен сделать вывод о развитии туристического бизнеса. Очевидно, ему следует собрать информацию обо всех туристических фирмах, действующих в данном регионе.

Во-вторых, полноту следует понимать и как охват наиболее существенных сторон явления, так как каждое изучаемое явление или совокупность носит достаточно сложный характер и имеет самые различные признаки. Если в процессе наблюдения за туристическими фирмами, например, не будут зарегистрированы финансовые результаты, то нельзя сделать окончательные выводы о развитии туризма. **В-третьих**, при изучении явления во времени полнота предполагает получение данных за максимально длительные периоды.

Достаточно полные статистические данные являются, как правило, массовыми, исчерпывающими. Они обеспечивают потребности комплексного статистического исследования.

На практике исследуемые социально-экономические явления чрезвычайно широки и многообразны, поэтому охватить все явления невозможно. Исследователь вынужден проводить сбор данных лишь по части совокупности. Выводы же делаются по всей совокупности. В таких ситуациях важнейшим требованием, предъявляемым к статистическому наблюдению, является ***обоснованный отбор*** той части совокупности, по которой собираются данные. Эта часть должна отражать основные свойства и специфические особенности явления и быть типичной (эти вопросы подробно рассмотрены в главе «Выборочное наблюдение»).

Можно отметить еще одно важное требование – это ***сопоставимость данных***, или ***единообразие***.

Сопоставимость

Чтобы данные об отдельных явлениях можно было обобщать, они должны **быть сопоставимы друг с другом**: собираться в одно и то же время, по единой методике. Кроме того, должна быть обеспечена сравнимость с прошлыми исследованиями, чтобы можно было понять, как изменяется явление. Для этого должна быть полная ясность организации и методологии статистического наблюдения, чтобы были понятны характер и причины различий в данных наблюдений, если таковые были вызваны именно организационно-методологическими факторами.

Сравнимость данных разных наблюдений выполняется, если использовались одно и то же определение единицы наблюдения, одна и та же методика регистрации первичных признаков и методика расчета, вторичных признаков, таких, как себестоимость, производительность труда, рентабельность, ликвидность и т.д.

Важным условием сравнимости является сохранение времени проведения наблюдения и периода или момента, к которому относятся регистрируемые данные. Например, численность студентов университета определяется на начало учебного года, стипендиальный фонд – на полгода (или год) и т.д. Обычно рекомендуется, чтобы данные соответствовали хотя бы одному полному циклу изучаемого процесса, например, учебному, хозяйственному или финансовому году и т.д. Если сильно влияет сезонность, данные должны собираться по месяцам или по кварталам. Время наблюдения выбирается таким образом, чтобы наблюдаемый объект находился в наиболее стабильном состоянии.

В условиях рынка возрастает значение еще одного требования к собираемым в результате наблюдения данным – **своевременности**. Практический менеджмент нуждается в постоянно пополняемых статистических данных. Достоверная, полная, но запоздалая информация оказывается практически ненужной.

2.6. Виды статистического наблюдения и их особенности

Отметим определенное многообразие видов и организационных форм или способов проведения статистического наблюдения. Это позволяет исследователю выбрать наблюдение, которое соответствует поставленным целям и задачам, учитывает особенности изучаемого объекта или явления, соотносится с реальными условиями места и времени, обеспечивается имеющимися ресурсами. Выбор и обоснование характера статистического наблюдения является важнейшим вопросом исследования.

Статистическое наблюдение подразделяется:

• **по охвату единиц** совокупности на сплошное и не сплошное;

• **по времени проведения** на непрерывное (текущее), единовременное и периодическое;

• **по способу организации** на специально организованное статистическое наблюдение и отчетность;

• **по источникам сведений** на непосредственное наблюдение, документальное наблюдение и опрос (рис. 2.4).

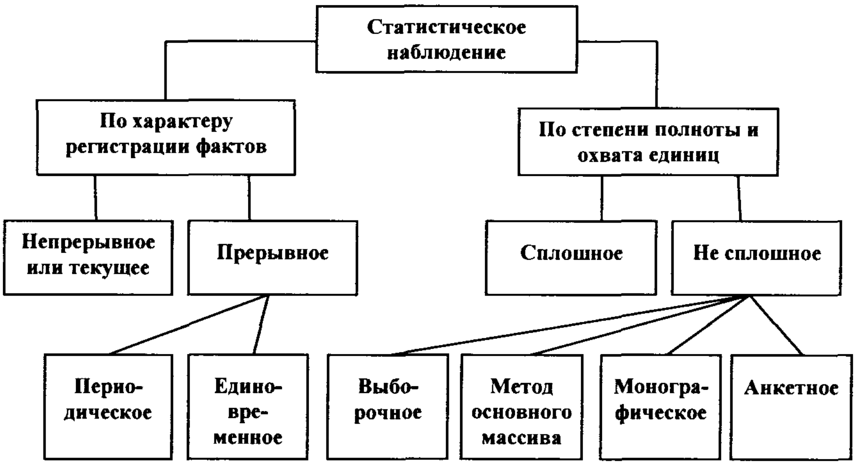


Рис. 2.4. **Основные виды статистического наблюдения**

Такое подразделение не является абсолютным, так как происходит развитие статистического наблюдения, совершенствуется методология сбора данных, получают распространение новые современные методы сбора статистических данных и информации в целом. Рассмотрим это подробнее.

Непрерывное статистическое наблюдение

Это так называемое еще текущим, проводится, когда необходимо зарегистрировать все единицы, случаи и т.п. по мере их возникновения. Это систематическая регистрация фактов. Так, непрерывно регистрируются все дорожно-транспортные происшествия, противоправные акты. Территориальными органами производится постоянная регистрация актов гражданского состояния – рождений, браков, смертей. Страховые компании регистрируют по мере возникновения все несчастные случаи и другие неблагоприятные случайные события. Получаемые в результате текущего наблюдения статистические данные богаче, так как дают непрерывную картину явления, а не моментальный статистический снимок.

При текущем наблюдении нельзя допускать значительного разрыва между моментом возникновения факта и моментом его регистрации.

Прерывное (периодическое) нблюдение

Это такое наблюдение, которое повторяется через определенные промежутки времени. Примером периодического наблюдения являются ежегодные переписи скота, проводимые по состоянию на 1 января, регистрация цен ярмарочной торговли на сельскохозяйственные продукты, осуществляемая 25-го числа каждого месяца.

По времени проведения специально организованное наблюдение также может быть разовым, или единовременным, и периодическим. ***Единовременное наблюдение*** может проводиться по мере возникновения потребности (в таком случае оно является прерывным). Потребность в разовом наблюдении возникает из конкретных условий. Если иным путем необходимую информацию получить невозможно, то принимается решение о проведении единовременного наблюдения. Это может быть «школьная перепись», учет состояния жилых зданий или неходовых товаров, сырья. Для изучения рынка труда в РУз актуальным является изучение статуса занятости населения на основе специальных обследований домашних хозяйств. Это позволило бы скорректировать официальную статистику безработных, которая не дает объективную информацию.

Специально организованное статистическое наблюдение может быть также периодическим. Таким наблюдением является перепись населения. По рекомендации международных статистических органов переписи населения в странах должны проводиться не реже, чем через 10 лет, в год, оканчивающийся на ноль или смежный, т.е. оканчивающийся на единицу или девятку.

На практике постоянно возникают задачи, для решения которых имеющаяся информационная база недостаточна или практически отсутствует. Это особенно ощущается сейчас, поскольку, с одной стороны, общий объем статистических данных резко сократился. Многие явления и процессы стали практически неуправляемыми. С другой стороны, центр тяжести в управлении переносится на региональный уровень. Это увеличивает потребность в дополнительной информации в регионах, возникает необходимость более глубокого исследования отдельных и новых социально-экономических явлений и процессов. Для этих целей используется **специально организованное статистическое наблюдение.** Вот некоторые примеры, свидетельствующие об актуальности и необходимости таких наблюдений.

Так например, часть предприятий страны имеет значительную задолженность по выдаче средств на заработную плату. Это отражается на уровне жизни населения, подчас значительно дестабилизирует общество. Возникает вопрос об изучении влияния этого фактора на изменение реальной заработной платы, качество жизни. Поэтому необходимо проводить специальное статистическое наблюдение отдельно по предприятиям, регулярно выплачивающим заработную плату, и предприятиям, имеющим задолженность по выплате заработной платы. Полученная информация позволила бы принять важные решения, разработать правительственные меры по снижению социальной напряженности в обществе.

Сплошное наблюдение

При ***сплошном наблюдении*** регистрации подлежат все без исключения единицы совокупности. Оно применяется, например, при переписи населения, сборе данных в форме отчетности, охватывающей предприятия разных форм собственности, учреждения и организации и т.д.

Развитие многоукладной экономики увеличило число объектов экономической деятельности. Это способствовало расширению практики *не сплошного наблюдения,* которое, в свою очередь, подразделяется на способ *основного массива, выборочное* и *монографическое.*

Выборочные наблюдение

Это наблюдение, при котором характеристика всей совокупности фактов дается по некоторой их части, отобранной в случайном порядке. При правильной организации оно дает достаточно достоверные данные, вполне пригодные для характеристики всей изучаемой совокупности. Для того чтобы понять, хороший напиток или нет, не обязательно выпить целую бутылку, то же можно сказать в отношении проверки качества любой продукции. В решении такого рода задач, да и во многих других случаях, может помочь только выборка. Выборочный метод играет все большую роль в отечественной статистике, поэтому планирование выборки, методы отбора, оценки ее репрезентативности специально рассматриваются в гл. 7.

Способ основного массива

При способе *основного массива* обследованию подвергается основной массив – та часть единиц, которая вносит наибольший вклад в изучаемое явление. Часть совокупности, о которой заведомо известно, что она не играет большой роли в характеристике совокупности, исключается из наблюдения, т.е. при этом методе отбираются и обследуются наиболее крупные единицы. Логика метода состоит в том, что крупные единицы могут практически определять интересующие нас статистические показатели. Например, вследствие концентрации производства в отрасли несколько наиболее крупных предприятий могут давать основной объем продукции, в то время как большая масса мелких предприятий выпускает ее незначительную часть. Это бывает при высоком уровне монополизма в отрасли экономики, особенно в условиях региона.

Монографическое наблюдение

Монографическое наблюдение или монографическое обследование – это подробное описание отдельных единиц наблюдения в статистической совокупности. Перед монографическим наблюдением не ставится цель дать характеристику всей совокупности.

Оно соответствует решению задач по более глубокому исследованию отдельных единиц совокупности. Поэтому монографическое наблюдение обычно проводится в отношении типичных единиц или характерных типов явлений.

Это может быть описание бюджета семьи шахтера или безработного, молодого фермера или обанкротившегося предприятия, выставляемого на продажу. Программа монографического наблюдения предусматривает определенную свободу действий исследователя. Это означает, что в процессе наблюдения не только даются ответы на поставленные вопросы, но и фиксируются признаки, стороны деятельности, которые могут представлять интерес для дальнейшего изучения или составления программы наблюдения уже для всей совокупности.

Анкетное обследование

В ***анкетном обследовании*** сбор данных основан на принципе добровольного заполнения адресатами анкет (листов опроса).

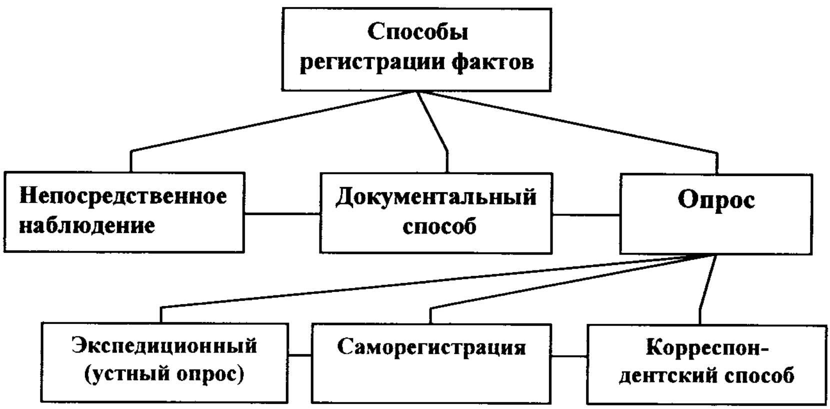


Рис. 2.5. **Виды наблюдения по источникам сведений**

Как правило, заполненных анкет возвращается меньше, чем рассылается. Кроме того, проверить достоверность собранного материала очень сложно. Поэтому такой способ наблюдения может применяться в тех случаях, когда не требуется высокая точность сведений, а нужны приблизительные характеристики.

К нему прибегают при проведении социологических обследований, в статистике **связи,** в библиотеках для опроса читателей, в торговле для изучения спроса населения на отдельные товары и т.д.

По способу регистрации фактов различают следующие виды статистического наблюдения (рис. 2.5).

Непосредственное наблюдение

Непосредственным является такое, наблюдение, при котором сами регистраторы путем замера, взвешивания или подсчета устанавливают факт, подлежащий регистрации, и на этом основании производят записи в формуляре наблюдения. Так, при учете остатков товаров в торговле за основу берется их инвентаризация. При переписи оборудования сведения заносятся в формуляр на основе личного осмотра машин и т.д.

Документальный учет фактов

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

Так, при переписи оборудования необходимые сведения могут быть получены на основании технических паспортов. В торговле источником таких сведений является паспорт торгового предприятия, содержащий достаточно полную и достоверную информацию о самых разнообразных сторонах его коммерческой деятельности.

Опрос

Это наблюдение, при котором ответы на изучаемые вопросы записываются со слов опрашиваемого. К опросу, например, прибегают при проведении переписи населения.

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

Экспедиционный способ

При этом специально подготовленные работники, которых обычно называют счетчиками, или регистраторами, сами устанавливают учитываемые факты путем непосредственного наблюдения на основании документов или опроса соответствующих лиц и сами заполняют формуляр наблюдения. Этот способ обеспечивает получение более доброкачественных материалов. Важнейшие статистические обследования населения проводятся экспедиционным способом.

При ***способе саморегистрации (само исчислении***) соответствующие документы заполняют сами опрашиваемые. Обязанность счетчиков (регистраторов) здесь состоит в раздаче бланков наблюдения опрашиваемым, инструктаже их и затем в сборе заполненных формуляров, которые при этом проверяются.

Корреспондентский способ

Суть данного способа заключается в том, что сведения в органы, ведущие наблюдение, сообщают их корреспонденты. Этот способ не требует больших затрат, но он не обеспечивает высокого качества материалов, так как проверить точность сообщаемых сведений непосредственно на месте не всегда представляется возможным. В современных условиях получает распространение специальное организованное систематическое наблюдение за состоянием явлений и процессов, объектов совокупности – мониторинг. Мониторинг используется для характеристики и слежения за социальными индикаторами, позволяющими исследовать, например, качество жизни. Получает распространение мониторинг окружающей среды. Данные мониторинга обобщаются. Он позволяет получать оперативную информацию для принятия решений. На практике мониторинг обычно выходит за рамки традиционного статистического наблюдения. Тем не менее, в каждом конкретном случае он может являться важным источником статистических данных, информации.

Различные виды статистических наблюдений в действительности могут сочетаться, взаимно дополняя друг друга и формируя эффективную статистическую информацию. Так, на начальном этапе исследования рынка труда была введена система регистрации безработных службами занятости. Это был первый и единственный источник данных о рынке труда. В настоящее время исследование рынка труда опирается на два основных источника. Это специальное обследование рабочей силы и ведомственная статистика служб занятости.

Теория и практика сбора данных для комплексного экономико-статистического исследования социально-экономических явлений и процессов и нужд управления постоянно совершенствуется. В целом, в настоящее время остается актуальной задача создания интегрированной системы, обеспечивающей комплексный подход к формированию информационной базы статистики, для чего намечается широкая программа совершенствования статистического наблюдения.

2.7. Ошибки статистического наблюдения и контроль точности информации наблюдения

Всякое статистическое наблюдение ставит задачу получения таких данных, которые точнее бы отображали действительность. Точность и достоверность собираемой статистической информации – важнейшая задача статистического наблюдения. Под точностью статистической информации понимается уровень соответствия величины изучаемого показателя показателю, получаемому посредством статистического наблюдения, действительному его значению. Чем ближе величина показателей, полученных в результате статистического наблюдения, к фактическим их значениям, тем выше точность статистического наблюдения.

Отклонения или разности между исчисленными показателями и действительными (истинными) величинами исследуемых явлений находят отражение в показателях, называемых ошибками или погрешностями. Чтобы предупредить их возникновение или уменьшить их размеры, необходимо в процессе подготовки и проведения наблюдения предусмотреть и осуществить ряд мероприятий. **Во-первых,** необходимо обеспечить правильный подбор и обучение персонала, на который будут возложены проведение наблюдения, систематический контроль за ходом наблюдения, широкая разъяснительная работа. **Во-вторых**, следует предусмотреть соответствующие меры во избежание сознательного искажения фактов, приписок и т.д., что является не только нарушением государственной дисциплины, но и прямым преступлением, наносящим вред интересам дела.

В зависимости от характера и степени влияния на конечные результаты наблюдения, а также исходя из источников и причин возникновения неточностей, допускаемых в процессе статистического наблюдения, обычно выделяют ошибки регистрации и ошибки репрезентативности (представительности) (рис. 2.6).

Ошибка статистического наблюдения представляет собой расхождение между величиной какого-либо показателя, установленной посредством наблюдения, и действительными его размерами.

Ошибки регистрации Это отклонения между значением показателя, полученного в ходе статистического наблюдения, и фактическим, действительным его значением.

Они могут быть как при сплошном, так и при не сплошном наблюдении и возникают в наших условиях чаще всего непреднамеренно, помимо воли и желания лиц, производящих наблюдение. Причиной их может быть отсутствие твердых знаний и навыков для проведения статистического наблюдения, неточности округления, описки. В отдельных случаях могут иметь место и преднамеренные ошибки, если лица, производящие наблюдение, по каким-либо причинам заинтересованы в скрытии или искажении действительности. Непреднамеренные ошибки регистрации могут носить случайный, или систематический характер.

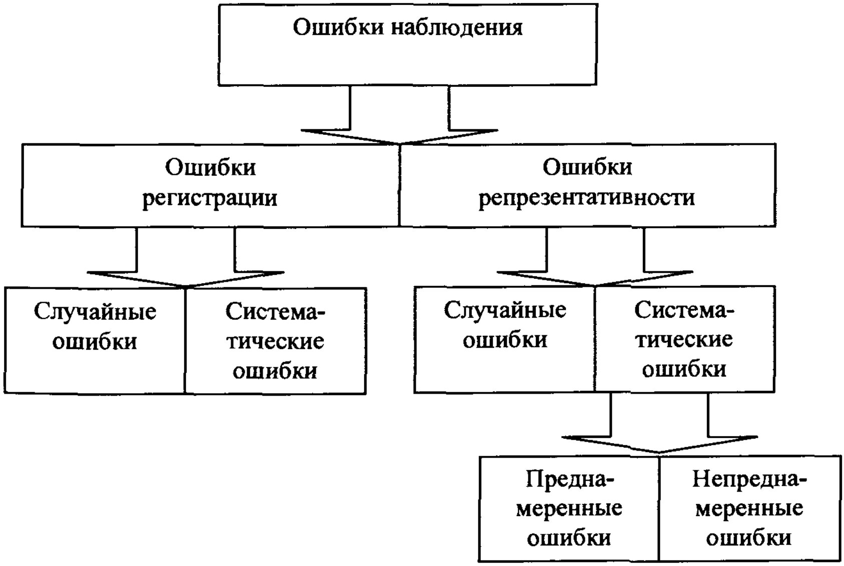


Рис. 2.6. **Ошибки статистического наблюдения**

Ошибки репрезентативности Они могут быть только при выборочном наблюдении и возникают в результате того, что состав отобранной для обследования части единиц совокупности недостаточно полно отображает состав всей исследуемой совокупности в целом, хотя регистрация фактов по каждой отобранной для обследования единицы была произведена точно. Подробнее вопрос об ошибках репрезентативности излагается в 7 главе учебника.

Случайные ошибки

Такого рода ошибки возникают при регистрации фактов по чисто случайным причинам: усталость учетчика, описка и т.п. При случайных ошибках результаты наблюдения могут быть искажены с одинаковой вероятностью, как в сторону преувеличения, так и в сторону преуменьшения фактов. При достаточно большом числе наблюдений случайные ошибки могут взаимно погаситься и не оказать существенного влияния на итоговые результаты наблюдения.

Систематические ошибки

Это непреднамеренная ошибка возникает по какой-либо определенной причине и искажает показатели в одну сторону: либо занижения, либо завышения. Они являются неслучайными и имеют определенную направленность. Такие ошибки очень опасны, так как приводят к искажению результатов статистического исследования. Эти ошибки, как правило, являются преднамеренными. Известно, например, что люди предпочитают преуменьшать свои доходы, округлять возраст, стараются показать большую осведомленность в области культуры, науки, чем это есть на самом деле. Предприятия также могут внести элементы недостоверности в свою информацию, особенно в те характеристики, от которых зависят величина налоговых платежей, расчеты с кредитора и т.п. Таким образом, **систематические ошибки репрезентативности** – это отклонения, возникающие вследствие нарушения принципов случайного отбора единиц изучаемой совокупности. Размеры систематической ошибки репрезентативности не поддаются количественной оценке.

2.8. Контроль материалов наблюдения

Материалы, собранные в результате статистического наблюдения, подвергаются всесторонней проверке и контролю.

Они должны быть проверены с точки зрения полноты охвата всех единиц совокупности наблюдения и правильности заполнения документов.

С целью выявления и устранения допущенных при регистрации ошибок производится проверка собранного материала посредством счетного и логического контроля.

Счетный контроль

Основан на жесткой связи между признаками, которая может быть проверена арифметическими действиями: сложением, вычитанием, умножением, делением. Связь такого рода часто отражается в заголовках граф отчетности и в подсказах такого рода: графа X равна графе У плюс графа 2 или графа X равна графе Y, деленной на графу Z, и т.д. Счетный контроль используется для проверки итоговых сумм. Если представленное число слагаемых не является полным, то сумма слагаемых должна быть меньше либо равна общему итогу, но не может превышать его.

Логический контроль

Состоит в сопоставлении ответов на вопросы и выяснении логической их совместимости. В процессе логического контроля могут быть установлены нереальные или мало правдоподобные ответы. Приемы контроля чрезвычайно разнообразны в зависимости от особенностей изучаемого явления, формы наблюдения, способа сбора сведений и целого ряда других факторов. Наиболее общими приемами логического контроля являются следующие:

1.Сопоставление ответов на различные взаимосвязанные вопросы одного и того же документа. Так, например, если в годовом отчете имеются сведения о посевных площадях картофеля, а расход картофеля на семена не показан, то где-то допущена ошибка. Или, например, в переписном листе значится, что четырехлетний ребенок учится в седьмом классе средней школы. Ясно, что два ответа на взаимосвязанные вопросы логически несовместимы. Подобные записи требуют уточнения сведений и исправления допущенных при регистрации ошибок.

2. Сравнение записей в проверяемом формуляре с аналогичными сведениями в других документах. Например, сведения о заработной плате производственного персонала имеются в отчете по труду и в отчете по себестоимости продукции.

3. Сопоставление отчетных показателей за смежные периоды. В частности, данные о закупках сельскохозяйственных продуктов в отчетах даются нарастающими итогами. Это обеспечивает определенную логическую связь показателей за отчетный период с показателями за предшествующий период: объем закупок зерна на 20 августа не может быть меньше, чем было закуплено на 10-е число этого же месяца и зафиксировано в представленном ранее отчете. Полезно сопоставление показателей за два смежных года и привлечение для анализа дополнительного материала, имеющего непосредственное отношение к проверяемым показателям. Так, при резком увеличении закупок зерна в отчетном году по сравнению с прошлым годом следует сравнить показатели о производстве зерна за эти же два года и убедиться в реальной возможности такого роста продажи зерна государству.

4. Сравнение данных наблюдения с действующими нормами, плановыми показателями или фактическими средними показателями за прошлые периоды. Предположим, что в отчете предприятия имеются данные о фонде заработной платы и числе работающих. Поделив первый показатель на второй, получили, что средняя заработная плата на одного работающего за месяц составила 3 млн сум. Полученная цифра нереальна, не соответствует тарифным ставкам. Следовательно, есть все основания полагать, что в отчете либо занижен фонд заработной платы, либо неверно показана численность работающих. Сведения обязательно нужно уточнить на предприятии. Аналогично при значительных отклонениях данных наблюдения от норм или плановых показателей сведения должны быть взяты под сомнение и подвергнуты всесторонней проверке.

5. В ряде случаев при контроле данных применяется метод балансовой увязки показателей. Наиболее часто в практике встречается такое балансовое равенство: наличие на начало периода, плюс поступление, минус выбытие в течение отчетного периода должно быть равно наличию на конец отчетного периода.

Указанные приемы проверки статистических данных путем арифметического и логического контроля применимы при любых формах наблюдения. Они широко используются при проверке, как отчетности, так и материалов специальных статистических наблюдений.

Обычно для проверки поступающего материала наблюдения составляется ***схема контроля,*** в которую включаются все увязки между вопросами программы наблюдения: как арифметические, так и логические.

Никогда не следует произвольно вносить исправления в формуляр. Необходимо либо самому статистику провести повторное наблюдение (повторный опрос и т.д.), либо обратиться к лицам, отвечающим за представленную информацию (директору, главному бухгалтеру предприятия).

Данные наблюдения считаются принятыми, если они прошли контроль и если потребовалось, в них внести исправления. Проверкой собранных данных завершается начальный этап статистического исследования.

Интеллектуальный тренинг

1. В чем сущность и особенности статистического наблюдения?

2. Что имел ввиду английский политик Б. Дизраэли (1804-1881) когда говорил: «В жизни, как правило, преуспевает тот, кто располагает лучшей информацией»?

3. Почему немецкий философ Гегель (1770-1831) говорил: «Факты до того плохие, что в отдельных случаях не подчиняются теории»?

4. «Статистика – это цифры. А цифры управляют миром». Кому принадлежат эти слова?

5. Кто сказал: «Есть ложь, есть наглая ложь, а есть и статистика». Что он имел ввиду?

6. Можно свободно говорить непонятными мыслями, однако факты не допускают к этому. Правда ли это?

7. Назовите основные программно-методологические вопросы статистического наблюдения.

8. Назовите основные организационные формы наблюдения.

9. Что такое единица и объект наблюдения?

10. Какие вы знаете статистические наблюдения:

а) по формам организации;

б) по времени регистрации фактов;

в) по способу регистрации;

г) по охвату изучаемого объекта наблюдения;

д) по источнику данных.

11. Что такое критический момент наблюдения и для чего он устанавливается?

12. Что такое статистическая отчетность, ее назначение?

13. Какие задачи решаются на основе специальных статистических наблюдений?

14. Какие ошибки могут возникать при статистическом наблюдении?

15. Какие вы знаете способы контроля материалов наблюдения?

Использованная и рекомендуемая специальная литература

1. Воронов Ю.П. Методы сбора информации в социологическом исследовании. М.: Статистика, 1974.

2. Деев Г., Крутова Т. Метод основного массива в статистических наблюдениях// Вестник статистики. 1992. № 5.

3. Деев Г., Мухин П. Не сплошное статистическое наблюдение: исторический опыт, практика, перспективы // Вопросы статистики. 1996. № 3. С. 21-27.

4. Джессен Р. Методы статистических обследований // Финансы и статистика. 1985.

5. Маргенштерн О. О точности экономико-статистических наблюдений. М.: Статистика, 1968.

6. Об ответственности за нарушение порядка представления государственной статистической отчетности // Вестник статистики. 1992. №1.

7.Статистический словарь. М.: Финстатинформ, 1996.

Глава III
СВОДКА И ГРУППИРОВКА

Дорожная карта

3.1. Понятие о статистической сводке

Сводка – второй этап статистического

Первичные данные, получаемые в результате статистического наблюдения, представляют собой обычно сырой материал, который подвергается научной обработке с целью получений сводной характеристики изучаемого явления. Так, в итоге проведения переписи населения поступают сведения о каждом жителе страны по ряду признаков, которые сами по себе не являются еще статистической характеристикой населения страны, а представляют лишь мате риал для нее. Чтобы получить на основе этих данных характеристику населения страны, надо свести их по ряду признаков, например, подсчитать численность населения отдельных административно-территориальных единиц (республик, краев, областей, округов и т.д.), численность мужчин и женщин, численность населения, проживающего в городах и сельской местности и т.п.

Получаемая в процессе статистического наблюдения информация об отдельных единицах статистической совокупности характеризует их, как правило, с различных сторон. Например, при изучении торговли района собранные статистические данные о коммерческой деятельности отдельных торговых предприятий содержат соответствующую оценку работы каждого из них.

Однако сообщающую характеристику по торговым предприятиям в целом можно получить, систематизируя и обобщая полученную информацию, а также сводку, являющуюся второй стадией статистического исследования, в процессе которого осуществляется научная обработка собранного материала. В результате этого этапа индивидуальные данные превращаются в упорядоченную систему статистических показателей, дающих возможность в целом оценить коммерческую деятельность торговых предприятий, выявить закономерности их развития.

Таким образом, ***статистическая сводка*** – систематизация единичных фактов, позволяющая перейти к обобщающим показателям, относящимся ко всей изучаемой совокупности и ее частям, и осуществлять анализ и прогнозирование изучаемых явлений и процессов.

Сводка и группировка материалов являются вторым этапом статистического исследования.

В результате сводки ***множество индивидуальных показателей,*** относящихся к каждой единице объекта наблюдения, ***превращаются в систему статистических таблиц и итогов,*** проявляются типические черты и закономерности изучаемого явления в целом.

Статистические сводки различаются по ряду признаков (рис. 3.1).

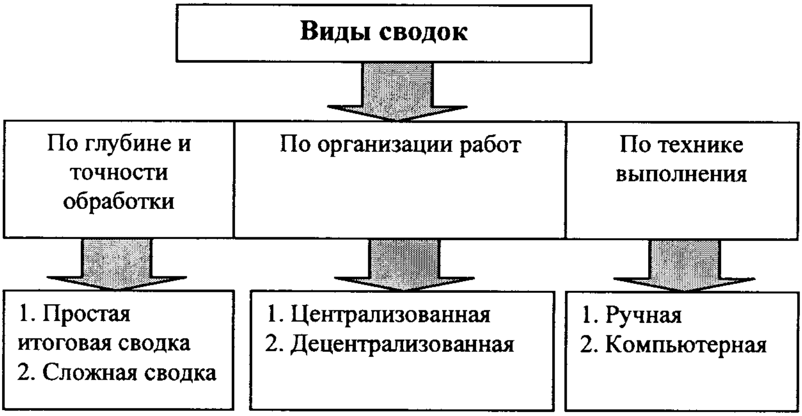


Рис. 3.1. **Классификация сводок**

Простая итоговая сводка проводится без распределения полученных сведений на группы.

Она просто суммирует сведения, содержащиеся во всех формулярах, подводит общий итог единиц совокупности или измеряет общий объем изучаемого показателя.

Например, чтобы получить общую численность студентов достаточно сложить данные по всем высшим учебным заведениям. Или же для получения данных о количестве безработных складывается общее число зарегистрированных безработных.

Сложная сводка, т.е. сводка в широком смысле, представляет сложную операцию научной обработки первичных статистических данных, которая включает группировку материала, разработку системы показателей для характеристики типичных групп и подгрупп, подсчет (подведение) итогов по группам и по совокупности в целом и изображение сгруппированных материалов в виде таблиц.

Результаты сводки и группировки всегда излагаются в виде статистических таблиц.

При **централизованной сводке** материалы наблюдения сосредоточиваются в одном центральном органе, т.е. материалы наблюдения поступают в одну вышестоящую организацию, например, в Госкомстат Республики Узбекистан.

При **децентрализованной сводке** обобщение собранных сведений производится на местах – в районных, городских, областных, краевых органах статистики по единому плану и вышестоящему органу статистики передаются уже сводные итоги для дальнейшего их обобщения.

По технике, или способу, выполнения сводка может быть ручной и компьютерной. В современных условиях доминирующей является компьютерная сводка. Ручная сводка (с применением счетов, арифмометров) применяется крайне редко, как исключение.

Компьютерная разработка материалов статистического материала, т.е. сбор, обработка и хранение информации осуществляется с помощью современных компьютеров.

Вся многомерная и сложная работа по статистической сводке исходной информации подразделяется на следующие этапы:

✓ формулировка задачи сводки на основе целей статистического исследования;

✓ формирование групп и подгрупп, определение группировочных признаков, числа групп и величины интервала; решение вопросов, связанных с осуществлением группировки, включая выделение существенных признаков, установление специализированных интервалов, построение комбинированных группировок;

✓ осуществление технической стороны сводки, т.е. проверка полноты и качества собранного материала, подсчет различных итогов и исчисление необходимых показателей для характеристики всей совокупности и её частей.

В связи с введением рыночной экономики меняются основные способы осуществления статистической сводки:

- ✓ приобретает все большее распространение отчетность внутри предприятия, нежели отчетность на уровне государства;
- ✓ все чаще применяются выборочное обследование, единовременные учеты и другие данные для управления процессами реальных секторов экономики и сферы услуг;
- ✓ качественные изменения претерпевает и централизованная сводка;
- ✓ все с большим размахом происходит применение компьютерных технологий в обработке и хранении информации статистической сводки, что значительно уменьшает время обработки и повышает качество хранимой информации.

Успешное осуществление задач сводки возможно лишь при научном обосновании и правильной организации. Наоборот, при неумелой, непродуманной сводке и группировке первичный материал, каким бы богатым он ни был, теряется, обесценивается «...и экономист, получает в своё распоряжение рутинные, бессмысленные столбцы цифр, статистическую «игру в циферки» вместо осмысленной статистической обработки материала».[1]

План статистической сводки

Даже превосходные данные, правильно характеризующие каждую единицу наблюдения, в результате неумелой сводки могут не дать правильной картины изучаемого явления. «При неудовлетворительной сводке, при неправильной или недостаточной группировке, – может получиться – и постоянно получается при обработке современных переписей – такой результат, что необыкновенно детальные, великолепные данные, имеющиеся о каждом отдельном предприятии, исчезают, теряются, пропадают в целом, когда речь идет о миллионах хозяйств всей страны».[2]

---

[1] Общая теория статистики / Под ред. Т.И.Козлова. М.: Статистика, 1967. С. 86-87.
[2] Статистика: Учебник / Под ред. С.Г. Струмилина. М.: Статистика, 1969. С.69.

Чтобы сводка статистических материалов принесла положительные результаты, она должна базироваться на правильных исходных теоретических положениях. Чтобы выделить и охарактеризовать на основе конкретных данных типы хозяйств, необходимо иметь четкое и ясное представление о том, что представляет собой в социально-экономическом отношении тип хозяйства.

Например, известно, что все фермы как тип хозяйства однородны. Все они – сельскохозяйственные предприятия, но существенно отличающиеся друг от друга количеством обрабатываемой земли, производственным направлением, численностью работников, результатами своей деятельности и т.д.

Наконец, успех сводки во многом зависит <u>от дорожной карты ее проведения</u>. Как всякая большая и сложная работа, статистическая сводка должна осуществляться по заранее составленному плану. В данном случае это особенно важно потому, что в статистической сводке, как правило, участвует большое число лиц. План должен содержать указания о последовательности и сроках выполнения отдельных частей сводки и изложения ее результатов (рис. 3.2).

Дорожная карта

<u>Перечень организационных вопросов</u>:

✓ о последовательности
✓ о сроках выполнения отдельных частей сводки
✓ о ее исполнителях
✓ о порядке изложения и представления результатов

Рис. 3.2. **Понятие о статистической группировке**

Группировка – научная основа сводки

В сводке статистического материала вопрос о группировке представляет самое важное звено, так как простой подсчет итогов без распределения единиц совокупности на группы по тем или иным признакам не дает полной характеристики объекта изучения. Например, если при сводке материалов переписи населения ограничиться только подсчетом итогов, то будут получены лишь сведения об общей численности населения; если при сводке материалов отчетности промышленных предприятий о выполнении плана по выпуску продукции ограничиться простой итоговой сводкой, то будет получен лишь общий итог объема продукции промышленности и т.д. Хотя эти данные дают общую характеристику населения страны, продукции промышленности и т.п., но их крайне недостаточно для всестороннего изучения рассматриваемых объектов.

Чтобы охарактеризовать население подробно и глубоко, нужно знать, как оно распределяется по полу, возрасту, общественным группам, занятиям и т.д. Для анализа выполнения плана выпуска продукции промышленностью нужно знать, как распределяются предприятия по степени технической оснащенности, по уровню специализации, ритмичности выпуска продукции и т.п.

Распределение единиц изучаемого объекта на однородные группы по существенным для них признакам называется статистической группировкой. Группировки позволяют выявить характерные свойства тех или иных типов явлений, обнаружить закономерности изучаемых процессов. Метод статистических группировок позволяет так разработать первичный статистический материал, чтобы все существенные черты и особенности изучаемых общественных явлений получили четкое выражение. Этим определяется роль группировок как научной основы сводки.

В научных исследованиях и в практической экономической работе группировки применяются давно. В подтверждение можно сослаться на таких известных прогрессивных общественных деятелей и ученых, как А.И. Радищев, Н.П. Огарев, Д.П. Журавский и ряд других, которые обращались к методу группировок при изучении общественных явлений. Д.П. Журавский отмечал, что статистическое исчисление носит характер счета по группам, категориям. Известно, что он и статистику определял как науку о «категорическом исчислении».

Способствуя выявлению наиболее существенных черт и признаков изучаемых явлений, выделению и характеристике основных общественно-экономических типов предприятий, хозяйств, групп населения, метод группировок служит вместе с тем основой научного применения других методов статистики – средних величин, индексов и т.д. Характерное для статистики исчисление сводных, обобщающих показателей лишь в том случае не приводит к затушевыванию существенных различий внутри изучаемых совокупностей явлений, когда оно производится на основе научно обоснованной группировки материала. Обобщающие показатели, затушевывающие существенные различия в разнородной по своему составу совокупности, будут, по словам статистиков, огульными, фиктивными.

Группировка статистических данных – не самоцель. Приступая к группировке статистического материала, нужно четко определить ее цели и задачи, так как группировка ради группировки не имеет никакого научного значения и является, по выражению статистиков, бессмысленной «игрой в циферки».

3.3. Задачи статистических группировок и их виды

Содержание и приемы группировок многообразны. Различны и задачи, выполняемые ими.

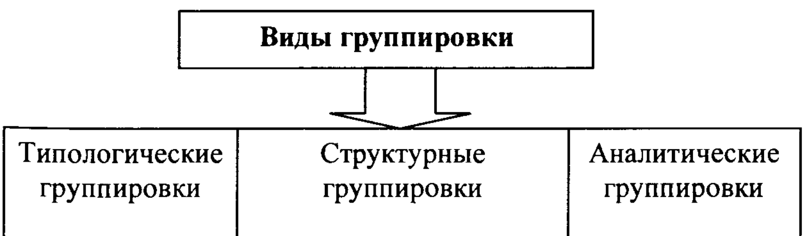


Рис. 3.3. **Виды статистических группировок**

Однако принято выделять следующие основные задачи, решаемые с помощью метода статистических группировок: образование социально-экономических типов явлений; изучение строения изучаемых явлений и структурных изменений, происходящих в них; выявление связи между изучаемыми признаками.

Для решения этих задач, соответственно, применяют типологические, структурные и аналитические группировки (рис. 3.3).

Типологическая группировка

Типологической называется группировка, задача которой состоит в выделении и характеристике сложившихся в реальной действительности типов общественных явлений. Важнейшим их содержанием является выделение из множества признаков, характеризующих изучаемые явления, основных типов в качественно однородные.

Например, предприятия могут быть сгруппированы по видам собственности: государственная, частная, смешанная; группировки населения по половозрастному или по уровню экономической активности и т.п.

Приведем пример типологической группировки (табл. 3.1).

Таблица 3.1

Основные показатели рынка труда

(в среднем за год, тыс. чел.)[3]

№	Группы населения по уровню активности	2016	2017	В % к итогу	
				2016	2017
1	Экономически активное население	13505,4	114357,3	72,5	73,6
2	Экономически неактивное население	4542,6	4309,0	24,2	22,1
3	Безработные	687,0	837,0	3,7	4,3
	Всего				

[3] Социальное развитие и уровень жизни в Узбекистане. Статсборник. Госкомстат РУз. Т., 2018. С. 48.

В данной таблице выделены различные типы (группы) населения по уровню активности в участии рынка труда. Подавляющая часть населения активно участвует в сфере рынка труда. Только 837 тыс. человек являются безработными. Подобное сопоставление позволяет сделать анализ социально-экономических типов трудоспособного населения.

Структурная группировка

Структурной называется группировка, которая служит для характеристики структуры изучаемой совокупности, то есть удельного веса в ней отдельных групп и подгрупп (табл. 3.1).

Данные таблицы показывают изменения, которые произошли за 2016-2017 гг. в структуре занятых на рынке труда Узбекистана. Так, повысился удельный вес экономически активного населения на 1,5 процентного пункта. Это повышение сопровождалось понижением уровня экономически неактивного населения на 2,1 процентного пункта. Однако, как это не приятно, уровень безработных увеличился на 0,6 процентного пункта.

Аналитическая группировка

Под аналитической понимается группировка, в задачу которой входит выявление и характеристика взаимосвязи между признаками изучаемых явлений. При этом один признак рассматриватся как фактор, другой – как результат. Аналитическую группировку строят по факторному признаку (причина – первична, а следствие – вторично).

Факторные признаки обусловливают изменения явлений, а результативные изменяются под влиянием факторных. Например, повышение квалификации рабочих (факторный признак) приводит к росту производительности труда (результативный признак).

Ниже приведена аналитическая группировка, в которой в качестве группировочного (факторного) признака взята урожайность картофеля в фермерских хозяйствах Хорезмской области, а в качестве результативного – его себестоимость (табл. 3.2).

Таблица 3.2

Зависимость себестоимости 1 ц картофеля от урожайности в фермерских хозяйствах Хорезмской области (2019 г.)

Группы фермерских хозяйств по уровню урожайности картофеля	Количество фермерских хозяйств	Средняя себестоимость 1 кг (сум)
I. До 110	13	602
II. 110-140	46	502
III. 140-170	58	470
IV. 170-200	28	413
Более 200	17	353
Всего	162	454

Из таблицы видно, что урожайность картофеля значительно влияет на его себестоимость. Взаимосвязь признаков обратная: с ростом факторного признака (урожайности) значения результативного признака (себестоимость) уменьшаются. Таблица показывает далее, что снижение себестоимости происходит непропорционально росту урожайности: от I группы ко II она снижается на 100 сум, а от IV к V группе хозяйств всего на 60 сум. Это свидетельствует о криволинейной (гиперболической) форме зависимости. Более подробное и точное исследование формы и тесноты статистических зависимостей ведется методом корреляционного анализа (см. главу 8).

Следует отметить, что разделение группировок на эти три вида по выполняемым ими задачам имеет некоторую цельность, поскольку они на практике применяются в комплексе. Одну и ту же группировку можно представить и как типологическую и как структурную. Традиционно обе эти группировки характеризуют состав совокупности как в абсолютном, так и относительном измерении. Например, распределение населения по уровню доходов.

Для ***типологической*** группировки характерным будет такое распределение:

- малообеспеченное население;

- население со средним достатком (средний класс);
- население с высоким уровнем дохода.

Для *структурной* группировки те же группы населения будут представлены в виде относительных величин, в процентах к итогу.

3.4. Группировочные признаки и их отбор

Принципы отбора группировочного признака

Признаки, положенные в основание группировки, называются группировочными.

Первым и наиболее сложным вопросом теории группировок является правильный выбор признаков, по которым будет производиться группировка. Сложность вопроса заключается в том, что политико-экономические и социальные понятия следует перевести на язык цифр. Для этого нужно найти такие статистические признаки, которые бы правильно и наиболее полно отражали эти понятия. При отборе признаков можно руководствоваться следующими положениями:

1. В качестве основания группировки необходимо брать типичные, существенные признаки изучаемого явления в соответствии с целями проводимой статистической работы.

2. К вопросу о выборе группировочных признаков нельзя подходить догматически. Одни и те же признаки могут быть положены в основание группировки в одних случаях и не годятся в других. Поэтому при выборе группировочных признаков должны быть приняты во внимание конкретные условия места и времени.

3. При изучении сложных явлений общественной жизни группировку следует проводить не по одному, а по нескольким существенным, характерным признакам. Это позволяет наиболее полно охарактеризовать изучаемое явление.

По числу группировочных признаков различают следующие виды группировки (рис. 3.4).

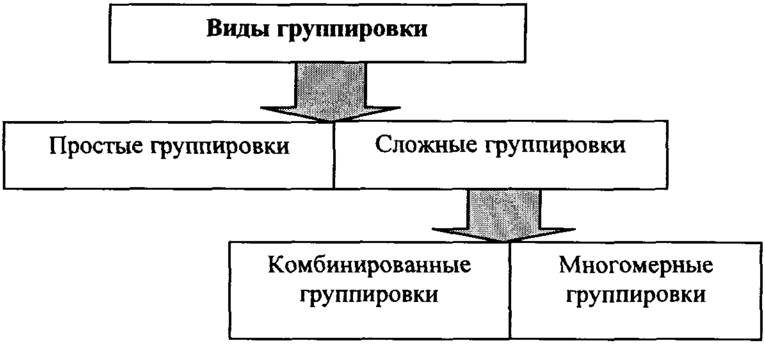


Рис. 3.4. **Виды группировок по числу группировочных признаков**

1. ***Простые группировки*** – имеют один группировочный признак.

2. ***Сложные группировки*** – имеют два и более группировочных признака. Они подразделяются на подвиды:

➢ ***комбинационные группировки*** – имеют от двух до четырех (включительно) группировочных признаков и они строятся по следующему принципу: сначала группы формируются по первому признаку, затем эти группы делятся на подгруппы по второму признаку, которые, в свою очередь, делятся по третьему признаку и, наконец, – по четвертому.

Комбинационные группировки позволяют проводить комплексный анализ сложных социально-экономических явлений. Однако чрезмерное дробление информации нередко затушевывает проявление закономерностей, что может сделать подобный анализ малоэффективным, поэтому на практике обычно используют не более трех группировочных признаков;

➢ ***многомерные группировки*** – имеют свыше четырех группировочных признаков (пять, шесть и т.д.), т.е. осуществляются по целому комплексу признаков. Они получили распространение благодаря использованию компьютеров и программных продуктов, позволяющих перерабатывать любые объемы информации с различной степенью детализации.

По сравнению с простыми комбинационные группировки обладают дополнительными аналитическими свойствами. Они помогают выявить такие различия и связи между исследуемыми признаками, которые нельзя обнаружить при простой группировке.

Отбор группировочных признаков

Отбор группировочных признаков проходит следующие стадии: вначале определяется цель, познавательная задача предполагаемой группировки, затем определяется специфическое содержание признаков, которые должны быть положены в основание группировки, устанавливаются число групп и количественные границы признаков. Определение числа групп и количественных границ признаков зависит от цели группировки и от того, с какими признаками приходится иметь дело.

Совокупности, изучаемые статистикой, характеризуются многими свойствами и выражаются различными признаками. Различают четыре вида ***группировочных признаков: атрибутивные, количественные, факторные и результативные*** (рис. 3.5).

Группировочными называются признаки, которые положены в основу группировки.

Атрибутивным называется такой признак, который характеризует свойство, качество данного явления и не имеет количественного выражения. При группировке по такому признаку группы отличаются друг от друга не по величине, а по характеру признака.

Число групп, на которые делится изучаемая масса, часто определяется числом разновидностей атрибутивного признака.

Так, группировка населения по полу дает две группы. Группировка же населения по национальности допускает образование стольких групп, сколько различных национальностей имеется среди изучаемого населения. Примером атрибутивной группировки может служить табл. 3.1.

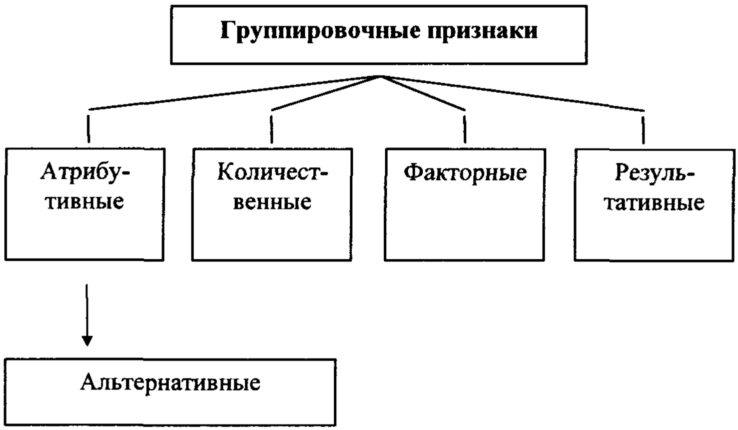


Рис. 3.5. **Основания** (признаки) **группировки**

Частным случаем атрибутивной группировки является *альтернативная группировка,* когда имеются два варианта атрибутивного признака и при этом один из них исключает другой. Например, одни имеют высшее образование, а другие не имеют; одни знают иностранный язык, а другие – нет, одни получают стипендию, а другие – нет и т.д.

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

При группировке по количественным признакам изучаемую совокупность разделяют по уровню, или величине, признака, Например, при группировке вагонов – по числу осей, автомашин – по грузоподъемности, рабочих – по тарифному разряду, семей – по числу членов и т.д.

На основе типологических и структурных группировок могут выявляться и изучаться взаимосвязи между факторными и результативными признаками. Так, например, в торговле и сфере быта встречается большое разнообразие взаимосвязей между признаками, выступающими в роли причины или следствия явления. Из них можно выделить следующие:

1) когда фактором выступает количественный признак, а результативным – качественный (стаж работы и квалификация продавца, продолжительность договорных связей между поставщиками товаров и предприятиями торговли, с одной стороны, и качеством товаров – с другой);

2) когда в основу группировки положен качественный признак, а результативным – представлен количественный (например, квалификация продавцов и производительность их труда);

3) когда в роли фактора и результата выступает качественный признак (например, категории работников торговли и их образование);

4) когда в группировке факторный и результативный показатели представлены количественным признаком (например, производительность труда и заработная плата).

Вопросы изучения взаимосвязи общественных явлений более подробно рассматриваются в гл. 8.

3.5. Принципы образования групп и интервалов группировки

Основные требования к образованию групп

Следующим важным шагом после определения группировочного признака является распределение единиц совокупности по группам. Здесь встает вопрос о количестве групп и величине интервала, которые между собой взаимосвязаны. При прочих равных условиях, чем больше число групп, тем меньше величина интервала и наоборот. Одним из основных требований, возникающих при решении данного вопроса, является выбор такого числа групп и величины интервала, которые позволяют более равномерно распределить единицы совокупности по группам и достичь при этом их представительности, качественной однородности. Оптимальная наполняемость интервалов является важным критерием правильности группировки.

Различают следующие виды интервалов (рис. 3.6).

Величиной интервала называется разность между максимальным и минимальным значениями признака в каждой группе. В зависимости от характера распределения совокупности по данному признаку интервалы по величине могут быть ***равными*** и ***неравными.***

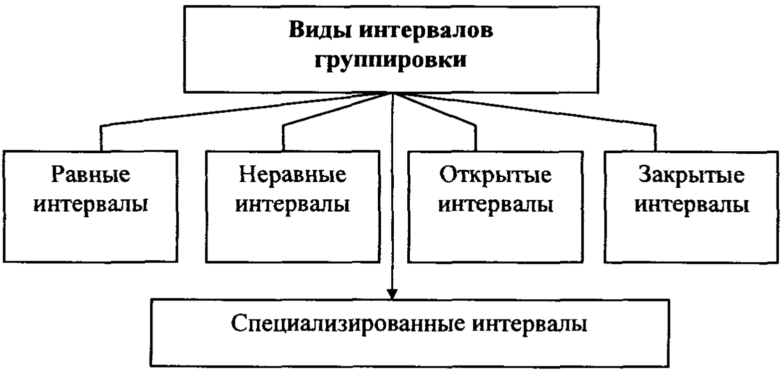


Рис. 3.6. **Виды интервалов группировки**

Равные интервалы

Это интервалы, размеры которых имеют во всех группах одну и ту же величину. Они применяются тогда, когда группировочный признак изменяется более или менее равномерно в небольших пределах. Величина равных интервалов определяется по формуле

$$i_p = \frac{p_1}{p_0}, \qquad (1)$$

где X_{max} и X_{min} – соответственно, максимальное и минимальное значения признака в данной совокупности, а n – число образуемых групп. Если, например, минимальная урожайность составляет 25 ц/га, максимальная – 55 ц/га, а число намеченных групп – 5, то величина интервала будет равна:

$$\frac{\sum p_1 q_1}{\sum p_0 q_1} \text{ ц/га.}$$

Прибавляя к минимальному значению признака (в данном примере 25 ц/га) найденное значение интервала, получаем верхнюю границу первой группы: 25 + 6 = 31 ц/га. Прибавляя далее величину интервала к верхней границе первой группы, получаем верхнюю границу второй группы и т.д. Окончательно группировка примет следующий вид (табл. 3.3).

Равные интервалы применяются обычно при относительно узких пределах вариации признака и распределении единиц, близком к равномерному, например, при группировке рабочих по размерам заработной платы, студентов по размерам стипендии и т.д.

Таблица 3.3

Группы фермерских хозяйств по уровню урожайности хлопка (ц/га)

№	Группы хозяйств по уровню урожайности (ц/га)	Количество фермерских хозяйств
I	25 – 31	5
II	32 – 38	19
III	39 – 45	50
IV	46 – 52	12
V	53 – 59	4

Число групп тесно связано с объемом совокупности. Здесь нет строго научных приемов, позволяющих решать этот вопрос при любых взаимосвязях названных величин. Всякий раз эта задача решается с учетом конкретных обстоятельств. Однако при равенстве интервалов для ориентировки существует формула, предложенная американским ученым Стерджессом, с помощью которой можно наметить число групп n при известной численности совокупности N:

$$n = 1 + 3{,}322 \lg N.$$

При 200 единицах совокупности число групп определяется следующим образом:

$$1 + 3{,}322 \lg 200 = 9.$$

Неравные интервалы Это интервалы, размеры которых изменяются по величине от группы к группе. Неравные интервалы применяются для отграничения групп в тех случаях, когда группировочный признак изменяется неравномерно или в больших пределах. Они применяются чаще равных интервалов и делятся на интервалы ***возрастающие*** и ***убывающие***. Возрастающие интервалы увеличиваются от одной группы к другой, а убывающие – уменьшаются.

Например, известно, что концентрация производства приводит к увеличению объема продукции. Об этом свидетельствуют данные нижеприведенной табл. 3.4.

Таблица 3.4

Группировка промышленных предприятий по общему объему продукции за год (в % к итогу)

	Число предприятий	**Объем продукции**	**Среднегодовая численность рабочих**
Предприятия, состоящие на самостоятельном балансе (без электростанций, электросетей и теплосетей)	100,0	100,0	100,0
В том числе с объемом продукции, млн сум			
До 500	18,1	0,3	1,1
501 –1000	12,0	0,7	1,8
1 001 – 5 000	34,6	6,3	12,3
5 001 – 10 000	12,5	6,6	9,8
10 001 – 50 000	16,9	27,6	29,9
50 001 – 100 000	3,2	16,2	15,2
100 001 – более	2,7	42,3	29,9

Эта группировка показывает высокую степень концентрации в промышленности страны, т.е. сосредоточения производства в крупных предприятиях. Предприятия с объемом продукций 100 трлн. сум в год и более составляли за год 2,7 % всего числа промышленных предприятий, но на них производилось почти более 42 % промышленной продукции. Кроме того, данные таблицы в какой-то степени характеризуют проявление закономерности состоящей в том, что концентрация производства гораздо сильнее, чем концентрация рабочих, потому что труд в крупных заведениях гораздо производительнее. Действительно, на этих крупнейших предприятиях сосредоточено, как мы видим, почти 60 % производства продукции, в то время как промышленных рабочих сосредоточено меньше – около 45 %. Обычно при исследовании экономических явлений применяются неравные, чаще всего прогрессивно увеличивающиеся интервалы.

Широкое применение неравных, прогрессивно увеличивающихся интервалов объясняется тем, что для большинства экономических явлений количественные изменения размера признака имеют неодинаковое значение в низших и высших по размеру признака группах. Так, если разница в 50, а тем более в 100 рабочих мест существенна для мелких промышленных предприятий, на которых насчитывается всего 100–200 рабочих, то для предприятий, имеющих 10 000 и более рабочих, такая разница существенного значения не представляет. То же можно сказать и о площади обрабатываемой земли при группировке фермерских хозяйств и т.п.

Открытые и закрытые интервалы

Различают ***открытые*** и ***закрытые*** интервалы. Открытые интервалы имеют одну какую-либо обозначенную границу, верхнюю или нижнюю, или неопределенные границы, закрытые – и верхнюю, и нижнюю. Для того чтобы четко установить числовые границы интервалов, даются дополнительные указания к ним, которые выражаются словами «до», «от – до», «свыше», «более», «менее» и т.д. (табл. 3.3 и 3.4).

Для решения ряда задач применяются так называемые ***специализированные интервалы*** и ***принцип равных частот.***

Специализированные интервалы

Это интервалы, различные для разных отраслей и производств, группируемых по одному и тому же признаку, с учетом особенностей каждой отрасли, производства. Например, группировка заводов по числу рабочих в автомобильной промышленности будет иметь одни интервалы, а в легкой – другие, или группировка предприятий по размеру основного капитала в химической промышленности будет иметь одни интервалы, а в пищевой – другие.

Принцип равных частот

При небольшом числе наблюдений для образования групп применяется ***принцип равных*** частот. В этом случае все элементы совокупности располагаются в порядке возрастания признака и распределяются равномерно, так что в каждую группу входит одинаковое количество единиц. Это исключает образование групп с незначительным количеством единиц совокупности.

После того как выбран группировочный признак, намечено число групп, необходимо составить систему показателей, которыми они будут характеризоваться, и определить их величину по каждой группе.

3.6. Вторичная группировка

Необходимость вторичной группировки

Как уже отмечено выше, в процессе обработки первичных данных в ряде случаев может быть образовано больше групп, чем это необходимо для социально-экономической характеристики исследуемых процессов. В этих случаях возникает необходимость в образовании новых групп, обычно укрупненных, на основе ранее осуществленной группировки.

Процесс образования новых групп на основе группировки, произведенной по первичным данным, называется вторичной группировкой.

Необходимость во вторичной группировке, кроме отмеченного случая, возникает также и тогда, когда требуется свести воедино или сопоставить между собой данные, полученные в результате применения различных приемов группировки. Если, допустим, при группировке рабочих по производственному стажу на одном предприятии образовано шесть групп с одними интервалами, а на другом предприятии выделено девять групп с другими интервалами, то данные о распределении рабочих, фонда заработной платы и других показателей по первому предприятию будут непосредственно несопоставимы с данными по второму предприятию. Чтобы сделать эти данные сопоставимыми, необходимо, очевидно, выделить по обоим предприятиям одно и то же число групп с одними и теми же интервалами, т.е. прибегнуть к вторичной группировке.

Вторичная группировка может быть осуществлена по признаку, положенному в основу первичной группировки, или по удельному весу групп в общей численности единиц совокупности, подвергнутой первичной группировке. При этом в обоих случаях показатели новых групп, т.е. групп, образованных при вторичной группировке, могут быть получены путем простого суммирования показателей первичной группировки или в результате специальных расчетов.

Проиллюстрируем сказанное на примерах. Сначала произведем вторичную группировку по группировочному признаку первичной группировки, использовав для этой цели результаты группировки, изложенные в следующей табл. 3.5.

Чтобы данные первого и второго заводов сделать сопоставимыми между собой, необходимо сопоставлять одни и те же группы рабочих по производственному стажу. Возьмем в качестве основы группы рабочих первого завода. В этом случае данные по-второму заводу надо подвергнуть вторичной группировке.

Выделить группу рабочих с производственным стажем до двух лет по второму заводу не представляет особого труда. Для этого надо объединить две первые группы рабочих этого завода со стажем до 1 года и 1–2 года.

Просуммировав соответствующие данные этих двух групп, найдем, что удельный вес рабочих объединенной группы составляет 18% (8+10), а удельный вес фонда заработной платы равен 13% (5 + 8).

Таблица 3.5

Группировка рабочих двух заводов по производственному стажу

Завод № 1			**Завод № 2**		
группы рабочих по производственно-му стажу	**количество рабочих в процентах к итогу**	**заработная плата в процентах к итогу**	**группы рабочих по производственному стажу**	**количество рабочих в процентах к итогу**	**заработная плата в процентах к итогу**
До 2 лет	15	10	До 1 года	8	5
2– 5 лет	20	16	1–2 года	10	8
			2–4 года	18	16
5–10 лет	30	30	4–6 лет	20	18
10–15 лет	20	24	6–8 лет	15	17
			8–10 лет	15	18
15–20 лет	10	12	10–15 лет	8	10
Свыше 20 лет	5	8	15–20 лет	4	5
			Свыше 20 лет	2	3
Итого	100	100	Итого	100	100

Для образования группы рабочих со стажем 2–5 лет по второму заводу необходимо произвести некоторые расчеты. В эту группу целиком войдет группа рабочих второго завода со стажем 2–4 года и часть группы со стажем 4–6 лет. Эту последнюю группу необходимо раздробить на две следующие подгруппы: 4–5 и 5–6 лет. Первая подгруппа войдет в группу рабочих со стажем 2–5 лет, а вторая – в группу рабочих со стажем 5–10 лет. Показатели (в нашем примере процент рабочих и процент фонда заработной платы) этих подгрупп обычно определяют таким образом, что в каждую из них относят ту часть величины показателя дробимой группы, которая взята от величины интервала этой группы. В рассматриваемом случае интервал дробимой группы, равный двум годам (4–6), расчленяется на две равные части. Поэтому и проценты рабочих и фонда заработной платы необходимо расчленить также на две равные части. Для названных показателей это составит, соответственно, 10% (20/2) и 9% (18/2). После этого нетрудно уже подсчитать проценты к итогу для группы рабочих со стажем 2–5 лег второго завода. Процент рабочих этой группы равен 28 (18+10), процент фонда заработной платы – 25 (16+9).

Подсчитав аналогичным способом показатели по группе рабочих второго завода со стажем 5–10 лет, получим следующие сопоставимые данные по двум заводам (табл. 3.6).

Таблица 3.6

Группировка рабочих по производственному стажу двух заводов

Группы рабочих по производственному стажу		Завод № 1		Завод № 2	
		количество рабочих в процентах к итогу	заработная плата в процентах к итогу	количество рабочих в процентах к итогу	заработная плата в процентах к итогу
До 2	лет	15	10	18	13
2–5	-"-	20	16	28	25
5–10	-"-	30	30	40	44
10–15		20	24	8	x0
15–20	-"-		12	4	6
Свыше	20 лет	5	8	2	3
	Итого	100	100	100	100

Изложенный выше прием расчета показателей при дроблении групп основан на предположении, что в пределах каждой группы численность единиц распределяется равномерно. В действительности это может иметь место лишь в редких случаях. В приведенном примере процент рабочих к итогу на втором заводе с увеличением производственного стажа сначала возрастает, а затем, достигнув максимума при стаже 4–6 лет, уменьшается. В группе рабочих со стажем 4–6 лет на половину интервала, равную одному году, приходится 10% рабочих, а в следующей группе на точно такую, часть интервала – только 7,5%. Это означает, что при дроблении первой, из только что упомянутых групп, по изложенному способу несколько преуменьшается процент рабочих вновь образованной по второму заводу группы рабочих с производственным стажем 2–5 лет и несколько преувеличивается группа рабочих со стажем 5–10 лет.

Неточности подобного рода имеют место также при дроблении данных по другим показателям. Эти неточности имеют тем больший размер, чем неравномернее распределяются единицы совокупности по группам.

3.7. Статистические ряды распределения

Дальнейшим развитием статистической группировки является статистический ряд распределения.

Ряд распределения – упорядоченное распределение единиц на группы по группировочному признаку. Ряды распределения могут быть образованы по атрибутивному и количественному признакам.

1. ***Атрибутивные ряды распределения*** – ряды распределения, образованные по атрибутивным (описательным) признакам. Например, по видам собственности.

2. ***Вариационные ряды распределения*** – ряды распределения, образованные по количественному признаку. При построении рядов распределения проводят ранжирование.

Ранжированный ряд – ряд единиц статистической совокупности, построенных по рангу, в порядке возрастания (или убывания) изучаемого признака.

Вариационные ряды по способу построения подразделяются на подвиды:

1. ***Дискретный ряд распределения*** – вариационный ряд, в котором группы составлены по дискретному признаку (целые числа), имеющему довольно ограниченное число вариантов. Например, ряд распределения студентов по их оценкам; ряд распределения рабочих по их разрядам; ряд распределение семей по числу детей и т.д.

2. ***Интервальный ряд распределения*** – вариационный ряд, в котором группы составляются с использованием интер-

валов, при этом целесообразность их построения объясняется непрерывностью признака (дробные числа) или большим числом вариантов дискретного признака.

Например, интервальные ряды распределения предприятий по объемам выпускаемой продукции, по среднесписочной численности рабочих, по фондам оплаты труда и т.п.

Конструктивно ряд распределения определяют два элемента:

•***вариант (x)*** – отдельное значение группировочного признака в ряду распределения.

•***частота (f)*** – число повторений вариантов (*frequency* (англ.) – частота). Частота показывает, как часто встречается тот или иной вариант. Сумма частот составляет объем ряда распределения (или объем статистической совокупности – суть одно и то же):

$$\sum_{j=1}^{k} f_i = n,$$

где j – порядковый номер группы в ряду распределения; k – число групп в ряду распределения.

Объем признака для вариационного ряда распределения будет определяться как

$$\sum_{j=1}^{k} x_j f_i.$$

Ряд распределения может быть дополнен ***частостями (w)*** – частотами, выраженными в виде относительных величин структуры – процентах или долях единицы.

Сумма частостей равна 1 (или 100%).

$$\sum_{j=1}^{k} w_j = 1(100\%).$$

Обычно используются ряды с равными интервалами. Если вариационный ряд содержит неравные интервалы, то частоты интервалов становятся несопоставимы, и тогда для статистического анализа рассчитывают плотность распределения: абсолютную или относительную.

Абсолютная плотность распределения – частота, приходящаяся на единицу длины интервала в вариационном ряду распределения:

$$f'_j = \frac{f_j}{h_j},$$

где h_j – длина j-го интервала в вариационном ряду распределения с неравными интервалами.

Относительная плотность распределения – частость, приходящаяся на единицу длины интервала в вариационном ряду распределения:

$$w'_j = \frac{w_j}{h_j}.$$

3.8. Статистические таблицы

Общее понятие о статистических таблицах

Результаты сводки статистических данных изображаются в виде таблиц. Но не всякая таблица является статистической.

Статистическими называются такие таблицы, которые дают сводную количественную характеристику статистической совокупности. Статистические таблицы представляют собой форму наиболее рационального, наглядного и систематизированного изложения результатов сводки и обработки статистических материалов. Они дают возможность компактно и выразительно отражать цифровые характеристики исследуемых общественных явлений и, процессов, показывают состояние изучаемых явлений, служат основой для сравнения, сопоставления, анализа и определенных выводов. Отличительной особенностью статистических таблиц является наличие частных и общих итогов или возможность их получения.

Статистическая таблица представляет собой комбинации горизонтальных строк и вертикальных граф (колонок, столбцов). Чаще всего строки и графы отделяются друг от друга, прямыми линиями, пересечение которых образует остов, *костяк* таблицы, а каждое такое пересечение – клетку таблицы (рис. 3.7).

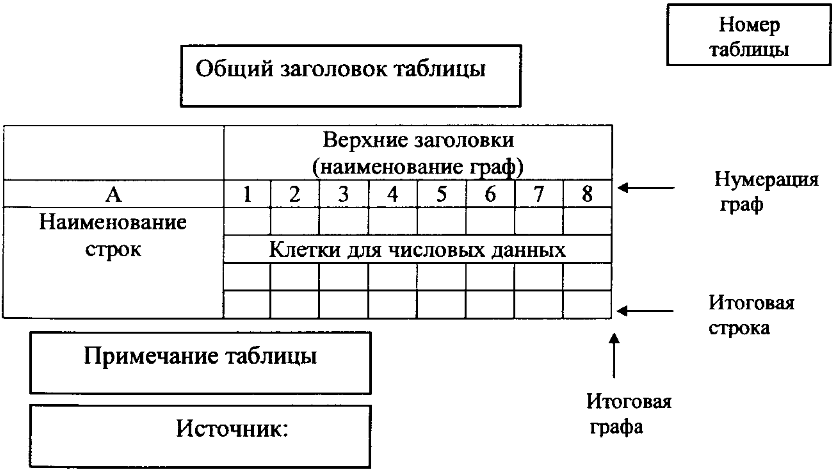


Рис 3.7. **Макет таблицы**

Составными элементами таблицы являются общий, боковые и верхние ***заголовки.*** Каждая таблица должна **иметь** краткое, но точное и ясное ***заглавие,*** полностью ***отражающее*** ее основное содержание. В общем заголовке должен быть выражен главный смысл помещенных в таблице цифровых данных, а также указано, к какой территории и к какому периоду времени они относятся. Кроме того, в названии заглавия даются единицы измерения. Остов таблицы в котором заполнены общий, боковые и верхние заголовки, но только без числовых данных, называется ***макетом*** таблицы: Если данные отличаются какими-либо особенностями, например, относятся только к части территории, являются предварительными и т.д., то делаются надлежащие оговорки или пояснения в виде примечаний и сносок. Примечания, как правило, относятся ко всей таблице, а сноски – к отдельным строкам, графам или клеткам.

Статистическую таблицу можно представить себе как своеобразное логическое предложение. Поэтому каждая таблица имеет статистическое ***подлежащее*** и ***сказуемое.***

Подлежащее и сказемое таблицы

Подлежащим называются статистические совокупности, которые характеризуются различными показателями, ***сказуемым*** – показатели, характеризующие эти совокупности. Подлежащее, как правило, помещается в горизонтальных строках таблицы, а сказуемое – в вертикальных графах.

По строению подлежащего различают простые, групповые и комбинационные таблицы.

Подлежащее ***простой таблицы*** представляет собой перечень дат или отдельных единиц совокупности. Соответственно, простые таблицы могут быть перечневыми, динамическими и территориальными. Часто они строятся в различном сочетании (перечневых и динамических, территориальных и динамических).

Так как простые таблицы дают лишь итоговую сводку и недостаточны для выявления типа изучаемого явления, его структуры и взаимосвязей, то применяются групповые и комбинационные таблицы.

Если в подлежащем таблицы даны группы по одному определенному признаку, то такая таблица называется ***групповой.*** Если же каждую группу разделить еще на подгруппы по одному или нескольким другим признакам, то получим ***комбинаиионную таблицу***.

По строению сказуемого различают простую и комбинированную группировку. При простой группировке показатели сказуемого разрабатываются отдельно друг от друга.

Комбинированная разработка сказуемого дает возможность выявить не только число родившихся по полу, но и отдельно распределение родившихся по полу в городской и сельской местностях.

Необходимо уметь читать и анализировать таблицы. Вначале знакомятся с названием таблицы, заголовками подлежащего и сказуемого, определяют, какие явления характеризуются таблицей, устанавливают единицы измерения, время, к которому относятся данные таблицы, а затем приступают к анализу. Анализ таблиц начинают обычно с рассмотрения общих итогов, затем групповых, частных итогов, и только после этого переходят к анализу отдельных строк и граф.

Общие правила составления таблиц

Чтобы таблица была наиболее выразительна, необходимо при ее составлении придерживаться следующих *правил:*

1. Таблицы должны быть небольшими по размеру. При сложной характеристике совокупности следует избегать громоздких таблиц – лучше составить две или несколько более простых, взаимосвязанных между собой таблиц.

2. Общее наименование таблицы, заголовки подлежащего и сказуемого, отдельных строк и граф должны быть сформулированы четко и кратко.

3. В таблице должны быть указаны место и время, к которому относятся данные, а также единицы их измерения.

4. Показатели подлежащего и сказуемого необходимо расположить в определенной логической последовательности и согласовать между собой, так как строки подлежащего пересекаются с графами сказуемого.

5. Если число показателей подлежащего и сказуемого таблицы значительно, то строки и графы таблиц следует пронумеровать, причем графы подлежащего обозначать буквами («а», «б» и т.д.), а графы сказуемого – цифрами. Это значительно облегчает пользование таблицей, а в ряде случаев дает возможность пояснить, какие действия надо произвести, чтобы получить те или иные показатели (например, графа 2 + графа 3, графа 5 : графу 8 и т.д.).

6. Необходимо строго соблюдать условные обозначения. Грамотно составленная таблица не должна иметь пустых незаполненных клеток. Если сведений нет, проставляются точки (...) или пишется «нет сведений», если явление отсутствует, ста-

вится прочерк (–). При наличии клеток, не подлежащих заполнению, в них проставляется знак (х), если явление есть и данные о нем есть, но значение их крайне малы и с принятой в таблице точностью превращаются в нуль, то пишется 0 (0,0 или 0,00 и т.д.).

7. Данные всех граф и строк таблицы должны приводиться с одинаковой степенью точности. Если числовые значения меньше принятой в таблице точности, проставляется 0,0.

8. Таблицы могут иметь сноски, в которых указываются источники приводимых данных, примечания, дающие пояснения, расшифровку показателей и т.д.

9. Таблицы, как правило, должны быть замкнутыми, т.е. иметь итоги по группам, подгруппам («всего») и в целом по таблице («итого»).

10. Анализ таблицы начинают с чтения общего итога. Затем переходят к анализу промежуточных итогов, причем сначала обращают внимание на наиболее значительные среди них в количественном измерении. Далее проводят анализ отдельных строк и граф, при этом сначала рассматриваются экстремальные, значения (min, max).

Не следует забывать, что один и тот же материал дает диаметрально противоположные выводы при различных приемах группировки.

3.9. Статистические графики

Сущность и назначение графиков

Полученный в результате разработки статистический материал, расположенный в таблицах, часто нуждается в наглядном изображении с помощью построения статистических графиков.

В статистике ***графиками*** называют наглядные изображения статистических величин в виде различных линий, геометрических фигур или географических картосхем. Главное достоинство графиков – наглядность. При правильном построении графика статистические показатели привлекают к себе внимание, становятся выразительными, лаконичными и запоминающимися.

Графики дают *целостную картину* явлений и процессов, *обобщающее представление* о них и помогают осмыслить статистический материал. При графическом изображении статистических данных становятся особенно отчетливыми и наглядными взаимная связь между явлениями и процессами общественной жизни, основные тенденции их развития, степень распространения в пространстве.

Графики позволяют мгновенно охватить и осмыслить совокупность показателей – выявить наиболее типичные соотношения и связи этих показателей, определить тенденции развития, охарактеризовать структуру, степень выполнения плана, оценить географическое размещение объектов. Этим объясняется широкое применение графиков для пропаганды статистической информации, характеризующей результаты развития различных сфер национальной экономики и социальных отношений.

Применение графиков в статистике насчитывает более чем двухсотлетнюю историю. Основоположником графического метода в статистике коммерческой деятельности считают английского экономиста У. Плейфейра (1731-1798). В его работе «Коммерческий и политический атлас» (1786) впервые были применены способы графического изображения статистических данных (линейные, столбиковые, секторные и другие диаграммы).

Основные элементы статистического графика

В статистическом графике различают следующие основные элементы:

- графический образ (столбики, линии, точки и т.д.);
- поле графика;
- пространственные ориентиры (система координат);
- масштабные ориентиры;
- экспликация (словесное описание содержания графика – название графика, подписи к масштабным шкалам и пояснения к отдельным частям графика).

Графический образ Это символические знаки, с помощью которых изображаются статистические данные. Они весьма разнообразны: линии, точки, плоские геометрические фигуры (прямоугольники, квадраты, круги и т.д.). В качестве графического образа выступают и объемные фигуры. Иногда в графиках используются негеометрические фигуры в виде силуэтов или рисунков предметов.

Одни и те же статистические данные можно изобразить с помощью различных графических образов. Поэтому при построении графика важен правильный подбор графического образа. Он должен наиболее доходчиво отображать изучаемые показатели и соответствовать основному предназначению графика.

Система координат ***Пространственные ориентиры*** определяют размещение графических образов на поле графика. 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. Но могут быть и графики, построенные по принципу полярных координат (круговые графики).

В так называемых ***статистических картах*** средствами пространственной ориентации выступают географические ориентиры (контуры суши или линии рек, морей и океанов и т.д.). Пространственные ориентиры позволяют определять расположение графических образов на поле графика.

Масштабные ориентиры Масштабные ориентиры статистических графиков – это масштаб, масштабные шкалы и масштабные знаки. Масштаб – это условная мера перевода числовой, величины в графическую и обратно. При построении графика масштаб выбирается такой, чтобы подлежащие нанесению на график данные поместились на поле. На вертикальной шкале графика должна быть обязательно нулевая линия. В тех случаях когда минимальное значение признака намного выше нуля, целесообразно дать разрыв вертикальной шкалы с тем, чтобы поле графика было заполнено равномерно (рис. 3.8).

Это линия, разделенная на отрезки точками. Расстояние между соседними точками называют графическим интервалом. Наиболее часто в статистических графиках используются прямолинейные масштабные шкалы, располагающиеся по осям координат.

Они могут быть равномерными или неравномерными. В равномерной масштабной шкале графические интервалы строго пропорциональны числам: если значение показателя возрастает в три раза, то отрезок шкалы должен быть увеличен также в три раза. Примером неравномерной шкалы является логарифмическая шкала, в которой графические интервалы пропорциональны не числам, а логарифмам этих чисел (рис. 3.8). Логарифмическая шкала применяется для изображения темпов роста за продолжительный период времени.

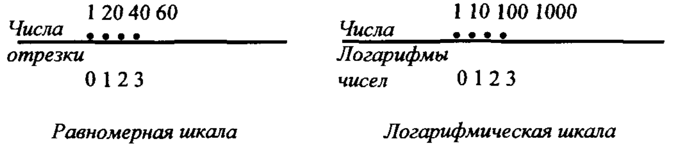


Рис. 3.8. **Прямолинейные шкалы**

Криволинейные шкалы имеются в секторных диаграммах – круг делится на 360°.

Масштабные знаки

Это эталоны величин, изображаемых на графике в виде отдельных графических знаков (квадратов, кругов, рисунков, силуэтов). Пользуются ими путем сравнения графических знаков со знаком-эталоном.

Допускается разрыв масштабной шкалы. Этот прием используется для изображения статистических данных, имеющих значения лишь в определенных пределах.

Экспликация графика

Экспликация графика – это пояснение его содержания, включает в себя заголовок графика, объяснения масштабных шкал, пояснения отдельных элементов графического образа.

Заголовок графика в краткой и четкой форме поясняет основное содержание изображаемых данных.

Помимо заголовка на графике дается текст, делающий возможным чтение графика. Цифровые обозначения шкалы дополняются указанием единиц измерения.

Заголовок графика должен кратко и точно раскрывать содержание графика, давать ответ на три вопроса – что, где, когда? Пояснительные надписи к отдельным элементам графика могут быть помещены на поле графика. В тех случаях когда пояснения длинны, а места на поле графика мало, следует выносить пояснения в виде ключа за пределы поля графика. Если на одном графике нанесено несколько линий, нужно указать название каждой линии в форме условного обозначения за пределами поля графика либо рядом с линиями. Все надписи на графике рекомендуется выполнять горизонтально. Не следует использовать для закраски графиков слишком яркие и пестрые цвета. По возможности цвета графика должны соответствовать цвету изображаемых предметов.

Общее правило построения графиков такое, что чем меньше на графике цифр, надписей и условных обозначений, тем он лучше воспринимается. Однако не следует забывать, что даже самый хороший график не заменит статистических данных. Поэтому если данные не приводятся в тексте, то их следует привести на графике.

Построение графика – всегда творческий процесс. Графики не следует оформлять сразу набело, необходим некоторый поиск. Лишь после составления и сравнения нескольких черновых вариантов можно наметить правильную композицию графика, установить масштабы и расположение знаков на поле графика.

В настоящее время разработаны пакеты прикладных программ компьютерной графики, которые облегчают задачу исследователя в практическом применении графиков. Наиболее распространенными пакетами прикладных программ являются: «Harvard graphics», «Statgraf», «Supercalc», «Excel».

Классификация статистических графиков

Для графического изображения статистических данных используются самые разнообразные виды графиков. Классификация графиков представлена на рис. 3.9.

На этой схеме графики подразделяются по двум признакам классификации: по способу построения и по цели использования.

Несмотря на многообразие видов графических изображений, при их построении выполняются общие правила. Так, **во-первых,** в соответствие с целью использования выбирается графический образ, т.е. вид графического изображения. **Во-вторых**, определяется поле графика – то пространство, в котором размещаются геометрические знаки.

В-третьих, задаются масштабные ориентиры с помощью масштабных шкал (равномерных или неравномерных). **В-четвертых**, выбирается система координат, необходимая для размещения геометрических знаков в поле графика. Наиболее распространенной системой координат при построении статистических графиков является система прямоугольных координат:

При этом наилучшее соотношение масштаба по осям абсцисс и ординат, равное 1,62:1, называется «Золотым сечением».

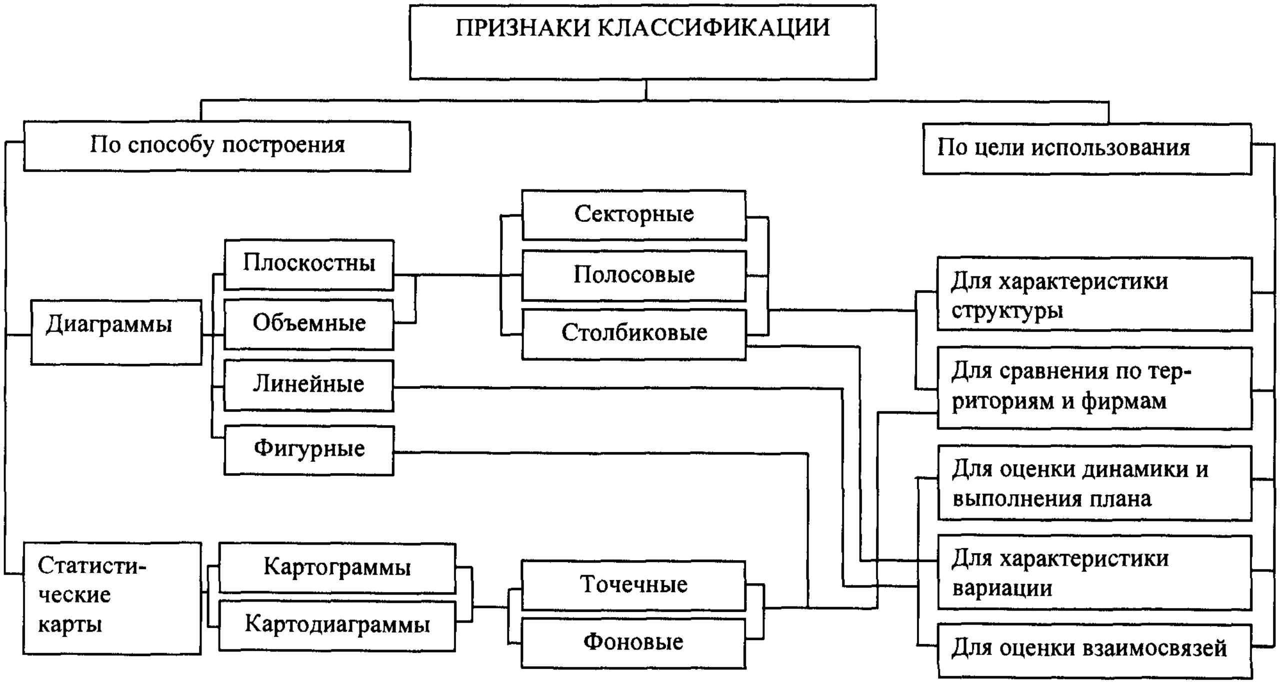
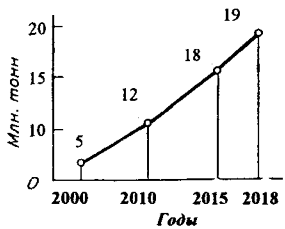


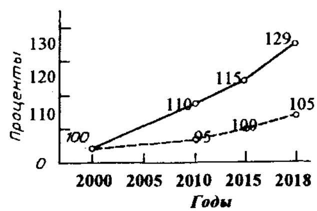
Рис. 3.9. **Классификация графиков**

Рассмотрение начнем с наиболее простых видов графиков

Линейные графики Линейные графики являются наиболее распространенными видами графических изображений при характеристике изменений явлении во времени, при изучении связей между явлениями.

Они незаменимы в тех случаях, когда на одном графике нужно показать динамику нескольких явлений.





——— темпы роста промышленной продукции

- - - - - темпы роста с/х продукции

Рис. 3.10. **Государственные закупки молока в регионе**

Рис. 3.11. **Темпы роста промышленной и сельскохозяйственной продукции** (в % к 2000 г.)

При построении линейной диаграммы динамики на оси абсцисс откладывают периоды времени или даты, а на оси ординат – величину изображаемого явления. Для начала отсчета на диаграмме должна быть обозначена нулевая линия. После этого на координатное поле наносят точки, соответствующие показателям изучаемого явления на определенные даты. Все точки соединяют и получают линию, характеризующую изменение исследуемого явления во времени.

Рассмотрим построение линейной диаграммы на основе данных о темпах роста промышленной и сельскохозяйственной продукции Республики Узбекистан за 2000–2018 годы (рис. 3.11).

На оси абсцисс откладываем четыре точки с учетом разной продолжительности периодов времени между приведенными годами. По оси ординат принимаем масштаб 1 см соответствует т. молока. С учетом максимального значения показателя наносим числа – масштаб на оси ординат. Из точек на оси абсцисс восстанавливаем перпендикуляры, высота которых пропорциональна государственным закупкам и принятому масштабу по оси ординат. Вершины перпендикуляров соединим отрезками прямых линий и получим кривую линию, характеризующую динамику государственной закупки молока (рис. 3.10).

Нередко на одном линейном графике приводится несколько кривых, которые дают сравнительную характеристику динамики различных показателей или одного и того же показателя в разных странах. В таких графиках линии всех показателей (или стран) расходятся из одной точки, принятой за 100%. На оси ординат дается разрыв масштабной шкалы для значений от 0 до 100%, в противном случае поле графика было бы заполнено неравномерно и график был бы невыразителен (рис. 3.11).

Столбиковые (ленточные) диаграммы

Столбиковые диаграммы

Столбиковыми диаграммами называются графические изображения статистических данных в виде столбиков-прямоугольников. Эти диаграммы широко используются для наглядного сравнения объемов изучаемых явлений во времени или пространстве, а также для изображения структуры явлений.

Технику построения столбиковой диаграммы покажем на примере данных о производстве электроэнергии в РУз: 2005 г. – 48; 2000 г. – 52; 2015 г. – 57; 2018 г. – 65 млрд кВт-ч. Для построения диаграммы берем систему, прямоугольных координат. На оси абсцисс на одинаковом расстоянии друг от друга наметим четыре отрезка равной длины – основания для столбиков. Высота столбиков определяется в соответствии с принятым масштабом по оси ординат и значениями показателей. Учитывая размер поля графика и максимальное значение показателя, установим такой масштаб, что каждым 200 млрд. кВт-ч электроэнергии соответствует отрезок по оси ординат в 1 см. Тогда высота столбиков составит для 2005 г. – 4,8 мм (48:10×10); для 2010 г. –5,2 мм (52:10×10); для 2015 г. – 5,7 мм и для 2018 г. – 6,5 мм.

Диаграмма имеет следующий вид (рис. 3.12).

Для наглядности столбики следует заштриховать, закрасить или нанести на них схематические рисунки изображаемых предметов.

Столбиковые диаграммы целесообразно применять для сравнения нескольких показателей. На рис. 3.13 посредством диаграммы показано изменение потребления основных продуктов питания населением. На этой диаграмме столбики располагаются вплотную. Масштаб принят такой, что каждым 40 кг соответствует отрезок по оси ординат в 1 см. На диаграмме сделаны необходимые надписи и условные обозначения.

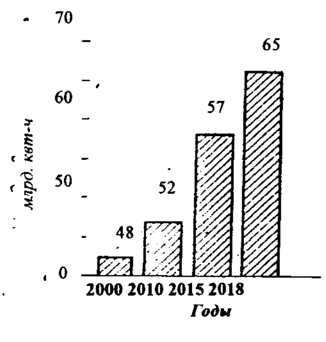


Рис. 3.12. **Производство электроэнергии в РУз**

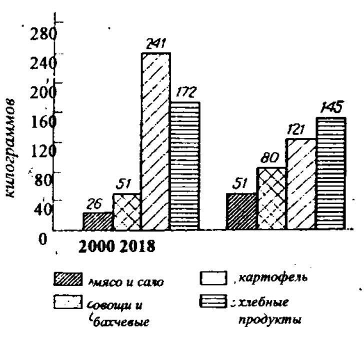


Рис. 3.13. **Потребление основных продуктов питания населением региона** (на душу населения в год, килограммов)

Диаграмма наглядно показывает не только изменение уровня потребления отдельных продуктов на душу населения, но и улучшение структуры питания населения. В рационе питания все большее место занимают мясные продукты и овощи. В 2018 г. по сравнению с 2000 г. потребление мяса возросло почти в два раза, овощей – в 1,6 раза. За этот же период потребление картофеля и хлебных продуктов значительно снизилось.

Если основания столбиков разместить по оси ординат, а значения уровней по оси абсцисс, то получим ленточные (полосовые) диаграммы.

Столбиковые диаграммы могут использоваться для характеристики структуры совокупности. Состав совокупности лучше воспринимается не в абсолютных, а в относительных величинах структуры. При таких данных все столбики в диаграмме имеют одинаковую высоту и соответствуют 100%. Каждый столбик разбивается на части пропорционально удельному весу отдельных частей во всей совокупности.

Секторные диаграммы

Секторные диаграммы представляют круг, разделенный на секторы. Применяются они главным образом для изображения структуры. Площадь всего круга принимается за 100%. Площадь каждого сектора характеризует часть целого и соответствует удельному весу этой части в целом.

Построим секторную диаграмму, характеризующую состав населения РУз за два года. Удельный вес городского населения в общей численности населения РУз составил в 1940 г. 25%, а в 2018 г. – 51%. Доля сельского населения равна, соответственно, 75 и 49% (рис 3.14).

Возьмем два одинаковых круга. Площадь каждого круга нужно разделить на два сектора. Известно, что площади секторов пропорциональны их центральным углам. Следовательно, для определения площади секторов нужно 360° распределить пропорционально величинам удельных весов. Поскольку 3,6° соответствует 1%, то центральные углы составят:

	городское население	**сельское население**
1940 г.	$3{,}6° \times 25 = 90°$;	$3{,}6°\times 75 = 270$;
2018 г.	$3{,}6°\times 51 = 184°$.	$3{,}6°\times 49 = 176°$.

На основании полученных данных при помощи транспортира делим каждый круг на сектора и получаем следующую диаграмму (рис. 3.15).

Секторные диаграммы выразительны в тех случаях, когда совокупность делится не более чем на 4-5 частей и наблюдаются значительные структурные сдвиги. Если совокупность делится на большее число частей и структурные изменения незначительны, то для изображения структуры целесообразнее применять ленточные диаграммы.

Рис. 3.14. **Удельный вес городского и сельского населения в общей численности населения, в процентах**

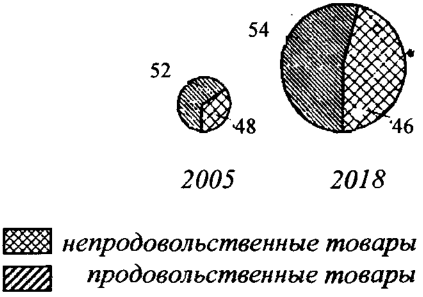


Рис. 3.15. **Розничный товарооборот государственной и кооперативной торговли**

Секторные диаграммы могут одновременно характеризовать изменение структуры и объема изображаемых явлений. Образец такой диаграммы представлен на рис. 3.15. Диаграмма показывает рост общего объема товарооборота в 2018 г. против 2005 г. в 19 раза – во столько раз площадь большого круга больше площади малого круга. Изменение соотношения площадей секторов показывает увеличение доли непродовольственных товаров во всем товарообороте с 48 до 54% при одновременном снижении доли продовольственных товаров с 52 до 48%.

Графики выполнения планового задания

Графический метод находит широкое применение в практике текущего контроля за выполнением плановых заданий. Применяемые формы графических изображений для сравнения плановых и фактических показателей весьма разнообразны. Здесь рассмотрены два вида графиков.

Для систематического контроля за выполнением планов одновременно по нескольким объектам используется контрольно-плановый график. Построение его покажем на основе приведенных в табл. 3.7. данных.

Контрольно-плановый график строится на сетке и имеет вид таблицы, в левой части которой записываются объекты (цехи, отделы, заводы), а в правой – периоды времени (дни, пятидневки, декады). На графике поле каждого периода соответствует 100% и делится на пять равных вертикальных полос. Каждая полоса соответствует 20% планового задания на этот период времени. Выполнение плана изображается прочерчиванием горизонтальной линией вертикальных делений. На графике проводятся две лилии.

Таблица 3.7

Выполнение плана товарооборота по молочному отделу магазина «Корзина» за октябрь

Декады	Товарооборот		Фактический товарооборот с начала месяца	Процент выполнения планового задания	
	плановое задание	фактически		за каждую декаду гр.2/гр.1×100%	нарастающим итогом с начала месяца гр.3/итог гр.1× ×100%
А	1	2	3	4	5
I	2 000	1 800	1 800	90	28,2
II	2 000	2 000	3 800	100	59,5
III	2 400	3240	7040	135	110,0
Итого	6 040	7 040	–	110	–

Тонкая линия дает характеристику процента выполнения плана за отдельный период, при этом она не должна выходить за пределы отведенных для *этого* периода границ. Двойная линия показывает процент выполнения месячного плана за истекшее время с начала месяца нарастающим итогом (рис. 3.16).

Объект	Декады														
	I					II					III				
	Проценты														
	20	40	60	80	100	20	40	60	80	100	20	40	60	80	100
Молочный отдел															

и т.д.

Рис. 3.16. **Выполнение плана товарооборота по магазину «Корзина» за октябрь 2018 г.**

За I декаду план выполнен на 90%. Горизонтальной тонкой линией прочеркиваются четыре узкие полосы полностью (соответствует 80%) и половина пятой полосы (соответствует 10%). За III декаду перевыполнение плана показано прочеркиванием всех пяти полосок (соответствует 100%) и проведением снизу дополнительной тонкой линии, длина которой соответствует 35%. Двойной линией внизу показано выполнение месячного плана товарооборота нарастающими итогами с начала месяца, при этом каждая декада отмечена штрихом.

Фигурные диаграммы

В фигурных диаграммах в качестве графического образа используются изображения самих предметов. Фигурные диаграммы можно строить фигурами различных размеров или различной численностью фигур одинакового размера.

Предположим, что в одной из областей было продано населению телевизоров в 2005 г. – 15,0; в 2010 г. – 64,5; в 2018 г. –90,1 тыс. шт. Построим на основе этих данных фигурные графики в двух видах.

При построении графика путем различной численности фигур одинакового размера, прежде всего, необходимо установить масштабный знак с таким расчетом, чтобы не получилось фигур слишком много, но в то же время и не слишком мало. И в том и в другом случае график будет невыразителен. В нашем примере можно установить такой масштаб: каждой фигуре соответствует продажа 15 тыс. телевизоров. Разделим приведенные показатели на 15 и получим число фигур для 2005 г. – 1; для . 2010 г. – 5,0; для 2018 г. – 7,0. Построим диаграмму, рис. 3.17.

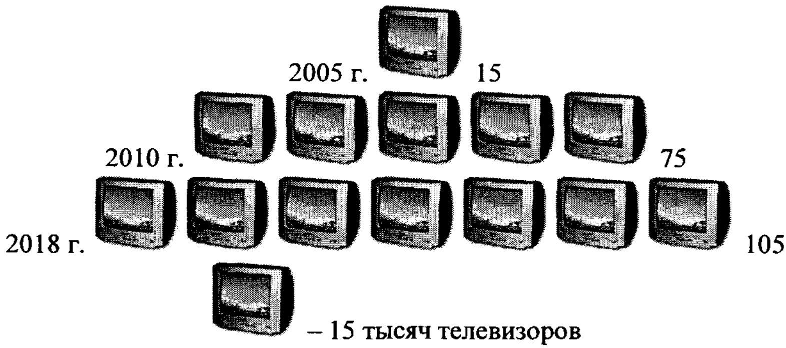


Рис. 3.17. **Продажа телевизоров населению по области** (тыс. шт)

Построение фигурного графика путем фигур различных размеров представлено на рис. 3.18. Размер фигуры за 2018 г. в шесть раз больше фигуры за 2005 г. Однако на глаз эти количественные соотношения воспринимаются трудно. Поэтому на графике приведены цифры.

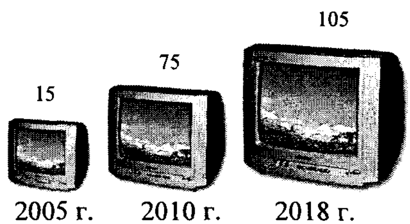


Рис. 3.18. **Продажа телевизоров населению по области** (тыс. шт)

Фигурные диаграммы используются главным образом, для популяризации статистических данных. Для аналитических целей их применение ограничено по причине их недостаточной точности.

При построении фигурных диаграмм следует иметь в виду, что на одном графике может быть отражена только одна тема. Не следует применять сложные рисунки, слишком яркие краски. Фигуры-символы должны быть лаконичными, понятными сами по себе и не должны требовать дополнительных объяснений.

Картограммы и картодиаграммы

Картограммы и картодиаграммы изображают распределение тех или иных явлений по территории. *Картограммой* называется изображение территориального размещения изучаемых совокупностей на географической карте при *помощи условных обозначений.* Например, с помощью точек изображают размещение посевных площадей сельскохозяйственных культур на территории области. В этом случае получают точечную картограмму. В области имеется 69 тыс. гектаров посевов картофеля. Условно принимают, что одна точка равна 1 тыс. гектаров. Точки наносят на контурную карту области в местах фактического размещения посевов картофеля (рис. 3.19).

Рис. 3.19. **Размещение посевов картофеля по территории области**

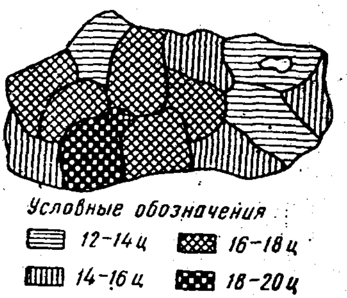


Рис. 3.20. **Урожайность зерновых в районах области**

Территориальное распределение статистических величин изображается *фоновой картограммой.*

Различные районы области группируются по величине показателя, и каждая группа заштриховывается или закрашивается особым способом. Обычно, чем выше значение статистической величины, тем гуще штриховка или темнее окраска.

Так построены известные в географии картограммы плотности населения. Таким же образом можно изобразить территориальные различия в урожайности в разных районах области (рис. 3.20).

Картодиаграммами называются графические изображения, нанесенные на географическую карту. С их помощью изображаются, например, структура посевных площадей в разных районах или областях, динамика валового сбора или динамика численности тракторов.

Диаграмма размещается на карте посреди той территории, к которой относятся изображаемые ею статистические величины.

Таблица 3.8

Распределение фермерских хозяйств области по числу отделений области

Число отделений	Число хозяйств
1	13
2	22
3	37
4	26
5	21
6	14
7	9
8	2
Всего	144

Рис. 3.21. **Распределение фермерских хозяйств области по количеству отделений**

Графическое изображение вариационного ряда осуществляется в прямоугольной системе координат. На оси абсцисс откладывают варианты признака, а на оси ординат – частоты. Ордината может быть изображена в виде столбика (этот способ применяется преимущественно для изображения интервальных вариационных рядов) или в, виде точек (для дискретных рядов). Построим график по данным табл. 3.8 (рис. 3.21).

Интервальный вариационный ряд графически построим по данным табл. 3.8.

В отличие от графического изображения динамического ряда промежутков между столбиками на подобных графиках оставлять не следует: непрерывность изображения соответствует непрерывному характеру вариации признака. Такой график называется *гистограммой* (рис. 3.22, табл. 3.9).

Графическое изображение взаимосвязи

Метод графического изображения играет важную роль при анализе взаимосвязи между признаками изучаемой совокупности, если эта связь имеет характер корреляции (см. главу 9). Имея данные о двух связанных между собой признаках по совокупности, можно построить график в системе прямоугольных координат. Величина факторного признака откладывается на оси абсцисс, а результативного признака – на оси ординат. Каждая пара значений признаков, свойственная отдельной единице совокупности, показывается точкой с данной абсциссой и ординатой. Общее расположение точек дает представление о характере зависимости. На рис. 3.23 графически представлена зависимость урожайности картофеля от количества, внесенных удобрений на гектар посевов.

Расположение точек на данном графике свидетельствует о наличии, прямолинейной зависимости урожайности от количества внесенных удобрений. Направление этой зависимости дано пунктирной линией, которая показывает, что увеличение значений одного признака (количества внесенных удобрений) влечет за собой увеличение значений другого признака (урожайности). На следующем графике (рис. 3.24) показана криволинейная обратная зависимость, при которой значения одного признака (себестоимости продукции) уменьшаются по мере увеличения другого признака (урожайности).

Таблица 3.9

Распределение фермерских хозяйств по урожайности озимых культур

Урожайность (ц с 1 га) (интервалы признака)	Число хозяйств (частота)
5-8	5
8-11	26
11–14	38
14-17	46
17-20	24
20-23	14
23-26	5
26-29	3
Итого	161

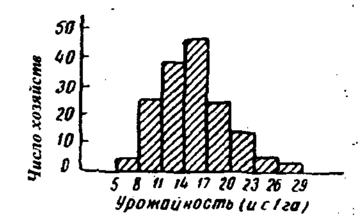


Рис. 3.22. **Распределение фермерских хозяйств по урожайности озимых зерновых**

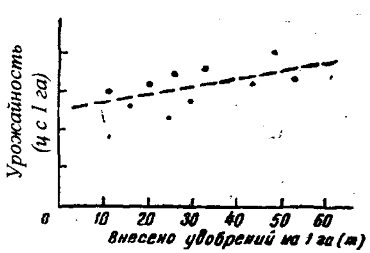


Рис. 3.23. **Изменение урожайности картофеля в зависимости от количества внесенных органических удобрений**

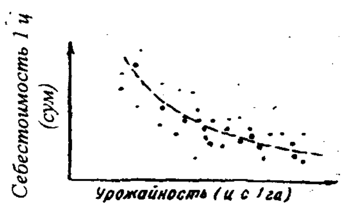


Рис. 3.24. **Изменение себестоимости картофеля в зависимости от его урожайности**

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

Диаграмма представляет собой прямоугольные фигуры для графического изображения *3 показателей, один из которых является произведением 2 других* (численность населения является произведением плотности населения на его территорию).

В каждом таком прямоугольнике основание пропорционально одному из показателей-сомножителей, а высота его соответствует 2-му показателю – сомножителю. Площадь прямоугольника равна величине 3-го показателя, являющегося произведением 2 первых. Располагая рядом нескольких прямоугольников, относящихся к разным показателям, можно сравнивать не только размеры показателя-произведения, но и значения показателей-сомножителей.

Пример

Валовой сбор с/х культуры равен произведению урожайности и посевной площади (рис. 3.25).

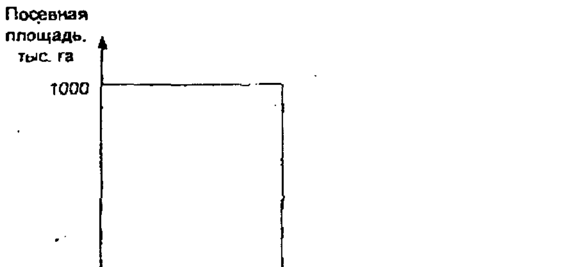


Рис. 3.25. **Зависимость валового сбора плодов и ягод от урожайности и отведенной посевной площади**

<u>На этом графике можно сравнить между собой:</u>

✓ урожайность (по длине основания);

✓ посевные площади (по длине боковой стороны);

✓ валовой сбор (по площади прямоугольника).

Для одновременного изображения на диаграмме динамики разных признаков (например, валового сбора или урожайности культур) можно применить два разных масштаба по ординате и две шкалы: справа и слева. Уровни валового сбора (объемный признак) лучше изобразить столбиками, а уровни урожайности (относительная величина) – точками (табл. 3.10, рис. 3.26).

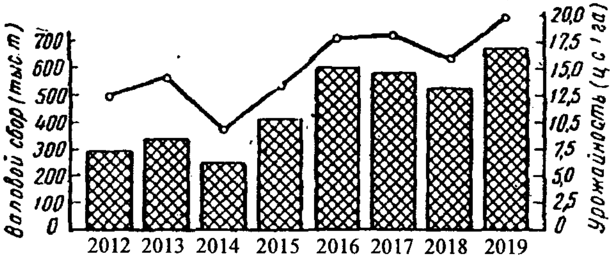


Рис. 3.26. **Валовой сбор и урожайность пшеницы фермерских хозяйств области на неорошаемых землях**

Таблица 3.10

Валовой сбор и урожайность пшеницы в фермерских хозяйствах области

Год	Валовой сбор	Урожайность (ц с 1 га)
2012	286	12,8
2013	345	14,5
2014	260	9,3
2015	418	13,5
2016	607	18,1
2017	574	18,6
2018	513	16,3
2019	685	19,9

При построении графика выбираем масштаб так, чтобы ширина графика в 1,5-2 раза превосходила его высоту: на оси абсцисс 1 год = 7,7 мм, на оси ординат *а* мм -100 тыс. тонн валового сбора пшеницы и 3 мм = 2,5 ц с 1 га урожайности.

На одном графике можно изобразить динамику 4-5 показателей, но не более, иначе график утратит наглядность.

Кроме рассмотренных видов диаграмм, картограмм и картодиаграмм на практике встречаются и другие, более сложные графические изображения статистических данных.

Интеллектуальный тренинг

1. В чем заключается содержание сводки статистических материалов?

2. «Один и тот же материал дает диаметрально противоположные выводы при различных приемах группировки». Кому принадлежат эти слова?

3. Вопрос о группировках статистического материала «...вовсе не является таким узкотехническим, узкоспециальным вопросом, каким он может показаться на первый взгляд». Кому эти слова принадлежат?

4. В чем состоит значение метода группировок в анализе статистических данных?

5. Какие основные задачи решаются исследователем с помощью метода группировок?

6. Какова роль и значение классификаций? Приведите примеры важнейших классификаций.

7. Какие группировки называют комбинационными? Приведите пример.

8. В чем состоит отличие комбинационной и многомерной группировки?

9. Какие основные проблемы подлежат решению при группировке статистических данных?

10. Чтобы выделить хозяйства разных типов, необходима «... во-первых, группировка хозяйств не по одному признаку (величина площади), а по нескольким признакам (количество машин, скота, площади под специальными культурами и пр.), а, во-вторых, комбинирование различных группировок, т.е. разделение каждой группы, напр., по величине площади, на подгруппы по количеству скота и т.д.» Кому относятся эти слова?

11. Что называется группировочным признаком?

12. Приведите пример группировки по атрибутивному признаку.

13. Как обозначаются границы интервалов при группировке по количесгвенному признаку?

14. Как обозначаются границы интервалов при группировке по непрерывно варьирующему количественному признаку?

15. Что называется рядом распределения, из каких элементов он состоит?

16. Что такое статистическая таблица? Какова их функция?

17. Как различаются статистические таблицы по подлежащему?

18. Назовите способы построения сказуемого таблицы.

19. Какие основные требования предъявляются к построению и оформлению статистической таблицы?

20. В чем заключается назначение статистических графиков?

21. Каковы основные элементы графика?

22. Перечислите основные виды статистических графиков.

23. Каково назначение и правила построения столбиковых диаграмм?

24. Каковы правила построения круговых и квадратных диаграмм?

25. Для каких целей строятся секторные диаграммы?

26. Назначение и правила построения линейных графиков (статистических кривых).

27. Что такое полигон, гистограмма? Для чего они применяются и каковы основные правила их построения?

Список использованной и рекомендуемой специальной литературы

1. Грачев Н.Г. Статистические группировки. М.: Госстатиздат, 1951.

2. Грачев Н.Г. Статистические группировки в изучении экономики. М.: Госстатиздат, 1958.

3. Герчук Я.П. Графические методы в статистике. М.: Статистика, 1968.

4. Ауэрбах Ф. Графические изображения М.-Л.: Госиздат, 1928.

5. Бринтон В.К. Графическое изображение фактов. М.: Экономическая жизнь, 1927.

6. Коротков В.А. Табличный метод в социально-экономической статистике. М.: МЭСИ, 1986.

7. Лившиц Д.Ф. Статистические таблицы. М.: Гостатиздат, 1968.

8. Плошко Б.Г. Группировка и системы статистических показателей. М.:Статистика, 1971.

9. Суслов И.П. Теория статистических показателей. М.: Статистика, 1975.

10.Смит М.Н. Группировка статистических данных. М.: Соцэгиз, 1931.

Глава IV
АБСОЛЮТНЫЕ И ОТНОСИТЕЛЬНЫЕ ВЕЛИЧИНЫ

Дорожная карта

4.1. Определение абсолютных величин

Результаты статистического наблюдения и сводки его материалов выражаются, прежде всего, в ***абсолютных (итоговых) величинах*** (показателях).

В зависимости от характера решаемых задач все статистические показатели классифицируются по следующим основным критериям.

✓ ***по форме выражения:***

• абсолютные;

• относительные;

• средние;

✓ ***<u>по охвату единиц совокупности:</u>***

• индивидуальные – характеризуют отдельную единицу совокупности;

• сводные – характеризуют группу единиц совокупности или всю совокупность в целом;

✓ **<u>по объектной и временной определенности</u>**.

• конкретные – характеризующие исследуемую величину в данном месте и в данное время;

• абстрактные – применяемые на этапах планирования.

Исходной, первичной формой выражения статистических показателей являются абсолютные величины, т.е. абсолютный показатель – это показатель в форме абсолютных величин.

Абсолютные величины показывают размеры (уровни, объемы); общественных явлений в данных условиях места и времени. Например, характеристику деятельности малых предприятий дают следующие абсолютные величины:

➤ объем производства продукции и услуг в стоимостном выражении;

➤ общая численность занятых;

➤ стоимость основных фондов;

➤ стоимость привлеченных инвестиций;

➤ объем экспорта и импорта товаров и др.

Или же характеристику выживаемости малых фирм можно получить только при наличии следующих абсолютных величин:

❖ сколько создано за год разных субъектов малого бизнеса;

❖ сколько из них прошли государственную регистрацию;

❖ сколько из них действующих и бездействующих;

❖ сколько за год ликвидировано субъектов малого бизнеса;

❖ сколько ими создано за год новых рабочих мест;

❖ средняя продолжительность живучести предприятий в годах и др.

Абсолютные величины необходимы для количественной характеристики общественных явлений, составления плановых заданий и проверки их выполнения, прогнозирования общественных явлений. Они дают представление о наличии тех или иных ресурсов – материальных, денежных, трудовых, о размерах производства различных видов продукции и т.п. Они важны также и как исходные данные для вычисления ряда других статистических показателей. Так, например, путем арифметических действий над абсолютными числами исчисляют относительные и средние величины, показателей рядов динамики, индексы и др.

Абсолютные величины имеют свою особенность. Они выражают размеры общественных явлений только в определенных единицах измерения и поэтому всегда представляют собой не отвлеченные (коэффициенты, проценты), а именованные числа. Например, численность постоянного населения Узбекистана в 2017 г. превысило 32,6 млн чел.

4.2. Виды абсолютных величин

По охвату единиц совокупности абсолютные величины подразделяются на индивидуальные и суммарные.

Индивидуальными называются абсолютные величины, выражающие размеры количественных признаков у отдельных единиц той или иной совокупности объектов (например, величина заработной платы отдельного рабочего, размер посевной площади пшеницы в отдельном хозяйстве и т.п.). Они получаются непосредственно в процессе статистического наблюдения и фиксируются в первичных учетных документах. Индивидуальные абсолютные величины служат основой любого статистического исследования.

В отличие от индивидуальных **суммарные абсолютные величины** характеризуют итоговую величину признака по определенной совокупности объектов, охваченных статистическим наблюдением. Они получаются либо путем прямого подсчета числа единиц наблюдения, либо в результате суммирования значений признака у отдельных единиц совокупности. Так, в результате подсчета зарегистрированных в процессе переписи населения лиц получаются итоговые, абсолютные данные о численности населения по отдельным областям и в целом по РУз; о численности мужчин и женщин и т.п.

В результате сводки годовых отчетов фермерских хозяйств получаются абсолютные итоговые данные о количестве фермерских хозяйств, о количестве полученной продукции и другие суммарные показатели.

Например, к суммарным абсолютным величинам относится и упоминавшийся в начале главы показатель численности населения страны на определенную дату. Численность населения нашей страны на 1 января 2018 г. составляла 32,6 млн человек, из них проживало в городах 16,5 млн человек, в сельской местности – 16,1 млн человек.

В ряде случаев суммарные абсолютные величины получаются не в результате сводки данных статистического наблюдения, а путем специальных расчетов.

При помощи таких расчетов определяется, например, перспективная численность населения, количество сельскохозяйственной продукции произведенной в подсобных хозяйствах населения отдельных областей (так, валовой сбор хлопка исчисляется путем умножения посевной площади хлопчатника на среднюю его урожайность с гектара и т.п.).

Практическая потребность в подобных расчетах обычно обусловливается отсутствием соответствующих данных в первичных учетно-статистических материалах.

4.3. Единицы измерения абсолютных величин

Абсолютные статистические величины всегда являются именованными числами, т.е. выражаются в определенных ***единицах измерения*** *(рис. 4.1)*.

Натуральными принято называть единицы измерения, выражающие величины, предметы, вещей и т.п. в физических мерах, т.е. в мерах длины, площади, объема, веса и т.п.

Рис. 4.1. **Единицы измерения абсолютных величин**

Натуральными единицами измерения пользуются для характеристики объема производства той или иной продукции, продажи товаров, потребления населением различных продуктов питания и т.п., например, производство чугуна и стали измеряют в тоннах, производство тканей – в погонных и квадратных метрах, производство газа – в кубических метрах, а электроэнергии – в киловатт-часах. Примером этому могут служить следующие фактические данные о производстве отдельных видов продукции в топливно-энергетических отраслях промышленности и сельского хозяйства Республики Узбекистан за 2005-2018 гг. (табл. 4.1).

Таблица 4.1

Производство промышленной и сельскохозяйственной продукции в натуральном измерении

	Виды продукции	**Ед. измерения**	**2005**	**2010**	**2014**	**2017**
1	Электроэнергия	Млн кВт.ч.	47665,1	51976,3	55776,0	62300,0
2	Нефть	Тыс. тонн	3448,7	2017,9	1031,3	...
3	Уголь	Тыс. тонн	3002,0	3629,4	4396,8	...
4	Хлопчатобумажные ткани	Млн м2	258,3	145,3	169,4	401,4
5	Обувь	Млн пар	2,3	3,3	8,7	...
6	Яйца	Млн штук	1966,7	3061,2	4950,0	6332,7

***Источник**: Статистический ежегодник Госкомстата РУз. Т., 2015. С. 207, 218, 2016.*

Данные о количестве различных продуктов, выраженные в натуральных единицах измерения, не допускают суммирования. Поэтому в статистике с целью получения общего итога близких по своему потребительскому назначению продуктов иногда используются **условно-натуральные единицы** измерения. Для этого, прежде всего, находятся так называемые переводные коэффициенты, выражающие соотношения между натуральными единицами измерения различных продуктов по какому-либо признаку (потребительским свойствам, трудоемкости, себестоимости и т.д.). Затем по найденным коэффициентам эти продукты пересчитываются на один продукт, натуральная единица которого принята в качестве условной единицы измерения.

Пример. За отчетный период предприятие выработало следующее количество мыла и моющих средств по видам (табл. 4.2). Требуется определить общее количество выработанной предприятием продукции в условно-натуральных единицах измерения. За условную единицу измерения принимается мыло 40% жирности кислот.

Таблица 4.2

Количество мыла и моющих средств, выработанных за отчетный год

	Виды мыла и моющих средств	**% жирных кислот**	**Количество, кг**	**Коэф. перевода (40%=1)**	**Кол-во продукции в условном исчислении**
1	2	3	4	5	6=5-4
1	Мыло хозяйственное	72	1000	1,8	1800
2	Мыло хозяйственное (специальное)	60	500	1,5	750
3	Мыло хозяйственное	40	250	1,0	250
4	Мыло туалетное	80	1500	2,0	3000
5	Стиральный порошок	10	2500	0,25	625
	Итого	х	х	х	6425

Решение

Чтобы определить общее количество продукции, выработанной предприятием, необходимо исчислить коэффициенты перевода. Если условной единицей является мыло 40% жирности, то оно принимается равным единице. Тогда коэффициенты перевода в условное мыло хозяйственное 72% жирности: 72/40 = 18; мыло хозяйственное (специальное) 60% жирности: 72/40 = 1,5; мыло туалетное 80% жирности; 80/40 = **2,0**; стиральный порошок **10%** жирности: = 10/40 = 0,25 (графа 5).

Далее определим общий объем производства мыла и моющих средств по видам (учитывая перевод в условные единицы в 40% исчислении). Таким образом, общий объем производства мыла и моющих средств в 40% исчислении составил 6425 кг (графа 6).

Производят также пересчет разных видов топлива – уголь, нефть, газ, торф и др. – в условное (7000 калорий) топливо. Это делается для того, чтобы привести к соизмеримому виду непосредственно несоизмеримые в силу качественных или количественных различий величины и определить их суммарную величину.

Условно-натуральные единицы измерения имеют довольно ограниченное применение в практике и совершенно непригодны для исчисления общего объема разнородной продукции.

Широкое применение в статистике имеют *денежные единицы измерения (сум, доллар и др.).* Они используются для характеристики в стоимостном (денежном) выражении объема производства ВВП, национального богатства, величины заработной платы рабочих, валовой доход населения и во многих других случаях (табл. 4.3).

Иногда применяются *комбинированные* единицы измерения. Это в тех случаях, когда учет в одной из возможных единиц измерения не дает достаточного представления о явлении, оно учитывается в двух единицах измерения. Так, например, недостаточно учесть количество выпущенных автомобилей в штуках, потому что они имеют различную мощность. Чтобы получить правильное представление об их производстве, надо использовать две единицы измерения в штуках и по мощности. Кожа учитывается в квадратных дециметрах и весовых единицах, стекло – в квадратных метрах и по весу, отработанное время – в человеко-днях и человеко-часах, потребление электроэнергии – в киловатт часах (см. табл. 4.1), грузооборот измеряется в тонно-километрах и т.п.

Трудовые единицы измерения – человеко-часы, человеко-дни и т.п. – применяются для определения затрат труда на производство продукции, на выполнение какой-либо работы, для характеристики трудовых ресурсов и их использования.

Таблица 4.3

Основные стоимостные показатели Республики Узбекистан за 2008-2017 гг. (трлн. сум)

№	Показатели	2005	2010	2014	2017
1	Валовой внутренний продукт (ВВП)	15,9	62,9	145,8	254,0
2	Основные фонды в экономике (на конец года)	25,7	85,9	179,8	...
3	Продукция промышленности	11,0	34,5	75,2	148,8
4	Продукция сельского хозяйства	6,0	16,8	36,9	69,5
5	Инвестиции в основной капитал	3,2	15,3	22,1	27,9
6	Иностранные инвестиции и кредиты	0,7	4,3	7,0	...
7	Строительные работы	1,4	8,2	20,0	...
8	Розничный товарооборот	5,6	21,9	58,1	105,2
9	Платные услуги населению	1,6	7,9	22,4	118,8
10	Внешнеторговый оборот	9,5	22,2	28,1	...
	В т.ч.				
	Экспорт	5,4	13,0	14,1	...
	Импорт	4,1	9,2	14,0	...

***Источник**: Статистический ежегодник Госкомстата РУз. Т., 2015. С. 28, 30.*

В ряде случаев в качестве единиц измерения в статистике используются единицы счета времени. Так определяется, например, срок службы сооружения или какого-либо изделия, например, продолжительность горения электрической лампочки в часах, средняя продолжительность жизни людей, измеряемая в годах, и др.

4.4. Определение относительных величин

Несмотря на большое самостоятельное значение в системе показателей, абсолютные величины сами по себе не всегда пригодны для глубокой характеристики общественных явлений. Нередко, чтобы правильно оценить тот или иной абсолютный показатель, требуется сравнить его с другим абсолютным показателем, например, с таким же абсолютным показателем, установленным по плану или относящимся к другому периоду времени.

В отличие от абсолютных статистических показателей относительные величины представляют собой производные, вторичные. Они получаются не в результате простого суммирования, а путем относительного (кратного) сравнения между собой двух величин:

- **в числителе – сравниваемая величина, т.е. та, *которую* сравнивают;**
- **в знаменателе – база сравнения, т.е. величина, *с которой* сравнивают.**

Относительные показатели, полученные путем сравнения двух абсолютных величин, называются ***относительными показателями первого порядка***, а полученные при сопоставлении относительных показателей – ***показателями высших (второго, третьего) порядков.*** Показатели четвертого порядка, ввиду сложности их построения, почти не применяются.

Например, прибыль от продаж и выручка от продаж – это абсолютные показатели. Отношение прибыли от продаж к выручке от продаж – рентабельность продаж – будет относительным показателем первого порядка. Если мы поделим рентабельность от продаж отчетного периода на этот же показатель прошлого периода, то мы получим коэффициент роста рентабельности продаж – относительный показатель второго порядка. Если же мы поделим коэффициент роста рентабельности продаж данного предприятия на коэффициент роста рентабельности продаж другого предприятия, то мы получим коэффициент опережения – относительный показатель третьего порядка.

Почти все виды относительных величин ***не имеют размерности.*** В этом смысле отличительной особенностью относительных статистических величин является то, что они обычно в отвлеченной, форме выражают соотношения суммарных или индивидуальных абсолютных величин. Так, удельный вес городского населения представляет собой отношение численности городского населения определенной территориальной единицы к численности всего населения, включая городское и сельское.

К относительным величинам в статистике относятся также некоторые именованные числа, например, выработка продукции на душу населения, плотность населения (число жителей на квадратный километр) и др. Подобного рода относительные величины показывают, сколько единиц одной именованной суммарной величины приходится на единицу другой, связанной с ней, именованной суммарной величины.

Одной из специфических черт относительных величин, обусловленных их характером, является то, что они позволяют отвлечься от конкретных различий абсолютных величин. В силу этого свойства они дают возможность сравнивать такие явления, абсолютные размеры которых непосредственно не сопоставимы. Так, например, отношение, выраженное тем, что 10% фермеров имеют 30% посева, абстрагирует различные абсолютные цифры и потому годно для сравнения со всяким пробным отношением любой местности. Но для такого сравнения надо выделить в другой местности тоже 10% фермеров, не больше и не меньше.

Относительные величины имеют важное значение и широко применяются в экономических расчетах. Они значительно усиливают наглядность статистических данных, помогают лучше разобраться в их особенностях и закономерностях. Покажем это на следующем примере.

Допустим, что фермерское хозяйство «Чашми Сафед» продало государству мясо в отчетном году 2500 ц. при плановом задании 2300 ц., а соседнее с ним фермерское хозяйство «Истиклол» – 2820 ц. при плановом задании 2750 ц.

Эти данные показывают, что второе хозяйство продало мяса больше, чем первое. Однако из приведенных абсолютных величин не видно, какое из этих хозяйств лучше выполнило свои обязательства перед государством. Для этого сравним по каждому фермерскому хозяйству объем продажи мяса государству с планом.

<u>*ф/х* «Чашми Сафед»</u>

$$\frac{2500\ ц}{2300\ ц} \times 100 = 108{,}7\%$$

<u>*ф/х* «Истиклол»</u>

$$\frac{2820\ ц}{2750\ ц} \times 100 = 102{,}5\%$$

Из этих данных видно, что ф/х «Чашми Сафед» свои обязательства перед государством выполнило значительно лучше (выполнение планового задания на 108,7%, а ф/х «Истиклол» – на 102,5%).

Относительные величины можно исчислить не только путем сравнения абсолютных величин. Сравниваемые величины могут быть абсолютными, средними и относительными.

Приведём следующий пример. В 2017 г. валовой доход на душу населения составило в Навоийской области 9,1 млн сум, а в республике Каракалпакстан 4,1 млн сум. Сравнив эти данные, получим 2,2 (9,1:4,1). Эта относительная величина показывает, что в расчете на душу населения валовой доход в Навоийской области больше, чем в РК в 2,2 раза. При этом относительная величина (2,2 раза) получена в результате сравнения двух других относительных величин, характеризующих долю в общей численности населения двух регионов.

4.5. Формы выражения относительных величин

В зависимости от характера изучаемого явления и конкретных задач исследования относительные величины могут иметь различную форму (внешний вид) выражения. Они могут выражаться в следующих видах (рис. 4.2).

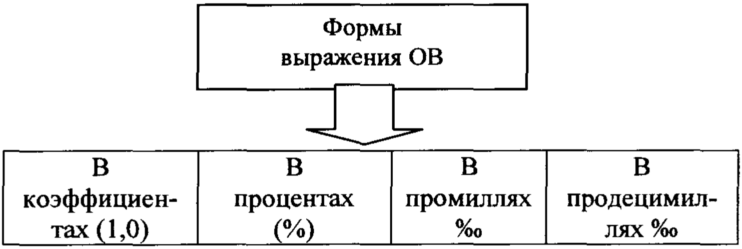


Рис. 4.2. **Формы выражения ОВ**

Относительные величины имеют форму ***коэффициентов,*** если они исчисляются простым делением сравниваемой величины на базу сравнения. При этом база сравнения условно понимается за единицу (одно целое). Коэффициент покажет, ***во сколько раз*** сравниваемая величина больше базы сравнения. Относительные показатели в форме коэффициентов удобно применять, когда сравниваемый показатель значительно больше чем базы сравнения.

Например, в 2000 г. численность занятых в экономике нашей страны составляла 3988 тыс. чел., а в 2017 г. – 13520 тыс. чел. Поделив 13520 на 8983, получим 1,51 – это коэффициент роста, показывающий, что численность занятых в экономике увеличилась за указанные годы в 1,51 раза.

Если базу сравнения принять за 100, то относительная величина будет выражена в ***процентах.*** Например, в нашем случае с занятыми в экономике если 1,51·100=151% т.е. получим, что в 2017 г. численность занятых в экономике нашей республики по сравнению с 2000 г. выросла на 151%. В ряде случаев относительные величины выражаются в ***промилле*** (когда базу сравнения принимают за 1000) и обозначаются ‰. Промилльные отношения применяются, чтобы избежать дробных показателей, которые трудно воспринимаются. Относительные **величины в промилле часто используют**

для характеристики демографических процессов. В частности, показатели рождаемости, смертности и др. обычно выражаются в промилле. Например, рождаемость в РУз в 1990 г. составляла 34,0‰, а смертность – 6,1‰, т.е. на каждую 1000 человек населения приходилось 34 родившегося и 6 умершего. Иногда при расчете относительных величин основание принимается за **10000**, за **100 000**, за **1 000 000**. Примером относительных величин, при расчете которых база принимается 10000 и 100000, могут служить следующие показатели (табл. 4.4).

Таблица 4.4

Основные показатели здравоохранения РУз за 2005-2017 гг.

№		Формы выражения	Человек населения			
			2005	2010	2014	2017
1	Численность врачей всех специальностей	‰о	29,1	27,4	26,4	26,1
2	Число больничных коек	‰о	54,1	47,9	42,2	41,6
3	Численность больных, состоящих на учете в лечебно-профилактических учреждениях	‰оо	319,7	332,4	354,0	296,2

Источник: *Статистический ежегодник Госкомстата РУз. Т., 2015. С. 118, 123.*

К относительным величинам, выраженным на **1 000, 10 000, 100 000** и т.п. единиц, прибегают в целях придания им более удобного для восприятия вида, так как, подобрав соответствующую базу сравнения, можно освободиться от очень маленьких дробных чисел. При сопоставлении разноименных величин получаются именованные относительные величины, выраженные в тех или иных именованных единицах измерения. Например, показатель плотности населения измеряется числом жителей, приходящихся на **1** $км^2$ территории страны (района). Какой из перечисленных форм выражения относительных величин следует воспользоваться в каждом конкретном случае, надо решать исходя из обрабатываемых материалов и результатов, которые получаются при сопоставлении одной величины с другой. Нужно брать такую форму, которая с большей ясностью и наглядностью выразит данное соотношение.

Приступая к исчислению относительных величин при изучении того или иного общественного явления, необходимо тщательно обосновывать выбор базы сравнения. Так, например, если изучают динамику урожайности, то нельзя вести сравнения с данными за тот год, когда в силу климатических условий была самая низкая урожайность. В этом случае всегда, получится приукрашенная картина. И наоборот, если вести сравнения с данными за высокоурожайный год, то мы также не получим реальной картины.

4.6. Виды относительных величин

Относительные величины различаются не только по форме выражения, но и по сущности выражаемых ими количественных соотношений. По этому признаку относительные величины можно подразделить на следующие виды (рис. 4.3).

Виды ОВ

- ОВ планового задания (ОВпз)
- ОВ выполнения планового задания (ОВвпз)
- ОВ динамики (ОВд)
- ОВ структуры (ОВ ст)
- ОВ координации (ОВк)
- ОВ интенсивности (ОВи)
- ОВ сравнения (ОВср)

Рис. 4.3. **Виды относительных величин**

Относительная величина планового

Относительная величина планового задания определяется как отношение планового задания к фактическому результату, достигнутому в предшествующий период, и выражается в процентах.

$$ОВпз = \frac{\textit{Плановое задание на предстоящий период}}{\textit{Фактическое выполнение за базисный период}} = \frac{Упз}{У_0} \cdot 100.$$

Это и есть темпы роста планового задания на предстоящий период. Это отношение показывает, во сколько раз плановое задание больше уровня, достигнутого в базисном периоде (или какую часть его составляет). Для того чтобы выразить плановое задание в процентах, достаточно данное отношение умножить на **100**. Так, например, в 2017 г. в республике произведено 254 трлн. сумов ВВП. А в 2018 г. было предусмотрено увеличить ВВП до 268 трлн. сум, то темп роста планового задания составит: 268:254 = 1,057 или 105,7%. Это означает, что плановым заданием было предусмотрено увеличение выпуска ВВП на 5,7%.

При разработке планового задания первоначально отталкиваются от уже достигнутых результатов, при этом учитываются все внутренние и внешние предпосылки, способствующие установлению плана на том или ином уровне: производственные возможности предприятия, его предстоящие реорганизации, результаты маркетинговых исследований и т.д.

Выше говорилось о том, что плановые показатели могут быть установлены в виде относительных величин. Однако наиболее часто их устанавливают в виде абсолютных или средних величин. Так, например, в абсолютных величинах обычно выражаются плановые показатели по производству продукции, по размерам посевных площадей под урожай соответствующего года и др. В виде средних величин в планах устанавливают такие показатели, как урожайность сельскохозяйственных культур (средняя урожайность), среднегодовая численность персонала и др.

Плановые задания в форме относительных величин обычно устанавливаются по таким показателям, как производительность труда, себестоимость продукции и др. Например, устанавливается задание: снизить себестоимость продукции на 5% по сравнению с прошлым периодом, повысить производительность труда на **8%** и т.п.

Относительная величина выполнения планового задания

Относительные величины выполнения планового задания выражают степень выполнения плановых заданий за определенный период времени и исчисляются как отношение фактически достигнутого уровня к плановому заданию в процентах:

$$ОВвпз = \frac{\textit{Фактическое выполнение}}{\textit{Плановое задание}} = \frac{У_{1(Ф)}}{Упз} \cdot 100.$$

Значение относительных величин выполнения планового задания в статистике определяется тем, что они являются необходимым орудием решения статистикой одной из наиболее важных ее задач – задачи контроля за ходом выполнения плановых заданий с помощью относительных величин по отдельным предприятиям, областям, районам и по секторам экономики в целом имеет большое практическое и познавательное значение. Относительные величины выполнения плановых заданий характеризуют ход и результат борьбы трудового коллектива за выполнение и перевыполнение принятых плановых заданий.

Здесь всякое отклонение от 100% (в любую сторону) означает нарушение оптимального процесса производства и реализации продукции (при условии, что план имеет реальную основу для его выполнения и является научно обоснованным).

Относительные величины динамики

Относительные величины динамики характеризуют изменения одноименных явлений во времени и получаются в результате сопоставления показателей каждого последующего периода с предыдущим или первоначальным. В первом случае получаем относительные величины динамики с переменной базой сравнения ***(цепные)***, во втором – относительные величины динамики с постоянной базой сравнения ***(базисные)***.

$$ОВд = \frac{y_1}{y_0} \text{ (базисные)}; \quad ОВд = \frac{y_w}{y_{i-1}} \text{ (цепные).}$$

Об этом подробно в гл. 9.

Основным условием правильного их вычисления является сопоставимость отчетной и базисной величин. Выбор базы для сравнения в каждом конкретном случае зависит от поставленной задачи.

Относительные величины динамики, планового задания и выполнения плана находятся в определенной взаимосвязи, а именно: произведение относительных величин выполнения плана и планового задания равно относительной величине динамики. Для доказательства этого обозначим фактически достигнутый уровень текущего периода через y_1, базисного периода – как y_0, уровень, предусмотренный планом – $y_{пл}$. Тогда $\frac{y_1}{y_{пл}}$ – относительная величина выполнения плана; $\frac{y_{пл}}{y_0}$ – относительная величина планового задания; $\frac{y_1}{y_0}$ – относительная величина динамики и очевидно, что

$$\frac{y_1}{y_0} = \frac{y_1}{y_{пл}} \times \frac{y_{пл}}{y_0}$$

Взаимосвязь относительных величин выполнения плана, планового задания и динамики дает возможность по двум имеющимся относительным величинам определять третью относительную величину.

Относительные величины структуры

Относительные величины структуры характеризуют состав изучаемой совокупности и показывают, какой удельный вес (какую долю) в общем итоге составляет каждая ее часть. Они получаются в результате деления значения каждой части совокупности на их общий итог, принятый за базу для сравнения

$$(ОВ_{СТ} = \frac{n}{\sum n}).$$

Сумма **относительных** величин структуры изучаемой совокупности всегда равна 100%, или 1.

Примером относительных величин структуры могут служить показатели, приведенные в 3 и 5 графах табл. 4.5.

Таблица 4.5

Посевные площади сельхозкультур РУз в 2014 г.

Культуры	2005 г.		2014г.		2014г. в % к 2005 г.
	посевная площадь (тыс. га)	в процентах к итогу	посевная площадь (тыс. га)	в процентах к итогу	
1	2	3	4	5	6
Зерновые	1616,1	44,3	1655,6	45,0	102,4
Хлопчатник	1472,3	40,4	13,01,5	35,4	88,4
Картофель и овоще-бахчевые культуры	222,7	6,1	324,6	8,8	145,8
Кормовые культуры	290,3	7,9	325,7	8,9	112,1
Прочие	46,7	1,3	70,8	1,9	153,6
Итого	3647,5	100,0	3678,2	100,0	100,8

***Источник**: Статистический ежегодник Госкомстата РУз. Т., 2015. С. 228.*

Первый из них (процент к итогу зерновых в 2014 г.) представляет собой отношение посевной площади под зерновыми (1655,6 тыс. га) к общей посевной площади (3678,2 тыс. га), выраженное в процентах (1655,6 / 3678,2 · 100 = 45%) (аналогично рассчитаны и остальные показатели 3 и 5 графы таблицы).

Показатель отношения части к целому называют удельным весом или долей. В нашем примере наибольшим удельным весом обладают зерновые культуры. Их доля в общей посевной площади РУз равна 45,0%.

Относительные величины, характеризующие внутреннюю структуру общественных явлений, широко применяются в различных отраслях статистики и в экономическом анализе. При их помощи устанавливается половозрастная и национальная структура населения, структура НБ, ВВП, госбюджета и основных фондов, инвестиции, структура товарооборота и т.п.

Сопоставление структуры той или иной совокупности за два или несколько последовательных периодов позволяет установить *структурные изменения,* происшедшие в ее составе, их направление и тенденцию.

Относительные величины структуры неразрывно связаны с группировкой и вычисляются на ее основе, причем не только по численности единиц совокупности и ее частей, но и по значениям различных других признаков. Примером этому могут служить относительные величины, приведенные в следующей табл. 4.6.

Таблица 4.6

Группировка промышленных предприятий РУз по объему валовой продукции за 2018 г.

Группы предприятий с объемом валовой продукции	В процентах по итогу		
	число предприятий	среднегодовая численность рабочих	валовая продукция
До 100	9,0	0,4	0,1
101–500	20,2	3,5	1,4
501–5 000	55,0	34,6	24,3
5 001–50 000	14,6	42,2	46,1
50 001 и выше	1,2	19,3	28,1
Итого	100,0	100,0	100,0

Приведенная таблица содержит интересные данные. В частности, можно отметить, что предприятия первой и второй группы составляют почти треть общей численности всех промышленных предприятий (29,2%), но производят лишь 1,5% общего объема промышленной продукции. Таким образом, их удельный вес в производстве продукции почти в 20 раз ниже удельного веса в общей численности предприятий. А вот предприятия последней группы, доля которых в общем числе предприятий составляет всего 1,2%, дают продукции в 18 раз больше, чем предприятия первой группы. Таблица показывает, что главными производителями промышленной продукции являются не масса мелких или даже средних предприятий с объёмом продукции до 5000 млн сум, а более крупные предприятия. Почти три четверти всего объема промышленной продукции (74,2%) дают крупные предприятия с объемом продукции более 5000 млн сум, составляющие 15,8% общего числа промышленных предприятий в стране.

Эта таблица позволяет проследить и другую важную закономерность, а именно, что доля крупных предприятий в объеме производства больше, нежели в численности рабочих, так как труд на крупных предприятиях более производителен. Действительно, доля предприятий с объемом продукции более 5000 млн сум в численности рабочих составляет 61,5%, а в производстве продукции – 74,2%, т.е. намного больше. Еще больше это превышение для группы наиболее крупных предприятий (с объемом продукции более 50 000 млн сум); доля этой группы в производстве продукции почти в 1,5 раза больше, чем ее же доля в численности рабочих.

Относительная величина координации (ОВк)

Относительная величина координации (ОВк) рассчитывается только для сгруппированных данных и показывает отношение между частями совокупности:

$$ОВк = \frac{\textit{Количество единиц одной части совокупности}}{\textit{Количество единиц другой группы совокупности}} = \frac{n_i}{n_j} \cdot 100.$$

По сути своей относительные показатели координации относятся к той же группе показателей, что и относительные показатели структуры, и дополняют их. Относительные показатели координации позволяют делать сравнения внутри изучаемой статистической совокупности.

К таким показателям относится число сельских жителей, приходящееся на 100 городских, или число лиц, занятых в финансовой сфере, на каждые 100 человек в реальном секторе и т.д. (табл. 4.7).

В 2017 г. в нашей республике было (включая все возрастные группы) 16124 тыс. женщин и 16532 тыс. мужчин. Найдем, что на 100 мужчин приходилось 97 женщин $\left(\frac{16124}{16532}\cdot 100\right)$. Благодаря урбанизации за рассматриваемый период число сельских жителей приходящиеся на 100 городских уменьшилось с 177 чел. до 97. Число лиц занятых в финансовой системе на каждые 100 занятых в промышленности остались неизменными, составив 4 чел.

Таблица 4.7

Динамика показателей ОВ к (2018 г.)

№	Показатели	2005	2010	2015	2017
1.	Число сельских жителей приходящиеся на 100 городских	177	95	97	97
2.	Число лиц занятых в финансовой сфере, на каждые 100 чел занятых в промышленности	4	4	4	4

Среди новорожденных наблюдается «избыток» мальчиков. Многолетние данные показывают, что обычно рождаются 105-106 мальчиков на каждые 100 девочек.

Относительные величины координации являются дополнением к характеристике структуры, они дают возможность осуществлять контроль за соблюдением необходимых пропорций между отдельными частями совокупности.

Относительные величины интенсивности (ОВи)

Относительными величинами интенсивности называют показатели, характеризующие распространение, развитие какого-либо явления в определенной среде. Они измеряют степень или интенсивность его распространения.

Относительные величины интенсивности представляют собой соотношение разноименных, но связанных между собой величин. В числителе отношения берется величина явления (показатель), степень распространения которого изучается, а в знаменателе объем той среды, в которой происходит развитие (распространение) этого явления.

Примером относительных величин интенсивности могут служить демографические коэффициенты, например, коэффициент смертности, показывающий, сколько умерших в течение года приходится на **1000** человек населения (в числителе берется число умерших за год на данной территории, а в знаменателе – средне-годовая численность населения той же территории (см. гл. 11, табл. 5.3).

К этому же виду относительных величин принадлежит показатель плотности населения, представляющий собой отношение численности населения данной страны, области, района и т.п. к соответственной территории и выражаемый числом жителей, приходящихся на 1 *км*^{2} площади. Приведем, для примера, данные о плотности населения по некоторым областям нашей страны (табл. 4.8).

Таблица 4.8

Плотность населения по некоторым областям РУз

№		2018
	Плотность населения (на 1 км^2, человек)	
1	Республика Узбекистан	73
2	РКК	11
3	Навоийская область (min)	9
4	Андижанская область (max)	705

К числу относительных величин интенсивности относятся показатели уровня экономического развития, т.е. показатели, характеризующие размеры производства различных видов продукции на душу населения.

Показатели уровня экономического развития играют весьма важную роль в социально-экономическом анализе. Без этих относительных величин нельзя обойтись в международных сравнениях при изучении уровней экономического развития. При расчете относительных величин уровня экономического развития в числителе берется годовой объем производства (количество) данного вида продукции, а в знаменателе – среднегодовая численность населения за тот же год, за которую обычно принимается полусумма численности населения на начало и конец года.

В отличие от других видов относительных величин показатели интенсивности являются именованными числами и имеют размерность тех абсолютных величин, соотношение которых они выражают.

При вычислении показателей интенсивности большое значение имеет правильный выбор базы, с которой следует сопоставлять изучаемое явление. За базу сравнения нужно принимать, как правило, только ту совокупность (среду), в которой может иметь место изучаемое явление. Например, выход молока и мяса крупного рогатого скота рассчитывается на **100** *га* всех сельскохозяйственных угодий, а выход свиного мяса – на **100** *га* пашни.

Приведем несколько примеров относительных показателей интенсивности, которые широко используются в области экономического анализа:

- ***фондоотдача,*** рассчитанная по показателю товарной продукции, показывает, сколько товарной продукции приходится на 1 сум основных производственных фондов;
- ***фондоемкость*** (обратный показатель фондоотдачи) показывает, сколько основных производственных фондов приходится на 1 сум товарной продукции;
- ***материалоотдача,*** рассчитанная по показателю товарной продукции, показывает, сколько товарной продукции приходится на 1 сум израсходованных материалов;
- ***материалоемкость*** (обратный показатель материалоотдачи) показывает, сколько израсходованных материалов приходится на 1 сум товарной продукции;
- ***затраты на 1 продажу*** показывают, сколько себестоимости проданных товаров, продукции, работ, услуг приходится на выручки от продажи товаров, продукции, работ, услуг;
- ***рентабельность продаж*** показывает, сколько прибыли от продаж приходится на выручку от продажи товаров, продукции, работ, услуг.

Все представленные относительные показатели интенсивности имеют размерность *сум/сум.*

Попутно отметим, что можно еще достаточно много построить показателей, родственных этим. Например, фондоотдачу (а также – фондоемкость, материалоотдачу, материалоемкость) можно рассчитать и по показателю реализованной продукции (выручке от продаж). При расчете фондоотдачи и фондоемкости мы можем брать не все основные производственные фонды, а только их активную часть, и рассчитанная таким образом фондоотдача будет характеризовать только активную часть основных производственных фондов.

Форма выражения у показателей интенсивности может быть разная. При этом размерность полученного показателя интенсивности, если и не указывается, однако обязательно подразумевается и легко определяется на основании исходных расчетных данных и экономической сущности самого показателя интенсивности. Так, например, если рентабельность продаж равна 20%, то это означает, что на каждый 1 сум выручки от продаж приходится 20 тийин прибыли от продаж. Или еще пример: если коэффициент рождаемости равен 22 ‰, то это означает, что на каждую тысячу человек населения приходится 22 родившихся.

Относительная величина сравнения (ОВср)

Относительная величина сравнения (ОВср) показывает отношение одноименных величин, хатеризующих разные объекты и относящихся к одному периоду:

$$ОВср = \frac{\textbf{Фактич. уровень явления на территории А за определенный период времени}}{\textbf{Фактич. уровень того же явления за тот же период времени на территории В}} = \frac{У_A}{У_B}$$

Другими словами относительными величинами сравнения называются относительные показатели, характеризующие сравнительные размеры одноименных величин, относящихся к одному и тому же периоду или моменту времени, но к различным объектам или территориям. Обычно они исчисляются в процентах или в кратных отношениях, показывающих во сколько раз одна сравниваемая величина больше (меньше) другой. Примером относительных величин сравнения могут служить, следующие данные о соотношении производства некоторых видов продукции в России и в Республике Узбекистан в 2014 г. (табл. 4.9).

Таблица 4.9

Производство важнейших видов промышленной продукции в Республике Узбекистан и в России в 2014 г.

№	Показатели	Узбекистан	Россия	Россия в разах к Узбекистану (гр. 2 / гр. 1)
А	Б	1	2	3
1	Производство электроэнергии (млрд. кВт-часов)	55,8	1064,0	19
2	Добыча нефти (млн т)	2,9	526,0	181
3	Производство зерновых и зернобобовых культур (млн т)	8,3	105,3	13

В последней графе таблицы приведены относительные величины сравнения, выраженные в разах. Они показывают, что Россия больше производит электроэнергию в 19 раз, нефти – 181 раз и зерновых культур 13 раз, чем Республика Узбекистан. Иначе говоря, Узбекистан производит 1/19 части электроэнергии (т.е. 5,2%), 1/181 части нефти (т.е. 0,6%) и 1/13 части зерновых (т.е. 7,9%) России.

Относительные величины сравнения широко применяются при сравнительной оценке показателей работы отдельных предприятий, городов, районов, областей. Для их расчета используют как абсолютные величины, так и относительные, например, сопоставляется объем промышленной продукции и США, темпы роста (относительные величины динамики) этих стран и т.п.

4.7. Общие принципы построения и использования абсолютных и относительных статистических величин

Правильное применение абсолютных и относительных величин для характеристики общественных явлений и процессов возможно только на основе соблюдения некоторых общих принципов.

Очень важным требованием экономико-статистического анализа является изучение абсолютных и относительных показателей в их единстве. За одними и теми же относительными показателями могут, скрываться разные абсолютные массы. Так, например, в ходе экономико-статистического анализа темпов развития не следует забывать, какие абсолютные величины скрываются за темпами роста и прироста. В частности, нужно иметь в виду, что при снижении (замедлении) темпов роста абсолютный прирост может возрастать, Возьмем, например, валовой сбор зерновых в Узбекистане (табл. 4.10).

Таблица 4.10

Производство зерна в РУз

Годы	Зерна (тыс. т)	Темпы прироста за 5 лет (%)	Абсолютный прирост за 5 лет (тыс. т)
2005	6401,8	–	–
2010	7404,1	15,7	1002,3
2015	8300,5	12,1	836,4

За 2005-2010 гг. т.е. за 5 лет производство зерна увеличилось на 15,7%, т.е. около 1,2 раза, а за последующее пятилетие только на 12,1%. Однако в абсолютном выражении прирост производства зерна за второе пятилетие был больше, чем за первое.

Так, в 2010 г. производство зерна по сравнению с 2005 г. возросло на 15,7%, или на 1002,3 тыс. т. Следовательно каждый из этих 15,7% прироста в абсолютном выражении составил 63,8 тыс. тонн (1002,3 : 15,7).

В течение следующего пятилетия производство увеличилась на 836,4 тыс. т. или на 12,1% уровня 2010 года. Каждый процент прироста составил уже 69,1 тыс. т. (836,4 : 12,1), т.е. 8,3% больше, чем в предыдущем пятилетии. Использованный нами показатель – абсолютное значение одного процента прироста.

Как видно из приведенного примера, абсолютное значение 1% прироста находится путем деления абсолютного прироста на выраженный в процентах темп прироста. Можно, однако, исчислить этот показатель технически более простым путем. Так как за 100% всегда принимается базисный уровень, то 1% будет в 100 раз меньше базисного уровня:

$$A = \frac{Убаз}{100}$$

Так, для первого пятилетия получим: А = 6401,8 : 100 = 64,0 (тыс. *т)*.

Указанными особенностями взаимосвязи между абсолютными и относительными величинами и обусловливается необходимость их комплексного использования в анализе. Взятые в отрыве одни от других, они не дают полного представления об изучаемых явлениях и процессах.

Особенно серьезное внимание при исчислении относительных показателей следует уделять вопросу *сопоставимости сравниваемых абсолютных величин*. В первую очередь, это относится к расчету относительных величин выиолнения плана, динамики и сравнения. Сопоставимость абсолютных величин должна быть обеспечена в различных отношениях. Прежде всего необходимо, чтобы все сравниваемые абсолютные величины характеризовали одно и то же явление и были однородны по содержанию и границам объекта, который они характеризуют. Так, сопоставляя производство электроэнергии за несколько лет или в нескольких странах, нельзя в одном случае включать, а в другом – не включать в общий итог электроэнергию, израсходованную на собственные нужды электростанций, электроэнергию, произведенную передвижными электростанциями, электростанциями на судах, в поездах и т.п. Очень важным условием сопоставимости абсолютных величин является одинаковая методология их исчисления. Известно, например, что средняя урожайность сельскохозяйственных культур может исчисляться либо в расчете на 1 *га* площади посева, которая имелась к концу весеннего сева, либо в расчете на 1 *га* фактически убранной площади, т.е. за вычетом площади посевов, погибших летом. Поэтому при сопоставлении средней урожайности той или иной культуры за несколько лет или по отдельным районам необходимо во всех случаях сравнивать урожайность, исчисленную одинаковым методом, т.е. в расчете либо на гектар посева, либо на гектар фактически убранной площади.

Необходимо также следить за тем, чтобы все сравниваемые величины были выражены в одинаковых единицах измерения. Нельзя, например, продукцию одной ткацкой фабрики, выраженную в погонных метрах, сопоставлять с продукцией другой фабрики, выраженной в квадратных метрах, и т.п.

Интеллектуальный тренинг

1. Какие величины называются в статистике абсолютными?

2. Назовите единицы измерения абсолютных величин.

3. Что такое относительные величины?

4. Назовите виды относительных величин.

5. Назовите важнейшие формы выражения относительных величин.

6. Что называется базой сравнения относительных величин?

7. Каковы основные условия правильного применения абсолютных и относительных величин?

8. «Отношение, выраженное, например, тем, что 10% дворов имеют 30% посева, абстрагирует различие абсолютных цифр и потому годно для сравнения со всяким подобным отношением любой местности». Но для такого сравнения надо выделить другой местности, сколько процентов дворов?

9. Так ли это, что изолированно взятый абсолютный или относительный показатель может очень односторонне отразить изучаемое явление или процесс, в то время как рассмотрение абсолютных и относительных показателей в их единстве устраняет такую односторонность. Поясните это не примере.

10. За одними и теми же относительными показателями могут скрываться разные абсолютные массы. Поясните это на примере.

11. Одни и те же абсолютные величины могут означать разные доли или разные интенсивности. Может ли быть так? Если да, то поясните это на примере.

12. В условиях инфляции стоимостные показатели требуются доводить до уровня сопоставимости. Но это их минус же. Почему?

13. Почти все виды относительных величин не имеют размерности. Так ли это? Если нет, то, какие относительные величины составляют исключение в этом правиле?

14. Какие относительные показатели позволяют делать сравнение внутри изучаемой статистической совокупности?

15. Для чего исчисляется показатель абсолютное значение 1% прироста?

16. В чем необходимость совместного применения абсолютных относительных величин?

17. Особенно серьезное внимание при исчислении относительных показателей следует уделять вопросу сопоставимости сравнения абсолютных величин. Почему? Поясните это на примере.

Использованная и рекомендуемая специальная литература

1. Суслов И.П. Теория статистических показателей. М.: Статистика, 1971.

2. Суслов И.П. Основы теории достоверности статистических показателей. Новосибирск: Наука, 1979.

3. Статистический словарь. М.: ВИТРЕМ, 2002.

Глава V
СРЕДНИЕ ВЕЛИЧИНЫ

Дорожная карта

5.1. Сущность и значение средних величин

Основные положения теории

Средняя – это один из распространенных приемов обобщений. Важность, средних величин для статистической практики и науки отмечалась в работах многих ученых. Так, английский экономист В. Петти (1623–1667) при рассмотрении экономических проблем широко использовал средние величины. В частности, он предлагал использовать в качестве меры стоимости затраты на среднее дневное пропитание одного взрослого работника. Его не смущала абстрактность средних, то, что данные, относящиеся к отдельным конкретным людям, могут не совпадать со средней величиной. Он считал устойчивость средней величины как отражение закономерности изучаемых явлений. Это положение можно проиллюстрировать на следующем примере.

Метод средних величин заключается в замене *индивидуальных значений* варьирующего признака единиц наблюдения, т.е. в замене x_1 x_2, x_3,..., x_n некоторой уравненной величиной

Например, индивидуальная выработка у пяти рабочих за месяц составила 135, 141, 153, 159, 162 детали. Чтобы получить среднюю выработку одного рабочего за месяц, нужно сложить эти индивидуальные показатели и полученную сумму разделить на пять:

$$\overline{X}=\frac{135+141+153+159+162}{5}=\frac{750}{5}=150 \text{ } \textit{деталей}.$$

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

Отличительной особенностью статистических средних является то, что в них взаимно погашаются индивидуальные различия признака у отдельных единиц изучаемой совокупности, в результате чего представляется возможность охарактеризовать общие черты и свойства массовых общественных явлений, закономерности их развития.

Отмеченное выше свойство средней констатировано в трудах немецкого экономиста К.Маркса (1818-1883). Труд, овеществленный в стоимости, – писал он в «Капитале», – есть труд среднего общественного качества, т.е. проявление средней рабочей силы. Но средняя величина есть всегда средняя многих различных индивидуальных величин одного и того же вида. В каждой отрасли промышленности индивидуальный рабочий, Петр или Павел, более или менее отклоняется от среднего рабочего. Такие индивидуальные отклонения, называемые на языке математиков «погрешностями», взаимно погашаются и уничтожаются, раз мы берем значительное число рабочих».[1]

Весьма широко применял средние и относительные величины английский ученый Г. Кинг (1648-1712) при анализе данных о населении Англии (средний доход на одну семью, среднедушевой доход и т.д.).[2]

Теоретические разработки бельгийского статистика А. Кетле (1796-1874), внесшего значительный вклад в разработки теории устойчивости статистических показателей, основаны на противоречивости природы социальных явлений – высокоустойчивых в массе, вместе с тем сугубо индивидуальных.

Согласно учению Кетле массовые процессы и явления формируются под влиянием двух групп причин. В первую группу общих для всех единиц массовой совокупности причин относятся причины, определяющие состояние массового процесса. Они формируют типичный уровень для единиц данной качественно-однородной совокупности и связаны с сущностью изучаемого явления.

Вторая группа (индивидуальных) причин формирует специфические особенности отдельных единиц массовой со-

---

[1] Маркс К. и Энгельс Ф. Соч. Изд. 2. Т. 23. С. 334.

[2] Плошко Б.Г., Елисеева И.И. История статистики. М.: Финансы и статистика, 1990. С. 23, 27.

вокупности, а следовательно, их отклонения от типичного уровня. Эти причины не связаны с природой изучаемого явления, их называют случайными причинами.

При исчислении средней величины до массы единиц влияние случайных причин взаимопогашается, и средняя, абстрагируясь от индивидуальных особенностей отдельных единиц совокупности, выражает общие свойства, присущие всем единицам. Кетле подчеркивал, что статистические средние представляют собой не просто метод математического измерения, а категорию объективной действительности. Принципиальная суть статистического познания состоит в погашении случайного, вызванного действием индивидуальных причин, и в выявлении закономерностей, обусловленных общими причинами.

Важный вклад в обоснование и развитие теории средних величин внесли последователи А.Кетле немецкий статистик В.Лексис (1837-1914), перенесший теорию «истинных величин» на экономические явления общественной жизни. Его теория известна под названием «теория устойчивости». Другая разновидность идеалистической теории средних основана на философии махизма. Ее основатель – английский статистик А. Боули (1869-1957); является одним из самых видных теоретиков новейшего времени в области теории средних величин. Его концепция средних величин изложена в книге «Элементы статистики». А. Боули рассматривает средние величины лишь с количественной стороны, тем самым отрывает количество от качества. Определяя значение средних или, как он выражается, «их функцию», Боули на первый план выдвигает махистский принцип мышлений. Так, он писал, что функция средних ясна: она заключается в том, чтобы выражать сложную группу при помощи немногих простых чисел. Ум не в состоянии сразу охватить величины миллионов статистических данных, они должны быть сгруппированы, упрощены, приведены к средним. Взгляд на метод средних как на технический прием упрощений цифровых материалов разделяли Р. Фишер (1890-1968), Дж. Юл (1871-1951), Фредерик С. Миллс (1892-1964) и др.

В 30-е и последующие годы средняя величина все чаще стала рассматриваться как социально значимая характеристика, информативность которой зависит от однородности данных. Однако зарубежная статистика не ставит вопрос о связи между средними величинами по разным признакам, не рассматривает системы средних.

Виднейшие представители итальянской школы Бенини (1862- 1956) и Коррадо Джини (1884-1965), считая статистику отраслью логики, расширили область применения статистической индукции. Причем познавательные принципы логики и статистики они связывали с природой изучаемых явлений, следуя традициям социологической трактовки статистики.[1]

Возможностью перехода от единичного к общему, от случайного к закономерному объясняется важность метода средних величин и его широкое применение в статистических исследованиях. Большой вклад в разработку средних величин и решение вопросов их практического применения внесли известные российские ученые, И.С. Пасхавер, А.Я. Боярский, Н.К. Дружинин и др. Средние величины применяются для оценки достигнутого уровня изучаемого показателя, при анализе и планировании производственно-хозяйственной деятельности, предприятий (объединений), фирм, банков и других хозяйственных единиц; средние используются при выявлении взаимосвязей явлений, при прогнозировании, а также расчете нормативов. Средняя величина всегда именованная, имеет ту же размерность (единицу измерения), что и признак у отдельных единиц совокупности.

Таким образом:

Средними величинами в статистике называют такие показатели, которые выражают типичные черты и дают обобщенную количественную характеристику уровня какого-то варьирующего признака по совокупности однородных общественных явлений. Метод средних величин представляет особую, форму статистического обобщения. Применение

[1] Плошко Б.Г., Елисеева И.И. История статистики. С. 164-165.

метода средних величин возможно только при наличии вариации размеров уровня какого-либо признака у совокупности однородных явлений.

Если бы все единицы изучаемых статистикой массовых явлений характеризовались одинаковыми размерами какого-то признака, то не было бы необходимости пользоваться средними величинами при изучении этого признака. Если бы выработка всех ткачей в какую-то единицу времени была одинаковой (например, 3,5 м тканей в час), то не нужно было бы прибегать к вычислению средней часовой выработки.

Средняя величина характеризует всю массу единиц изучаемой совокупности и, выражая то общее, что характерно для данной совокупности, непосредственно не характеризует отдельные единицы, из которых состоит эта масса общественных явлений.

Средние величины могут быть как абсолютными, так и относительными. Так, вышеприведенная средняя заработная плата рабочих является абсолютной средней величиной. Средний же процент выполнения плана реализации продукции по группе промышленных предприятий представляет собой относительную среднюю величину.

5.2. Условия правильного применения средних в статистике

Основными условиями научного использования средних являются:

Во-первых, средняя величина должна исчисляться лишь для совокупности, состоящей из единиц однородных, одного и того же вида. Только при этом условии средняя будет обобщающим показателем изучаемой совокупности.

Средняя, исчисленная для качественно неоднородной совокупности, скрывает коренные различия между ее отдельными единицами. Тем самым она не показательна, а следовательно, не имеет никакого реального смысла. Пример нетипичной средней хорошо показан в рассказе Глеба

Успенского «Живые цифры». Там средний доход определялся сложением 1 млн миллионера Колотушкина и 1 гроша просвирни Кукушкиной, и делением на два получалось, что он составил 0,5 млн руб. Например, если рассчитывать среднюю заработную плату в кооперативах и на госпредприятиях, а результат распространить на всю совокупность, то средняя фиктивна, так как рассчитана по неоднородной совокупности, и такая средняя теряет всякий смысл.

При помощи средней происходит как бы сглаживание различий в величине признака, которые возникают по тем или иным причинам у отдельных единиц наблюдения.

Например, средняя выработка продавца зависит от многих причин: квалификации, стажа, возраста, формы обслуживания, здоровья и т.д. Средняя выработка отражает общее свойство всей совокупности. Поэтому очень важное правило – вычислять средние величины лишь по однородной совокупности единиц. Только при выполнении этого условия средняя как обобщающая характеристика отражает общее, типичное, закономерное, присущее всем единицам исследуемой совокупности.

Во-вторых, совокупности, неоднородные в качественном отношении, необходимо расчленять на однородные группы и вычислять для них групповые, типические средние, характеризующие каждую из этих групп. В этом заключается связь средних и группировок. Иначе говоря, метод средних должен быть непосредственно связан с методом группировок.

Применение группировки при исчислении средней дает возможность избежать огульных средних и искажения действительности.

Качественная однородность той или иной массы единиц отнюдь не означает их тождественности. Наоборот, величина признака у отдельных единиц различна, она варьирует. Но эта вариация не нарушает качественной однородности данной массы явлений, если она не переходит известные границы, известную меру. Однако и в пределах этих границ уже происходит процесс нарождения косого качества. Диалектический процесс развития явлений приводит в конечном счете к распаду единой, качественно однородной группы явлений на несколько различных групп.

Носителем нового качества сначала являются единичные объекты, а затем количество этих объектов увеличивается, и новое становится массовым, типичным. Новаторы производства сначала обычно представляют собой единичное явление. Но затем метод их работы делается достоянием многих рабочих, и **единичное становится массовым**.

Статистика должна подмечать зародыш будущего и вместе с тем она должна выявлять отстающие участки. Это требует от нее наряду с исчислением средней, характеризующей «господствующую тенденцию», выявлять лучшие образцы и отстающие участки.

Следовательно, прежде чем вычислять средние величины, необходимо произвести группировку единиц исследуемой совокупности, выделив качественно однородные группы.

Средняя, рассчитанная по совокупности в целом, называется общей средней, средние, исчисленные для каждой группы, – групповыми средними. Общая средняя отражает общие черты изучаемого явления, групповая средняя дает характеристику размера явления, складывающуюся в конкретных условиях данной группы.

Например, статистическое изучение рождаемости и среднего количества детей в семье на территории Узбекистана проводится и в региональном аспекте (по областям республики). Традиционно более высокая рождаемость в Кашкадарьинской (22,5%) и Сурхандарьинской (24,6%) по сравнению с Сырдарьинской областью среднее количество детей в семье, исчисленное по каждому региону – это групповые средние, а соответственно, исчисленное по всей территории страны общая средняя.

Сравнительный анализ групповых и общих средних используется для характеристики социально-экономических типов изучаемого общественного явления. В частности, при изучении рождаемости важное значение имеет характеристика этого процесса по общественным группам населения региона.

Групповые средние используются для изучения закономерностей развития общественных явлений. Так, в аналитических группировках анализ групповых средних позволяет сделать вывод о наличии и направлении взаимосвязи между группировочным (факторным) признаком и результативным показателем (примеры рассмотрены в гл. 7).

Групповые средние широко применяются также при определении имеющихся неиспользованных резервов производства, когда наряду со средними величинами рассматриваются и индивидуальные значения признака (методика исчисления рассмотрена в главе 7).

В-третьих, как бы эффектно ни выглядели достижения передовых групп на фоне средних по стране показателей, нельзя забывать о том, что за средними скрываются также низкие показатели, Поэтому средними характеристиками нужно пользоваться с большой осторожностью, не преувеличивая их значения, так как средняя, являясь обобщающим показателем, погашает те количественные различия варьирующего признака, которые проявляются у **отдельных единиц совокупности**.

Общие средние, которыми не видно лица отдельных предприятий, фирм, лиц и т.д., нередко порождают самоуспокоенность и благодушие у хозяйственных руководителей, что приносит большой вред.

Вот почему научная статистика исходит из того, что средние должны дополняться конкретными **индивидуальными данными**. С помощью индивидуальных данных устанавливается и обобщается опыт передовиков, но с помощью этих же данных выявляются и отстающие участки, плохая работа.

Критика методов руководства, исходящих из так называемых общих данных, при которых конкретные факты подменяются средними показателями, содержится в ряде правительственных и ведомственных решениях.

5.3. Виды средних и исходная база их расчета

Исходные основвания расчета средних

В статистике применяют разные виды средних величин. Наиболее распространенными являются следующие (рис. 5.1).

Виды средних

Степенные	Структурные
• средняя арифметическая • средняя геометрическая • средняя гармоническая • средняя квадратическая • средняя хронологическая	• мода • медиана

Рис. 5.1. **Виды средних**

Вычисление средних в статистике существенно отличается от их вычисления в математике. В математике рассматриваются возможные формы средней, (арифметическая, геометрическая, гармоническая и др.), их свойства и способы расчета. При этом выводятся соотношения различных средних, полученных из одного и того же ряда чисел (вариантов признака).

По-иному обстоит дело в статистике. Изучая количественную сторону массовых общественных явлений в неразрывной связи с их качественной стороной, статистика исходит из принципа осмысленности результатов при суммировании или при взвешивании. Только тогда средняя применима правильно, когда получают величины, имеющие реальный экономический смысл. Поэтому в статистике введены следующие понятия и обозначения: признак, по которому находится средняя, называется ***осредняемым признаком*** и обозначается

$\bar{x}$; величина осредняемого признака у каждой единицы совокупности называется ***индивидуальным его значением*** или ***вариантами,*** и обозначается как $x_1, x_2, x_3, ..., x_n$; частота – это повторяемость индивидуальных значений признака, обозначается буквой f.

Трудным вопросом методологии является вопрос о выборе вида средней. Здесь важно твердо усвоить, что решение этого вопроса зависит от характера исходного соотношения, выражающего данную среднюю величину, от содержания осредняемого признака, его связи с другими признаками, а также от особенностей исходного материала. Каждый из видов средних величин может выступать либо в форме простой, либо в форме взвешенной средней.

Мы предлагаем читателю способ определения нужной формы средней, основанный на принципе выяснения сущности средней, ее социально-экономического содержания. Ведь средняя величина признака – это отношение. Поэтому, прежде чем оперировать цифрами, необходимо выяснить, соотношением каких показателей, каких величин (в конечном счете) является средняя в данном случае. Это надо попробовать записать словами в виде формулы, которая и будет *логической формулой средней.*

Иначе говоря исчисления средних лучше начинать с исходного соотношения средней (ИСС), а её логическая формула выглядит так:

$$\textbf{\textit{ИСС}} = \frac{\textbf{\textit{Суммарное (итоговое) значение (объем усредняемого признака)}}}{\textbf{\textit{Число единиц (объем совокупности)}}}$$

После того как словами записана логическая формула средней, которую надо вычислить, необходимо внимательно рассмотреть имеющиеся для вычисления данные и заменить словесные значения числителя и знаменателя логической формулы средней соответствующими числовыми значениями, после чего остается только рассчитать и получить ответ. Этот принцип обеспечит правильный выбор формы средней, а значит, и правильное определение величины средней. Еще одно важное свойство принципа логической формулы средней заключается в том, что здесь не возникает проблемы выбора весов средней, которая часто порождает ошибки.

Иллюстрируем сказанное примером.

Допустим, требуется определить средние затраты рабочего времени на изготовление одной детали по следующим данным (табл. 5.1).

Таблица 5.1

Затраты рабочего времени четырех рабочих, занятых изготовлением одного вида деталей и их выработка

№ п/п	Выработано деталей за отчетный месяц (f)	Отработано человеко-часов (x·f)	Затрачено человеко-часов на одну деталь ($\overline{x}$)
	1	2	3
1	30	174	5,8
2	28	154	5,5
3	29	174	6,0
4	25	155	6,2
	112	657	–

Искомая средняя может быть определена по данным двух граф таблицы, имеющих порядковую нумерацию. Так, если, допустим, известны данные первой и второй граф, то для определения средней достаточно итог второй графы (657) разделить на итог первой графы (112). Проделав это, найдем, что в среднем на изготовление одной детали затрачивается 5,9 человеко-часа.

Предположим теперь, что известны только данные первой и третьей граф. В этом случае для получения средней необходимо сначала рассчитать общие затраты рабочего времени на производство всех деталей, а затем найденную величину разделить на количество деталей, т.е. произвести следующие расчеты:

$$\frac{5{,}8\cdot 30+5{,}5\cdot 28+6{,}0\cdot 29+6{,}2\cdot 25}{30+28+29+25}=\frac{657}{112}=5{,}9$$ человеко-часа.

В том же случае, когда известны данные только второй и третьей граф, для получения средней необходимо рассчитать количество произведенных каждым рабочим деталей, для чего надо количество отработанных/человеко-часов каждым рабочим разделить на затраты человеко-часов на одну деталь, т.е. данные второй графы на данные третьей графы. После этого нетрудно уже найти искомую среднюю. Включая все операции, расчет средней в этом случае будет иметь следующий вид:

$$\frac{174+154+174+155}{\frac{174}{5{,}6}+\frac{154}{5{,}5}+\frac{174}{6{,}0}+\frac{155}{6{,}2}}=\frac{657}{112}=5{,}9$$

человеко-часа.

Из рассмотренного примера видно, что какой бы способ расчета средних затрат рабочего времени на производство одной детали ни применялся, в конечном счете дело сводилось к делению затрат рабочего времени на количество выработанных изделий. Это означает, что основой расчета средней в данном примере послужило количественное соотношение затрат рабочего времени и количества изготовленных деталей.

Подобного рода количественные соотношения, обусловленные характером осредняемого показателя (т.е. показателя, средние размеры которого определяются), *являются исходной базой расчета средних величин*. Зная количественное соотношение, всегда можно, при наличии надлежащих данных, определить тем или иным способом среднюю величину.

Формулы расчета статистических средних

Изложенные выше способы расчета статистических средних могут быть обобщены в виде определенных формул. Обозначив осредняемый признак через x, количество единиц, обладающих этим признаком, – через f, произведение xf – через w, а отдельные конкретные их значения – соответственно, через x, f и w со значками внизу 1, 2, 3,..., n, получим следующие три формулы расчета средних:

$$\bar{x}=\frac{\Sigma w}{\Sigma f};\ \bar{x}=\frac{\Sigma xf}{\Sigma f};\quad \bar{x}=\frac{\Sigma w}{\Sigma \frac{w}{x}},$$

где $\bar{x}$ – средняя, а Σ – знак суммы.

Приведенные выше формулы, отображающие различные способы расчета среднего размера осредняемого признака в зависимости от его характера и имеющихся в наличии данных, имеют свое наименование: первая $\left(\bar{x}=\frac{\Sigma w}{\Sigma f}\right)$ называется средней агрегатной, вторая $\left(\bar{x}=\frac{\Sigma xf}{\Sigma f}\right)$ – средней арифметической, третья $\left(\bar{x}=\frac{\Sigma w}{\Sigma \frac{w}{x}}\right)$ – средней гармонической.

Иллюстрируем расчет средней по приведенным формулам на конкретном примере.

5.4. Средняя арифметическая

Наибольшее распространение в статистической практике имеет средняя арифметическая.

Простая средняя арифметическая вычисляется в тех случаях, когда каждый вариант встречается в изучаемом явлении один или одинаковое число раз (табл. 5.1).

Простая средняя арифметическая представляет собой результат деления суммы варианта *(Σx)* на их число *(n)*. Ее можно исчислить по формуле:

$$\bar{x}=\frac{\Sigma x}{n};$$

где $\bar{x}$ – средняя величина;

Σ – знак суммирования (сигма);

x – варианты;

n – число варианта.

Средняя арифметическая взвешенная вычисляется в тех случаях, когда различные варианты встречаются в изучаемой совокупности неодинаковое число раз (табл. 5.1).

Определяют среднюю арифметическую взвешенную по формуле:

$$\bar{x} = \frac{\Sigma xf}{\Sigma f};$$

где f – частоты (веса).

Таким образом, для исчисления взвешенной средней арифметической необходимо *(табл. 5.1)*:

а) каждый вариант умножить на его вес (гр. 3 · гр. 1);

б) найти сумму этих произведений (итог 2 графы);

в) сумму произведений варианта на вес разделить на сумму весов (гр. 2 : гр. 1).

Средняя из вариантов, которые повторяются различное число раз или, как говорят, имеют различный вес, называется взвешенной. В нашем примере мы получили формулу взвешенной средней арифметической.

Как видно выше из приведенной формулы, средняя арифметическая взвешенная равна сумме произведений вариантов (х) на соответствующие им веса (f), деленной на сумму весов.

Взвешенная, средняя арифметическая применяется в тех случаях, когда показатель, находящийся в числителе исходного отношения средней, непосредственно неизвестен, а его величина равна сумме произведений вариантов осредняемого признака (х) на соответствующие слагаемые знаменателя исходного отношения (веса – f), т.е. исходной базы расчета средней.

5.5. Абсолютные и относительные веса средней

При расчете средних взвешенных можно вместо абсолютных данных о весах пользоваться относительными величинами, в частности, процентами к итогу. В этом нетрудно убедиться, если, например, числитель и знаменатель в формуле средней арифметической взвешенной разделить на сумму весов и одновременно умножить на 100.

$$\bar{x} = \frac{\Sigma xf}{\Sigma f} = \frac{\Sigma x \cdot \frac{1}{\Sigma f} \cdot 100}{100} = \frac{\Sigma xf'}{100},$$

где f' – процент к итогу.

Использование в качестве весов процентов к итогу обычно несколько упрощает технику расчета, средних взвешеенных. Иллюстрируем это на примере.

Допустим, требуется исчислить среднюю урожайность озимой пшеницы по следующим данным (табл. 5.2).

Таблица 5.2

Посевная площадь и урожайность озимой пшеницы

Участки	Посевная площадь		Урожайность (в ц с га)
	в га (f)	в процентах к итогу	
№ 1	1198,5	20,4	19,2
№ 2	1034,0	17,6	10,8
№ 3	1222,0	20,8	11,5
№ 4	2420,5	41,2	15,4
Итого	5875,0	100,0	–

Использовав в качестве весов проценты к итогу, найдем:

$$\bar{x} = \frac{19{,}2 \cdot 20{,}4 + 10{,}8 \cdot 17{,}6 + 11{,}5 \cdot 20{,}8 + 15{,}4 \cdot 41{,}2}{100} = \frac{1455{,}44}{100} = 14{,}6 \text{ ц}.$$

Этот же результат получим, если в качестве весов возьмем абсолютные данные. Однако в этом случае расчеты оказались бы несколько более трудоемкими, чем в первом случае. В этом легко убедиться всякий, кто сам произведет оба расчета.

5.6. Исчисление средних из средних

В практике расчета средних нередко приходится иметь дело с такими данными, которые представляют собой средние показатели. Так, рассчитывая среднюю урожайность озимой пшеницы (табл. 5.2) мы имели дело фактически с данными о средней урожайности. Например, средняя урожайность в размере 14,6 ц представляет собой среднюю урожайность нескольких звеньев участка №1. Точно так же и остальные варианты урожайности являются средними показателями. Следовательно, определяя среднюю заработную плату рабочих четырех участков находили среднюю из средних.

Техника исчисления средней из средних, как видно из рассматриваемого примера, та же самая, что и при исчислении ее из первоначальных значений признака: средние, используемые для исчисления общей для них средней, принимаются за варианты, из которых затем производится расчет обычными способами.

Таким образом, общая средняя равна средней из частных (групповых) средних.

5.7. Исчисление средней из величин интервального ряда

Иногда варианты признака, из которых исчисляется средняя, представлены в виде интервалов. Так, например, нам может быть известно, что на площади 100 *га* наблюдается урожай от 10 до 12 *ц*, на площади 120 га – от 12 до 14 *ц* и т.д.

В качестве конкретного варианта, который необходимо умножить на соответствующий вес, в таких случаях условно принимается середина каждого интервала, после чего производится взвешивание обычным порядком.

В приведенном примере такими срединными значениями будут:

$$\frac{10+12}{2}=11;\ \frac{12+14}{2}=13 \text{ и т.д.}$$

В ряде случаев задача исчисления средней по величинам интервального ряда осложняется тем, что не известна одна из границ начального и конечного интервала.

В этом случае, обычно, условно принимается, что расстояние между границами данного интервала такое же, как и в соседнем интервале.

Покажем способ исчисления средней по данным интервального ряда на следующем примере (табл. 5.3).

Таблица 5.3

Выполнение рабочими завода установленных норм выработки

Выполнение норм выработки в процентах	Число рабочих f	Середина интервала x	Произведение xf
До 80	30	70	2 100
80–100	50	90	4 500
100–120	120	110	13 200
120–150	150	135	20 250
150–200	100	175	17 500
200 и выше	50	225	11 250
	500	–	68 800

Средняя равна: $\frac{68800}{500} = 137{,}6\%$,

Середина первого интервала находится следующим образом. Величина соседнего с ним (второго) интервала равна 20. Допуская, что такова же величина и первого интервала, считаем, что его начальной точкой служит 60. Отсюда серединой интервала будет, величина 70. Таким же путем находим середину последнего интервала, принимая его границами 200–250. Из полученных величин и исчисляется арифметическая средняя. Необходимо иметь в виду, что изложенный прием исчисления средней является вынужденным выходом из такого положения, когда нет прямых данных о конкретной величине отдельных вариантов. Этот прием основан на условном предположении, что отдельные конкретные варианты равномерно распределены внутри интервала.

Однако в действительности распределение отдельных вариантов в пределах интервала может оказаться неравномерным, и тогда середина интервала может в той или иной степени отличаться от действительной средней величины признака в данном интервале.

Это может сказаться на точности общей средней из интервального ряда.

5.8. Средняя прогрессивная и регрессивная

Средняя прогрессивная

Средняя арифметическая отражает то общее, что характерно для всей совокупности в целом. Поэтому, исчисляя среднюю, учитывают все без исключения варианты. Например, при определении средней урожайности пшеницы в хозяйстве района принимают во внимание показатели урожайности всех фермерских хозяйств – передовых, средних и отстающих. Однако в статистической и плановой работе нельзя ограничиться лишь такими средними. Наши планы должны опираться на среднепрогрессивные показатели. Средняя прогрессивная дает обобщающую характеристику не всех, а только прогрессивных показателей. Порядок вычисления средней прогрессивной рассмотрим на следующем примере (табл. 5.4).

Таблица 5.4

Расчет средней урожайности пшеницы в фермерских хозяйствах района

Номера фермерских хозяйств	Посевная площадь (га) f	Урожайность (ц с 1 га) x	Валовой сбор пшеницы (ц) xf
1	864	19,4	16 761,6
2	818	20,9	17 096,2
3	946	24,2	22 893,2
4	891	25,3	22 542,3
5	919	19,4	17 828,6
6	962	25,8	24 819,6
7	871	19,1	16 636,1
8	848	22,3	18 910,4
9	828	21,2	17 553,6
10	926	19,1	17 686,6
11	819	23,6	19 328,4
12	880	18,2	16 016,0
13	936	20,8	19 468,8
14	884	18,8	16 619,2
15	952	23,0	21 896,0
Итого	**13 344**		**286 6,6**

1. Находим среднюю урожайность пшеницы в фермерских хозяйствах района:

$$\bar{x} = \frac{\Sigma xf}{\Sigma f} = \frac{286056{,}6}{13344} = 21{,}43$$
ц с 1 га.

2. Отбираем те хозяйства (передовые), в которых показатели урожайности пшеницы выше исчисленной средней урожайности, то есть выше 21,43 ц, и находим среднюю (табл. 5.5).

$$\bar{x} = \frac{\Sigma xf}{\Sigma f} = \frac{130389{,}9}{5418} = 24{,}07$$
ц с 1 га.

Таблица 5.5

Расчет средней урожайности пшеницы в передовых фермерских хозяйствах района

Порядковый номер	Урожайность (ц с 1 га)	Посевная площадь (га)	Валовой сбор зерновых (ц)
	x	f	xf
3	24,2	946	22 893,2
4	25,3	891	22 542,3
6	25,8	962	24 819,6
8	22,3	848	18 910,4
11	23,6	819	19 328,4
15	23,0	952	21 896,0

Таким образом, средняя урожайность пшеницы в фермерских хозяйствах района равна 21,43 ц, а средняя прогрессивная – т.е. средняя урожайность в передовых хозяйствах составляла 24,07 ц. Подобным способом определяется и по фермерским хозяйствам, у которых средняя урожайность ниже общей средней (21,43 ц/га).

5.9. Основные математические свойства средней арифметической

Средняя арифметическая обладает некоторыми свойствами, имеющими большое значение для понимания сущности средней и упрощения ее расчетов.

1. *Сумма положительных и отрицательных отклонений вариант от средней равна нулю.* Это свойство иногда называют «нулевым» свойством средних:

- $\Sigma(x-\bar{x})=0$;
- $\Sigma(x-\bar{x})f=0$.

Потому что:

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

Пользуясь этим свойством, можно проверить правильность вычисления средней. Приведем пример (табл. 5.6).

Таблица 5.6

Расчет среднесуточного удоя молока и проверка «нулевого» свойства средних

Суточный удой коровы (кг)	Число коров	xf	$x-\bar{x}$	$(x-\bar{x})f$
x	f			
8	16	128	–3	$\left.\begin{matrix}-48\\-14\end{matrix}\right\}-62$
10	14	140	–1	
11	12	132	0	0
14	10	140	+3	$\left.\begin{matrix}30\\32\end{matrix}\right\}+62$
15	8	120	+4	
Итого	60	660		0

По данным таблицы средний суточный удой молока от одной коровы составил 11 кг

$$I_{\bar{x}}=\frac{\sum x_1f_1}{\sum f_1}:\frac{\sum x_0f_0}{\sum f_0}=\left(\frac{\sum x_1f_1}{\sum f_1}:\frac{\sum x_0f_1}{\sum f_1}\right)\cdot\left(\frac{\sum x_0f_1}{\sum f_1}:\frac{\sum x_0f_0}{\sum f_0}\right).$$

Сумма отклонений $(\Sigma(x-\bar{x})f=(-62)+(+62))$ равна нулю. Следовательно, средняя величина удоя (11 кг) исчислена правильно. Это правило показывает, что средняя является равнодействующей.

2. *От уменьшения или увеличения частот каждого значения* f *признака x в n раз величина средней арифметической не изменится.* Если все частоты разделить или умножить на какое-либо число, то величина средней не изменится.

$$\bar{x}=\frac{\Sigma x_i f_i}{\Sigma f_i}=\frac{\Sigma x_i \dfrac{f_i}{d}}{\Sigma \dfrac{f_i}{d_i}}=\frac{\Sigma x_i f_i d}{\Sigma f_i d}$$

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

Это свойство дает возможность частоты заменить удельными весами, называемыми частостями, а также когда частоты всех вариант одинаковы, вычислять по формуле простой средней арифметической. Указанное свойство важно тогда, когда абсолютные числа – частоты не известны, а известны лишь удельные веса, т.е. относительные величины структуры совокупности. Тогда средняя вычисляется так:

$$\bar{x}=\frac{\Sigma xd}{100},$$

если d – в %.

Используя эти свойства, можно упростить технику вычисления средней путем сокращения громоздких весов, замены абсолютных значений весов относительными (коэффициентами, процентами).

Покажем это свойство средней арифметической на примере приведения частот (f) к частостям (w), т.е. к удельным весам; в этом случае A=100 (табл. 5.7).

Таблица 5.7

Элементы расчета дневной выработки ткачих
(2-е свойство средней арифметической)

Дневная выработка ткачих (м) x	**Число ткачих** f	$\frac{f}{500}\times 100$	$x\frac{f\times 100}{500}$
46	70	14	644
50	170	34	1700
54	200	40	2160
58	60	12	696
Итого	500	100	5200

Средняя дневная выработка ткачих будет равна:

$$\bar{x} = \frac{\Sigma xd}{100} = \frac{5200}{100} = 52 \text{ м.}$$ (табл. 5.8).

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

$$\bar{x} = \frac{\Sigma x_i f_i}{\Sigma f_i} = \frac{\Sigma (x_i \pm a) f_i}{\Sigma f_i} \pm a$$

Для получения средней, вычисленной по заданому вариационному ряду, нужно к средней, рассчитанной по уменьшенным на эту величину вариантам, прибавить или вычесть эту же величину А.

Покажем эти два свойства средней арифметической на примере, допустив, что А = 54 (табл. 5.8).

Таблица 5.8

Элементы расчета средней дневной выработки ткачих

(1-е и 2-е свойство средней арифметической)

Дневная выработка ткачих (м) x	Число ткачих f	Произведение вариант на частоты xf	Расчет средней при уменьшении вариант на число		Расчет средней при увеличении вариант на число	
			x–A	(x–A)f	x+A	(x+A)f
46	70	3 220	–8	–560	100	7 000
50	170	8 500	–4	–680	104	17 680
54	200	10 800	0	0	108	21 600
58	60	3 480	+4	+240	112	6720
Итого	500	26 000	–	–1000	–	53 000

$$\bar{x} = \frac{26000}{500} = 52;$$

$$\bar{x} = \frac{-1000}{500} + 54 = -2 + 54 = 52; \ \bar{x} = \frac{53000}{500} - 54 = 106 - 54 = 52.$$

Средняя дневная выработка ткачих, вычисленная непосредственно по заданным вариантам будет равна:

$$\bar{x} = \frac{\Sigma x_i f_i}{\Sigma f_i} = \frac{26000}{500} = 52 \ \text{м};$$

этой же величине равны средние, исчисленные с использованием первого и второго свойств средней арифметической (табл. 5.6 и 5.7).

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

$$\bar{x} = \frac{\Sigma x_i f_i}{\Sigma f_i} = \frac{\Sigma \frac{x_i}{A} f_i}{\Sigma f_i} \cdot A = \frac{\Sigma A x_i f_i}{\Sigma f_i} : A.$$

Для получения средней, вычисленной по заданному вариационному ряду, нужно среднюю, рассчитанную по деленным на эту величину вариантам, умножить или разделить на это же число.

5.10. Моментный способ исчисления средней

Изложенные выше свойства средней арифметической позволяют во многих случаях упростить ее расчеты: можно из всех значений признака вычесть произвольную постоянную величину, разность сократить на общий множитель, а затем исчисленную среднюю умножить на общий множитель и прибавить произвольную постоянную величину.

Формула средней арифметической взвешенной получит следующий вид:

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

Средняя m_1 из значения $x - A/i$ называется *моментом первого порядка,* а способ вычисления средней – *способом моментов.*

Иногда его также называют *способом отсчета от условного нуля.* Рассмотрим методику расчета по данным табл. 5.9.

Таблица 5.9

Распределение предприятий региона по объему товарооборота

Группы предприятий по объему товарооборота, млн сумм, x	Число, предприятий, f	Середина интервала, x'	x'f	x'–A	$\left(\frac{x'-A}{i}\right)$	$\left(\frac{x'-A}{i}\right)f$
1	2	3	4	5	6	7
До 400	9	350	3 150	–200	–2	–18
400–500	12	450	5 400	–100	–1	–12
500–600	8	550	4 400	0	0	0
600–700	9	650	5 850	100	I	9
Свыше 700	2	750	1 500	200	о	4
Итого	40	–	20 300	–	–	–17

Для вычисления средней величины надо в каждом варианте определить серединное значение x', после чего произвести взвешивание обычным порядком $x'f$. В закрытом интервале серединное значение определяется как полусумма значений нижней и *верхней* границ (гр 3. табл. 5.9). Иногда задача исчисления средней по величинам интервального ряда осложняется тем, что неизвестны крайние границы начального и конечного интервалов. В этом случае предполагается, что расстояние между границами данного интервала такое же, как и в соседнем интервале. В нашем примере:

в качестве A принимается одна из центральных вариант ряда $(x' = 500)$, если ряд имеет нечетное число признаков. При четном числе признаков берется среднее значение из двух вариант с наибольшей частотой. В качестве i берется общий (наибольший) делитель индивидуальных отклонений. В интервальном ряду в качестве i наиболее целесообразно использовать величину интервала (i–100). Тогда средний объем товарооборота предприятия составит:

$$\bar{x} = \frac{\Sigma\left(\frac{x-A}{i}\right)f}{\Sigma f} + A = \frac{-17}{40}\cdot 100 + 550 = 507{,}5$$

млн сум.

Необходимость использования упрощенных методов расчета в настоящее время ограничена, так как все более широко внедряются ЭВМ, позволяющее производить расчеты по индивидуальным значениям.

Необходимо отметить, что изложенный прием исчисления средней является вынужденным в случае, когда нет прямых данных о конкретной величине отдельных вариантов. Этот прием основан на предположении, что отдельные конкретные варианты равномерно распределены внутри интервала.

5.11. Средняя гармоническая

Учитывая, что статистические средние всегда выражают качественные свойства изучаемых общественных процессов и явлений, важно правильно выбрать форму средней, исходя из взаимосвязи явлений и их признаков. Средняя гармоническая – это величина, обратная средней арифметической, когда $z = -1$. Когда статистическая информация не содержит, частот по отдельным вариантам совокупности, а представлена как их произведение, применяется формула *средней гармонической взвешенной.*

Так. например, расчет средней цены выражается отношением

$$\textit{Средняя цена} = \frac{\textit{Сумма реализации}}{\textit{Количество реализованных единиц}}$$

Таблица 5.10

Порядок применения средней гармонической формулы

Город	Цена, тыс. сум, x_i	Сумма реализации, тыс. сум w_i	Частоты $f' = \frac{w_i}{x_i}$
А	30	600	20
Б	20	1000	50
В	35	350	10
Итого		1 950	80

Величина суммы реализации, т.е. показателя, который находится в числителе исходного отношения, известна. Для определения неизвестной величины – количества реализованных единиц – нужно отдельно по каждому виду товара разделить сумму реализации на цену (табл. 5.10).

Расчет примет следующий вид:

$$\bar{x} = \frac{\Sigma w_i}{\Sigma \frac{w_i}{x_i}} = \frac{600 + 100 + 350}{\frac{600}{30} + \frac{1000}{20} + \frac{350}{35}} = 24{,}3$$

тыс. сум

$$\bar{x} = \frac{\Sigma w_i}{\Sigma \frac{w_i}{x_i}} = \frac{w_1 + w_2 + \ldots + w_n}{\frac{w_1}{x_1} + \frac{w_2}{x_2} + \ldots + \frac{w_n}{x_n}}.$$

При определении средней цены, используя невзвешенную среднюю арифметическую, получим среднюю, которая не отражает объема реализации, т.е. нереальна.

$$\bar{x} = \frac{30 + 20 + 35}{3} = 28$$

тыс. сум.

Как видно, средняя гармоническая является превращенной формой арифметической средней. Вместо гармонической всегда можно рассчитать среднюю арифметическую, но для этого сначала нужно определить веса отдельных значений признака.

В том случае, если объемы явлений, т.е. произведения, по каждому признаку равны, применяется средняя гармоническая (простая).

Пример. Две автомашины прошли один и тот же путь: одна со скоростью 60 км/ч, а вторая – 80 км/ч, тогда средняя скорость составит:

$$\bar{x} = \frac{1+1}{\frac{1}{60} + \frac{1}{80}} = \frac{2}{\frac{80+60}{4800}} = \frac{9600}{140} = 68{,}6$$

км/ч.

тогда

$$\bar{x} = \frac{n}{\Sigma \frac{1}{x}},$$

где $\Sigma\frac{1}{x}$ – сумма обратных значений варианта; n – число вариантов.

5.12. Другие виды средних величин

1. Средняя геометрическая

Если имеется n коэффициентов роста, то формула среднего коэффициента роста будет иметь вид

$$\overline{K} = \sqrt[n]{K_1 \cdot K_2 \cdot \ldots \cdot K_n}.$$

Это формула средней геометрической, которую можно сформулировать следующим образом:

Средняя геометрическая равна корню степени n из произведения коэффициентов роста, характеризующих отношение величины каждого последующего периода к величине предыдущего.

Средний коэффициент роста можно определить также по данным последнего и первого уровней ряда. Если первый уровень ряда обозначить y_1, а последний – y_n, то

$$\overline{K} = \sqrt[n-1]{\frac{Y_n}{Y_1}}, \text{ так как } \sqrt[n]{\frac{Y_2}{Y_1} \cdot \frac{Y_3}{Y_2} \cdot \frac{Y_4}{Y_3} \cdot \ldots \cdot \frac{Y_n}{Y_{n-1}}} = \sqrt[n-1]{\frac{Y_n}{Y_1}},$$

где n – число лет, а не коэффициентов.

Приведенные формулы идентичны, но одна применяется в тех случаях, когда имеются текущие коэффициенты или темпы роста, а вторая – когда имеются абсолютные значения начального и конечного уровней ряда.

Вычисление средней геометрической связано с извлечением корней. Это, как известно, можно сделать при помощи таблицы логарифмов. Так,

$$\overline{K} = \sqrt[n-1]{\frac{Y_n}{Y_1}}, \text{ а } \lg\overline{K} = \frac{\lg n}{n}(n-1).$$

Этой средней удобно пользоваться, когда уделяется внимание не абсолютным разностям, а отношениям двух чисел.

Поэтому средняя геометрическая используется в расчетах, среднегодовых темпов роста. Об этом более подробно см. в гл. 9.

2. Средняя квадратическая

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

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

$$\bar{x} = \sqrt{\frac{x_1^2 + x_2^2 + x_3^2 + \ldots + x_n^2}{n}} = \sqrt{\frac{\Sigma x^2}{n}}.$$

Средняя квадратическая взвешенная равна

$$\bar{x} = \sqrt{\frac{\Sigma x^2 f}{\Sigma f}},$$

где f –веса.

3. Средняя хронологическая

Для общей характеристики какого-либо явления за определенный период рассчитывают средний уровень из всех членов динамического ряда.

Способы его расчета зависят от вида динамического ряда. Для интервальных рядов средняя рассчитывается по формуле средней арифметической, причем при равных интервалах применяется средняя арифметическая простая, а при неравных – средняя арифметическая взвешенная.

Для нахождения средних значений моментного рада применяют среднюю хронологическую:

$$\bar{y} = \frac{\frac{y_1}{2} + y_2 + y_3 + \ldots + y_{n-1} + \frac{y_n}{2}}{n-1}$$

Средняя хронологическая моментного ряда равна сумме всех уровней ряда, поделенной на число членов ряда без *одного, причем первый и последний члены ряда берутся в половинном размере.*

Если интервалы между периодами не равны, то применяется средняя арифметическая взвешенная, а в качестве весов берутся отрезки времени между датами, к которым относятся парные средние смежных значений уровня (более подробно в гл. 9).

5.13. Структурные средние величины

Для характеристики структуры совокупности применяются особые показатели, которые можно назвать *структурными средними.*

К таким показателям относятся мода и медиана:

Мода (Мо)

Модой (Мо) называется чаще всего встречающийся вариант, или модой называется то значение признака, которое соответствует максимальной точке теоретической кривой распределений.

Мода представляет наиболее часто встречающееся или типичное значение. Мода широко используется в коммерческой практике при изучении покупательского спроса (при определении размеров одежды и обуви, которые пользуются широким спросом), регистрации цен.

В дискретном ряду мода – это вариант с наибольшей частотой. Например, по приведенным ниже данным наибольшим спросом обуви пользуется размер 37 (табл. 5.11).

Таблица 5.11

Размер обуви	Число купленных пар
34	2
35	10
36	20
37	88
38	19
39	9
40	1

Таблица 5.12

Стаж (лет)	Число работников
До 2	4
2–4	23
4–6	20
6–8	35
8–10	11
Свыше 10	7

В интервальном вариационном ряду модой приближённо считают центральный вариант так называемого модального интервала, т.е. того интервала, который имеет наибольшую частоту (частость). В пределах интервала надо найти то значение признака, которое является модой.

Решение вопроса состоит в том, чтобы в качестве моды выявить середину модального интервала. Такое решение будет правильней лишь в случае полной симметричности распределения либо тогда, когда интервалы, соседние с модальными, мало отличаются друг от друга по числу случаев. В противном случае середина модального интервала не может рассматриваться, как мода. Конкретное значение моды для интервального ряда определяется формулой

$$Mo = x_{Mo} + i_{Mo}\frac{(f_{Mo} - f_{Mo-1})}{(f_{Mo} - f_{Mo-1})(f_{Mo} - f_{Mo+1})}$$

где x_{Mo} – нижняя граница модального интервала; i_{Mo} – величина модального интервала; f_{Mo} – частота, соответствующая модальному интервалу; f_{Mo-1} – частота, предшествующая модальному интервалу; f_{Mo+1} – частота интервала, следующего за модальным.

Эта формула основана на предположении, что расстояния от нижней границы модального интервала до моды и от моды до верхней границы модального интервала прямо пропорциональны разностям между численностями модального интервала и прилегающих к нему. В нашем примере (табл. 5.12) модальным интервалом величины стажа работников торгового предприятия будут 6–8 лет, а модой продолжительности стажа – 6,77 года.

$$Mo = 6 + 2\frac{35-20}{(35-20)(35-11)} = 6,77$$ года.

Мода всегда бывает несколько неопределенной, так как она зависит от величины групп, от точного положения границ групп.

Медиана (Me) Это величина, которая делит численность упорядоченного вариационного ряда на две равные части одна часть имеет значения варьирующего признака меньшие, чем средний вариант, а другая – большие. Понятие медианы легко уяснить из следующего примера. Для ранжированного ряда (т.е. построенного в порядке возрастания или убывания индивидуальных величин) с нечетным числом членов медианой является вариант, расположенный в центре ряда.

Например, в ранжированных данных о стаже работы семи продавцов – 1, 2, 2, 3, 5, 7, 10 лет – медианой является четвертый вариант – 3 года. Для ранжированного ряда с четным числом членов (индивидуальных величин) медианой будет средняя арифметическая из двух смежных вариантов. Если в бригаде продавцов из шести человек распределение по стажу работы было таким: 1, 3, 4, 5, 7, 9 лет, то медианой будет значение, равное: (4 + 5) : 2= 4,5 года, т.е.

$$Me = \frac{x_{Me} + x_{Me+1}}{2}.$$

В интервальном вариационном ряду порядок нахождения медианы следующий: располагаем индивидуальные значения признака по ранжиру; определяем для данного ранжированного ряда накопленные частоты; по данным о накопленных частотах находим медианный интервал. Медиана делит численность ряда пополам, следовательно, она там, где накопленная частота составляет половину или больше половины всей суммы частот, а предыдущая (накопленная) частота меньше половины численности совокупности.

Иллюстрируем технику нахождения медианы дискретного вариационного ряда на примере. Воспользуемся для этого данными о распределении рабочих нефтеперерабатывающей промышленности по тарифным разрядам (табл. 5.13).

Полусумма относительных численностей этого ряда равна 50%. Сумма накопленных численностей (процентов) большая 50% (гр.3 5.13 табл.), соответствует варианту 3. Следовательно, третий тарифный разряд является медианой. В том случае, если бы одна из накопленных сумм оказалась точно равной половине объема ряда, медиана была бы равна полусумме двух соседних вариантов, делящих численность ряда на две равные части.

Таблица 5.13

Распределение рабочих нефтеперерабатывающей промышленности по тарифным разрядам (на 2 августа 2019 г.) **и расчет накопленных численностей для определения медианы**

Тарифный разряд	**Численность рабочих (в процентах к итогу)**	**Суммы накопленных процентов**
А	1	2
1	5,4	5,4
2	15,9	21,3 (5,4+15,9)
3	30,7	52,0 (21,3+30,7)
4	27,4	–
5	16,3	–
6	4,3	–
Итого	100,0	–

Так, если бы оказалось, что в рассматриваемом примере сумма численности двух первых вариантов точно равнялась 50%, то в этом случае медиана была бы равна $2{,}5\left(\frac{2+3}{2}\right)$.

Медиана интервальных вариационных рядов определяется в указанной ниже последовательности:

1) находят медианный интервал, т.е. интервал, делящий численность ряда на две равные части (определяется так же, как и в случае определения медианы дискретного вариационного ряда);

2) рассчитывают конкретное значение медианы по формуле:

$$Me = x_0 + d\frac{\frac{1}{2}\Sigma f - S_{m-1}}{f_m},$$

где *Me* – медиана;

x_0 – начальная граница медианного интервала;

d – величина медианного интервала;

Σf – объем ряда;

S_{m-1} – сумма накопленных частот интервалов, предшествующих медианному;

f_m – частота медианного интервала.

Рассчитаем в качестве примера медиану по данным о распределении машинистов по стажу работы на локомотивах.

Для определения медианного интервала подсчитаем суммы накопленных частот (см. гр. 3) следующей таблицы (табл. 5.14).

Таблица 5.14

Распределение машинистов депо по стажу работы на локомотивах (на 1 июня 2019 г.) **и расчет накопленных численностей**

Стаж работы (лет)	Численность машинистов	Накопленные частоты
До 3	12	12
3–6	18	30
6–9	27	57
9–12	32	89
12–15	24	–
15–18	16	–
18–21	5	–
	134	

Медианному интервалу соответствует частота *fm*, удовлетворяющая условию

$$S_{m-1} < \frac{1}{2}\Sigma f < S_{m-1} + f_m$$

В рассматриваемом примере этому условию удовлетворяет частота интервала 9–12, равная 32.

Отсюда следует, что $x_0 = 9$, $d = 3$, $S_{m-1} = 57$, $f_m = 32$. Подставив эти значения в приведенную выше формулу; получим:

$$M_e = 9 + 3\frac{67-57}{32} = 9{,}9$$ года.

Из приведенных расчетов видно, что в основе определения конкретного значения медианы интервальных вариационных рядов лежит принцип пропорциональности, а именно: к начальной границе медианного интервала (x_0) прибавляется такая часть величины этого интервала (d), какая часть взята от его численности *(fm)* для получения половины объема ряда $\left(\frac{1}{2}\Sigma f\right)$, в рассматриваемом примере половина объема ряда составляет 67, а сумма частот первых трех интервалов – 57. Следовательно, для того чтобы получить половину объема ряда, надо взять 10 единиц из 32, составляющих численность медианного интервала. В соответствии с этим к начальной границе медианного интервала, равной 9, и прибавлено $\frac{10}{32}$ от величины этого интервала, равной трем.

Если предполагать, что внутри медианного интервала нарастание или убывание изучаемого признака происходит по прямой равномерно, то формула медианы в интервальном ряду распределения будет иметь следующий вид:

$$M_e = x_{Me} + i_{Me}\frac{\frac{\Sigma f}{2} - S_{Me-1}}{f_{Me}}$$

где x_{Me} – нижняя граница медианного интервала; i_{Me} – величина медианного интервала; $\Sigma f/2$ – полусумма частот ряда; $S_{Me\text{-}1}$ – сумма накопленных *частот,* предшествующих медианному интервалу; f_{Me} – частота медианного интервала.

Медиана ряда наблюдений может быть очень далека от его типичной величины и в действительности может не приближаться ни к одному из наблюдаемых объектов. Но, поскольку медиана является серединным (центральным) значением, это делает ее смысл вполне ясным. Медиана по своему положению более определенна, чем мода.

Медиана находит практическое применение вследствие особого свойства – сумма абсолютных отклонений членов ряда от медианы есть величина наименьшая Σ *(x* – Me) = min (табл. 5.15).

Таблица 5.15

№ п/п	Расположение магазинов от базы снабжения, км (x)	Отклонения от среднего значения $(x-\bar{x})$	Отклонения от медианного значения (x – Me)
1	2	3	2
2	3	2	1
3	4	1	0
4	6	1	2
5	10	5	6
Итого	25	13	11

$\bar{x} = 25 : 5 = 5$ км; Me = 4 км.

Вышеназванное свойство Me находит широкое практическое применение в маркетинговой деятельности.

Величины, приходящиеся на одной четверти и на трех четвертях расстояния от начала ряда, называются *квартилями,* на одной десятой – *децилями,* на одной сотой – *процентилями.*

Другими словами:

Квартель - значение признака, делящее совокупность на *4 равные части.*

Квинтель – значение признака, делящее совокупность на *5 равных частей.*

Децель – значение признака, делящее совокупность на *10 равных частей.*

При статистическом изучении совокупности правильно выбранная средняя обладает следующими свойствами: если в индивидуальном признаке явления есть какая-либо типичность, то средняя ее обнаруживает, но она учитывает и влияние крайних значений.

Если:

$$\bar{x} = Me = Mo$$

совпадают, то данная группа симметрична.

Если:

$$Me < \bar{x}$$

при исчисленной группе с очень высокими числами.

Если

$$\bar{x} < Me,$$

то нет очень больших чисел и данные концентрируются.

Если совокупность неоднородна, то мода трудно определяется. $\bar{x} < Me$, если имеется немногочисленная группа с высокими числами и *Мо* отчетливо выражена при однородности группы.

В статистической практике мода обычно используется в тех случаях, когда невозможно исчислить средние размеры признака по обычным формулам. Практически, например, невозможно рассчитать средние цены, по которым реализуются товары на дехканских рынках, ибо для этого надо было бы произвести не только регистрацию различных цен, по которым в каждом отдельном случае продавались товары, но и учесть объемы реализации во всех этих случаях. Поэтому регистрации подвергаются модальные цены, т.е. цены, которые чаще всего встречаются на данном рынке в период наиболее оживленной торговли.

Мода и медиана наряду со средней арифметической используются в анализе вариационных рядов (для характеристики их асимметрии).

Интеллектуальный тренинг

1. Что представляет собой средняя величина и в чем состоит ее определяющее свойство?

2. Говорят, что изучение работы лучших и худших предприятий – важный резерв развития общества ибо при рыночной экономике «сила примера впервые получает возможность оказать свое массовое действие». Так ли это? Дайте разяснение.

3. Кому принажлежат слова «вполне естественно, что в обществе разрозненных товаропроизводителей, связанных

лишь рынком, закономерность не может проявляться иначе как в средней, общественной, массовой закономерности при взаимопогашении индивидуальных уклонений в ту или другую сторону».

4. По мнению известного английского экономиста-статистика А. Боули, функция средней заключается именно в том, чтобы множество величин, которые человеческий ум не в состоянии охватить, свести к немногим средним показателям. «Ум не в состоянии, – пишет он, – сразу охватить величины миллионов статистических данных, они должны быть сгруппированы, упрощены, приведены к средним».[1]

Подобной точки зрения на природу средней придерживаются многие современные статистики. Например, французский профессор Р. Дюма по существу повторяет приведенные выше слова А. Боули, когда он связывает вопрос о необходимости «вывести среднюю» с невозможностью сделать выводы на основе массовых данных, сгруппированных и представленных в табличной форме.[2] Поясните сущность их теории.

5. «Соединение крупных и мелких заведений вместе дает совершенно фиктивные «средние» цифры, не дающие никакого понятия о действительности, затушевывающие кардинальные различия, изображающие в однородном нечто совершенно разнородное, разносоставное». Кому относятся эти слова? Поясните сущность высказанного.

6. Один теоретик сказал обработка статистических данных земскими статистиками сводится «к одному сплошному и невероятному злоупотреблению средними величинами» и с тонкой иронией отмечал, что «стоит только пользоваться всегда и исключительно «средними» данными о крестьянском хозяйстве, – и все «превратные идеи» о разложении крестьянства окажутся раз навсегда изгнанными». Кем сказаны эти слова? Поясните сующность высказанного.

7. Что такое осредняемый признак?

---

[1] Боули А. Элементы статистики. Ч. 1. М.: Госиздат, 1930. С. 125.

[2] Дюма Р. Предприятие и статистика. М.: Госстатиздат, 1958. С.48-49.

8. В чем заключается связь метода группировок и метода средних?

9. Какие формы средней вы знаете?

10. В каких случаях может применяться простая (невзвешенная) средняя?

11. Что такое средняя гармоническая?

12. На чем основывается выбор вида средней?

13. Что характеризуют мода и медиана?

14. Как подсчитываются мода и медиана для дискретного ряда с четным числом уровней ряда?

15. Как определяются мода и медиана дискретного ряда с с четным числом уровней?

16. Как обосновывается выбор весов при расчете взвешенных средних?

17. Для каких целей используется формула средней геометрической?

18. В чем различие между степенными и структурными средними?

19. Какие условия правильного применения средних Вы знаете?

20. Можно ли использовать относительные величины в качестве весов?

21. Как исчисляется средняя величина в интервальном ряду?

22. Что Вы понимаете под средней прогрессивной и регрессивной?

23. Что означает «нулевое» свойство средних?

24. Как исчисляется средняя моментным способом?

25. Что означает следующие равенства: $\bar{x} = Me = Mo$?

26. Что означает следующая разность: $\bar{x} - M_0$?

27. В каких случаях рассчитывается следующая формула и как она называется?

$$As = \frac{\overset{u}{\bar{x}} - M_0}{\sigma}$$

Использованная и рекомендуемая специальная литература

1. Гранков В.П. Средние величины в статистике. М.: Госстатиздат, 1957.

2. Джини К. Средние величины. М.: Статистика, 1970.

3. Пасхавер И.С. Средние величины в статистике. М.: Статистика, 1979.

4. Рябушкин Т.В. Средние в статистике. М.: Госстатиздат, 1954.

Глава VI
ПОКАЗАТЕЛИ ВАРИАЦИИ

Дорожная карта

6.1. Общее понятие о показателях вариации

Определение средней величины еще недостаточно для изучения варьирующего признака, так как в средней величине погашена колеблемость изучаемого признака у отдельных единиц, погашена вариация. В ходе анализа средних величин возникает вопрос о степени колеблемости, степени вариации, скрывающейся за средней величиной. Покажем это на примере табл. 6.1.

Таблица 6.1

Распределение рабочих, производящих однородные изделия, по величине сменной выработки

Выработка изделий в смену (шт.)	Число рабочих в смену	
	завод № 1	завод № 2
40	20	–
41	40	30
42	80	140
43	40	30
44	20	–
Итого	200	200

Вычислим среднюю сменную выработку изделий одним рабочим на каждом заводе

Завод № 1:

$$\bar{x} = \frac{\Sigma xf}{\Sigma f} = \frac{40\times 20 + 41\times 40 + 42\times 80 + 43\times 40 + 44\times 20}{20+40+80+40+20} = \frac{8400}{200} = 42$$ изделия.

Завод № 2:

$$\bar{x} = \frac{\Sigma xf}{\Sigma f} = \frac{41\times 30 + 42\times 140 + 43\times 30}{30+140+30} = \frac{8400}{200} = 42$$ изделия.

На обоих заводах средняя выработка рабочего в смену была одинаковой. Однако распределение рабочих по величине сменной выработки, колеблемость этого признака на каждом заводе были различны.

Как видно из табл. 6.1, колеблемость сменной выработки на обоих заводах резко различается: на первом заводе были рабочие со сменной выработкой по 40 и 44 изделия, а на втором заводе рабочих с такой выработкой не было. На втором заводе 140 рабочих из 200 вырабатывали по 42 изделия за смену, а на первом заводе такую сменную выработку имели также из 200 рабочих только 80. Таким образом, на втором заводе доля групп рабочих, вырабатывавших в смену 42 изделия, значительно больше, чем на первом (70% и 40%).

Все это говорит о том, что средние величины в ряде случаев необходимо дополнять показателями, характеризующими колеблемость – вариацию признака.

Вариация признака

Вариация признака (от лат. variatio – «изменение, колеблемость, различие») – различие индивидуальных значений признака внутри изучаемой совокупности, возникающее в результате того, что индивидуальные значения складываются под совокупным влиянием разнообразных факторов, которые по-разному сочетаются в каждом отдельном случае. Эти колебания обусловлены тем, что на величину признака у отдельных единиц изучаемого объекта оказывают влияние различные факторы, часто действующие в противоположных направлениях. Так, величина урожайности той или иной сельскохозяйственной культуры зависит от качества почвы, способов ее обработки, качества семян, количества внесенных удобрений; метеорологических условий, ухода за посевами, своевременной уборки урожая и ряда других факторов. В результате совместного действия всех этих факторов и складывается величина урожайности на отдельных участках посева.

Вариация (колеблемость) признаков может быть большей или меньшей. Это наглядно видно из приводимых ниже данных табл. 6.2.

Таблица 6.2

Распределение посевной площади фермерских хозяйств по величине урожайности озимой пшеницы

Урожайность (в ц с га)	Посевная площадь (в га)	
	первое отделение	второе отделение
12–14	12	–
14–16	25	–
16-18	43	68
18–20	64	67
20–22	30	34
22–24	16	36
Σ	190	205

Величина урожайности озимой пшеницы в первом отделении колеблется в пределах 12–24 ц, в то время как во втором отделении колеблемость не выходит за пределы 16–24 ц.

Не всякие различия принято называть вариацией. Под вариацией в статистике понимают такие количественные изменения величины исследуемого признака в пределах однородной совокупности, которые обусловлены перекрещивающимся влиянием действия различных факторов. **Например, работники предприятия одновременно различаются**:

✓ по уровню доходов;

✓ по занимаемой должности;

✓ по затратам времени на работу и др.

Вариации присущи явлениям природы и общества. **Вариация широко применяется в практике:**

➢ для оценки ритмичности работы промышленных предприятий;

➢ для определения устойчивости урожайности сельскохозяйственных культур тех или иных сортов;

➢ для разработки показателей, характеризующих социально-экономические явления и процессы.

Различают следующие вариаии признака:

- случайную;
- систематическую;
- общую.

Систематическая вариация – вариация, возникающая вследствие действия существенных факторов и носящая систематический характер (последовательное изменение вариантов признака в определенном направлении).

Случайная вариация – вариация, порождаемая случайными факторами. Здесь все изменения носят хаотичный характер, так как не наблюдается взаимосвязь факторов с единицами изучаемой совокупности.

Общая вариация – вариация, порождаемая всеми без исключения факторами. Это итог объединения систематической и случайной вариаций.

Вариация альтернативного признака

Вариация альтернативного признака – степень колебания, которой **обладают одни единицы совокупности и не обладают другие.** *Пример*: брак продукции, ученая степень преподавателя вуза, наличие красного диплома у студентов и др.

6.2. Показатели вариации и способы их расчета

Поскольку вариация признаков может быть различной по своим размерам, то возникает необходимость к ее *измерению.* В статистике для измерения вариации признаков применяют систему абсолютных и относительных показателей. К абсолютным показателям вариации относятся:

1) размах вариации $R = x_{max} - x_{min}$;
2) среднее линейное отклонение $\bar{d}$;
3) дисперсия σ^2;
4) среднеквадратическое отклонение $\sigma = \sqrt{\sigma^2}$

Эти показатели (кроме дисперсии) измеряются в тех же единицах, что и сам признак: в тоннах, метрах, секундах, рублях, сумах и т.п. К относительным показателям вариации относятся:

1) Коэффициент осцилляции $K_{осцилляции} = \frac{R}{\bar{x}}$;
2) линейный коэффициент вариации
$K_{линейной\ вариации} = \frac{\bar{d}}{\bar{x}}$;
3) простой коэффициент вариации $v = \frac{\sigma}{\bar{x}}$.

Эти показатели выражаются в процентах или относительных величинах.

6.3. Размах вариации

1. Размах вариации **(амплитуда колебаний)** показывает, насколько велико различие между единицами совокупности, находящимися на концах ранжированного ряда.

Размах вариации определяется как разность между максимальным и минимальным значениями признака в изучаемой совокупности.

Размах вариации – это **абсолютное отклонение,** сохраняющее размерность изучаемого признака.

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

В нашем примере размах вариации сменной выработки рабочих на первом заводе будет равен 4 изделиям (44–40), а на втором заводе – 2 изделиям (43–41). Таким образом, размах вариации на первом заводе в 2 раза больше, чем во втором.

Особенность показателя размаха вариации заключается в том, что он зависит лишь от двух крайних значений признака, так как при определении величины размаха вариации учитываются только два крайних значения признака, колеблемость же значений признака, заключенных между его крайними значениями, и распространенность (частоты) различных значений признака, в этом показателе не находит отражения.

Другими словами недостаток размаха вариации как показателя колеблемости заключается в том, что он не отражает всех значений признака, промежуточных между минимальным и максимальным значениями.

Не учитывает он и частот.

Поэтому размах вариации недостаточно правильно характеризует колеблемость признака.

В связи с тем, что каждое индивидуальное значение признака отклоняется от средней на определенную величину, очевидно, что мерой вариации может служить средняя из отклонений каждой отдельной варианты от их средней.

Такими показателями являются среднее линейное отклонение, дисперсия и среднее квадратическое отклонение.

6.4. Среднее линейное отклонение

Среднее линейное отклонение представляет собой среднюю из абсолютных значений отклонений отдельных вариант от их средней. Так как алгебраическая сумма отклонений индивидуальных значений признака от средней равна нулю, т.е. $\Sigma(x - x) = 0$ (второе свойство средней арифметической), при исчислении среднего линейного отклонения принимаются во внимание только абсолютные значения отклонений без учета знаков («+» или «–»). Среднее линейное отклонение рассчитывается по формуле средней арифметической простой:

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

или средней арифметической взвешенной:

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

Среднее линейное отклонение обладает значительным преимуществом перед размахом вариации в отношении полноты характеристики колеблемости признака. Среднее линейное отклонение как меру вариации признака применяют в статистической практике редко. Во многих случаях этот показатель не устанавливает степень рассеивания.

Как доказывается в курсах математической статистики, и среднее линейное отклонение является недостаточно точным показателем вариации. Более точный результат дает исчисление дисперсии и среднеквадратического отклонения.

6.5. Дисперсия и среднеквадратическое отклонение

Дисперсией *называется средний квадрат отклонений* **индивидуальных** *значений признака от средней величины. Дисперсия обозначается греческой буквой* σ *(сигма) в квадрате и равна*

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

При равенстве весов или когда они равны 1

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

Дисперсия имеет большое значение в статистическом анализе. Однако ее применение как меры вариации в ряде случаев бывает не совсем удобным, потому что размерность дисперсии равна квадрату размерности изучаемого признака. В таких случаях для измерения вариации признака вычисляют среднее квадратическое отклонение.

Среднее квадратическое отклонение равно корню квадратному из суммы квадратов отклонений индивидуальных значений признака от их средней, т.е. из дисперсии:

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

А при равенстве весов, или когда они равны 1

$$\sigma = \sqrt{\frac{\Sigma(x-\bar{x})^2}{n}} \text{ или } \sigma = \sqrt{\sigma^2}.$$

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

Порядок вычисления простого среднего линейного отклонения и простого среднего квадратического отклонения рассмотрим исходя из условий вышеприведенного примера. Для этого построим и заполним табл. 6.3.

Расчет простого среднего линейного отклонения по данным о сменной выработке изделий рабочими на первом заводе:

1. Средняя выработка изделий рабочим за смену нам уже известна; она равна 42 шт $\left(\bar{x} = \frac{\Sigma x}{n} = 210:5 = 42\right)$.

2. Находим отклонение каждого варианта от величины средней выработки рабочим:

- для первой группы $x_1 - \bar{x} = 40 - 42 = -2$ шт.;
- для второй группы $x_2 - \bar{x} = 41 - 42 = -1$ шт.;
- для третьей группы $x_3 - \bar{x} = 42 - 42 = 0$;
- для четвертой группы $x_4 - \bar{x} = 43 - 42 = +1$ шт.;
- для пятой группы $x_5 - \bar{x} = 44 - 42 = +2$ шт.

Таблица 6.3

Элементы расчета показателей вариации сменной выработки рабочих

Завод № 1				Завод № 2			
Номера групп рабочих	Сменная выработка изделий (шт.)	отклонение от средней	Квадрат отклонения от средней	Номера групп рабочих	Сменная выработка изделий (шт.)	отклонение от средней	Квадрат отклонения от средней
	x	$x-\bar{x}$	$(x-\bar{x})^2$		x	$x-\bar{x}$	$(x-\bar{x})^2$
1	40	–2	4	1	41	-1	1
2	41	– 1	1	2	42	0	0
3	42	0	0	3	43	+1	1
4	43	+ 1	1				
5	44	+2	4				
Итого	–	–	10	Итого	–	–	2
	$\bar{x}=42$		$\Sigma(x-\bar{x})^2$		$\bar{x}=42$		$\Sigma(x-\bar{x})^2$

Как видно, сумма положительных и отрицательных отклонений каждого варианта от средней величины сменной выработки изделий равна нулю (–2–1 + 1+2 – 0).

1. Находим сумму всех абсолютных отклонений (без учета знаков «плюс» и «минус»):

$$\sum|x-\bar{x}| = 2 + 1 + 1 + 2 = 6 \text{ шт.}$$

2. Делим сумму линейных отклонений на число вариант и получаем значение среднего линейного отклонения (**d**) на первом заводе:

$$\bar{d} = \frac{\sum|x-\bar{x}|}{n} = \frac{6}{5} = 1{,}2 \text{ шт.}$$

Соответственно, на втором заводе простое среднее линейное отклонение будет равно:

$$\bar{d} = \frac{2}{3} = 0{,}67 \text{ шт.}$$

Таким образом, на первом заводе среднее линейное отклонение в 1,8 (1,2:0,67) раза больше, чем на втором заводе, что свидетельствует о значительно большей вариации производительности труда на первом заводе по сравнению со вторым заводом.

Покажем теперь расчет простого среднего квадратического отклонения на первом заводе по тем же данным о сменной выработке рабочих.

1. Средняя выработка (42 шт.) и линейные отклонения от нее нам уже известны. Возведем в квадрат отклонения по каждому варианту:

- для первой группы $(-2)^2=4$ шт.;
- для второй группы $(-1)^2=1$ шт.;
- для третьей группы 0;
- для четвертой группы $(+1)^2 = 1$ шт.;
- для пятой группы $(+ 2)^2 = 4$ шт.

2. Находим сумму квадратов отклонений всех вариантов:

$$\sum(x-\bar{x})^2 = 4 + 1 + 1 + 4 = 10 \text{ шт.}$$

3. Делим сумму квадратов отклонений на число вариантов. Эта средняя из квадратов отклонений называется *дисперсией* или средним квадратом отклонений и обозначается симоволом σ^2 (сигма):

$$\sigma^2 = \frac{\sum(x-\bar{x})^2}{n} = \frac{10}{5} = 2 \text{ шт.}$$

Корень квадратный из дисперсии называется *средним квадратическим отклонением* и обозначается символом σ; по первому заводу оно составит:

$$\sigma=\sqrt{\sigma^2}=\sqrt{\frac{\Sigma(x-\bar{x})^2}{n}}=\sqrt{\frac{10}{5}}=\sqrt{2}=1{,}42 \text{ шт.}$$

Соответственно, среднее квадратическое отклонение на втором заводе будет равно:

$$\sigma=\sqrt{\frac{2}{3}}=0{,}82 \text{ шт.}$$

Среднее квадратическое отклонение на первом заводе в 1,7 (1,42:0,82) раза больше, чем на втором, что опять-таки свидетельствует о более сильной вариации уровней производительности труда на первом заводе по сравнению со вторым заводом.

Найденные нами средние отклонения (линейное и квадратическое) являются простыми, так как распространенность каждого из вариантов принималось одинаковой. В действительности же, как видно из исходной для данного примера таблицы, разные варианты имеют различную распространенность, различные веса (см. табл. 6.3). Поэтому надо рассчитать средние отклонения (линейное и квадратическое) взвешенные. Покажем порядок их расчета.

Расчет среднего линейного отклонения взвешенного:

1. Определяется средняя сменная выработка рабочего по формуле средней арифметической взвешенной:

$$\bar{x}=\frac{\Sigma xf}{\Sigma f}.$$

2. Находятся отклонения каждого варианта от средней арифметической взвешенной путем вычитания из значения каждого варианта величины средней арифметической:

$$x-\bar{x}.$$

3. Умножается абсолютное значение отклонений каждого варианта на веса (на число единиц в каждой группе):

$$(x-\bar{x})f.$$

4. Суммируются произведения взвешенных абсолютных значений этих отклонений:

$$\Sigma(x-\bar{x})f.$$

5. Находится среднее взвешенное линейное отклонение по селедующей формуле:

$$\bar{d}=\frac{\Sigma(x-\bar{x})f}{\Sigma f}.$$

Расчет среднего квадратического отклонения взвешенного производится в следующем порядке:

1 и 2. Эти элементы расчета совпадают с соответствующими элементами расчета среднего линейного отклонения взвешенного.

3. Отклонения по каждому варианту возводятся в квадрат.

4. Каждый квадрат отклонений умножается на число единиц в группе, т.е. на веса:

$$(x-\bar{x})^2 f.$$

5. Находится **сумма** взвешенных квадратов отклонений:

$$\Sigma(x-\bar{x})^2 f.$$

6. Сумма взвешенных квадратов отклонений делится на сумму весов. Такая средняя из взвешенных квадратов отклонений является взвешенной дисперсией, или взвешенным средним квадратом отклонений, и обозначается символом σ^2, т.е.

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

7. Среднее квадратическое отклонение взвешенное определяется путем извлечения квадратного корня из взвешенной дисперсии (из суммы взвешенных квадратов отклонений, деленной на сумму весов), т.е.

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

Рассмотрим порядок вычисления взвешенных среднего линейного и среднего квадратического отклонений, продолжая наш пример (табл. 6.4).

Средняя **сменная** взвешенная **выработка одного рабочего, как** нам уже известно, на обоих заводах одинакова: 42 шт.

На основе данных предыдущей таблицы определяем среднее линейное отклонение взвешенное по каждому заводу. Оно равно:

- по первому заводу $\bar{d} = \frac{\Sigma(x-\bar{x})f}{\Sigma f} = \frac{160}{200} = 0{,}8$ шт.;
- по второму заводу $\bar{d} = \frac{60}{200} = 0{,}3$ шт.

Таблица 6.4

Элементы расчета показателей вариации сменной выработки рабочих

Номера групп	Сменная выработка изделий (шт.)	Число рабочих в смене	Отклонение от средней	Взвешенное абсолютное отклонение	Квадрат отклонений	Взвешенный квадрат отклонений
	x	f	$x-\bar{x}$	$(x-\bar{x})f$	$(x-\bar{x})^2$	$(x-\bar{x})^2 f$
Завод №1						
1	40	20	–2	40	4	80
2	41	40	–1	40	1	40
3	42	80	0	–	0	–
4	43	40	+1	40	1	40
5	44	20	+2	40	4	80
Итого	–	200	–	160	–	240
		Σf		$\Sigma(x-\bar{x})f$		$\Sigma(x-\bar{x})^2 f$
Завод № 2						
1	41	30	–1	30	1	30
2	42	140	0	–	0	–
3	43	30	+1	30	1	30
Итого	–	200	–	60	–	60
		Σf		$\Sigma(x-\bar{x})f$		$\Sigma(x-\bar{x})^2 f$

Среднее линейное отклонение взвешенной показывает, что вариация сменной выработки рабочих на первом заводе была в 2,7 (0,8 : 0,3) раза больше, чем на втором заводе.

На основании данных предыдущей таблицы произведем также расчет среднего квадратического отклонения взвешенного:

- по первому заводу $\sigma = \sqrt{\frac{\Sigma(x-\bar{x})^2 f}{\Sigma f}} = \sqrt{\frac{240}{200}} = \sqrt{1{,}2}$ =1,1 шт.;
- по второму заводу $\sigma = \sqrt{\frac{60}{200}} = \sqrt{0{,}3} = 0{,}55$ шт.

Таким образом, среднее квадратическое отклонение взвешенное по первому заводу в 2 (1,10:0,55) раза было больше, чем по второму заводу.

Среднее линейное отклонение взвешенное и среднее квадратаческое отклонение взвешенное, кроме случаев равной значимости отдельных вариантов, дают более правильную характеристику колеблемости признака, чем среднее линейное отклонение простое и среднее квадратическое отклонение простое.

В случае же одинаковой значимости отдельных вариантов расчет средних отклонений надо вести по формуле простой (равновзвешенной). Что же касается соотношения среднего линейного и среднего квадратического отклонения, то, как уже указывалось выше, теория математической статистики доказывает, что более правильную характеристику степени колеблемости, степени вариации признака дает среднее квадратическое отклонение.

Среднее линейное отклонение и среднее квадратическое отклонение как простое, так и взвешенное, выражены в той же единице измерения, в которой выражен варьирующий признак.

В нашем примере варьирующий признак был выражен в штуках и, соответственно, в штуках были выражены и величины линейных и квадратических отклонений.

6.6. Относительные показатели вариации

Для характеристики меры колеблемости, изучаемого признака, исчисляются показатели колеблемости в относительных величинах. Они позволяют сравнивать характер рассеивания в различных распределениях (различные единицы наблюдения одного и того же признака в двух совокупностях, при различных значениях средних, при сравнении разноименных совокупностей). Расчет показателей меры относительного рассеивания осуществляют как отношение абсолютного рассеивания к средней арифметической, умножаемое на 100%.

Коэффициент осцилляции

Коэффициент осцилляции отражает относительную колеблемость крайних значении признака вокруг средней.

$$K_o = \frac{R}{\bar{x}} \cdot 100\%.$$

Коэффициент осцилляции сменной выработки рабочих, исчисленный на основе размаха вариации, составит:

- по первому заводу $Ko = \frac{R}{\bar{x}} = \frac{4}{42} = 0{,}095$ или 9,5 %;
- по второму заводу $Ko = \frac{R}{\bar{x}} = \frac{2}{42} = 0{,}048$ или 4,2 %.

На заводе 1 разница между крайними значениями на 9,5% превышает среднее значение сменной выработки.

В то же время на заводе 2 этот показатель составляет 4,8% среднего значения.

Относительное линейное отклонение

Относительное линейное отклонение характеризует долю усредненного значения абсолютных отклонений от средней величины.

$$K_d = \frac{\bar{d}}{\bar{x}} \cdot 100\%.$$

На заводе 1 он составил 2,8% против 1,6%.

$$K_{d_1} = \frac{1{,}2}{42} \cdot 100 = 2{,}8\%. \quad K_{d_2} = \frac{0{,}47}{42} \cdot 100 = 1{,}6\%.$$

Коэффициент вариации – относительный показатель вариации, который определяется как отношение среднего квадратического отклонения к средней арифметической изучаемого показателя.

Коэффициент вариации

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

где v – коэффициент вариации;

σ – среднее квадратическое отклонение;

$\bar{x}$ – средняя арифметическая.

Коэффициент вариации может быть исчислен по среднему квадратическому отклонению простому и взвешенному.

Коэффициент вариации сменной выработки рабочих, исчисленный на основе среднего квадратического отклонения простого, составит:

- по первому заводу $v = \frac{\sigma}{\bar{x}} = \frac{1{,}42}{42} = 0{,}034$ или 3,4%;
- по второму заводу $v = \frac{\sigma}{\bar{x}} = \frac{0{,}82}{42} = 0{,}0195$ или 1,95%.

Сравнение этих коэффициентов вариации показывает, что колеблемость сменной выработки рабочих на втором заводе меньше, чем на первом.

Коэффициенты вариации, исчисленные на основе среднего квадратического отклонения взвешенного, составят:

- по первому завод $v = \frac{\sigma}{\bar{x}} = \frac{1{,}1}{42} = 0{,}026$, или 2,6%;
- по второму заводу $v = \frac{\sigma}{\bar{x}} = \frac{0{,}55}{42} = 0{,}013$, или 1,3%.

Сравнение этих коэффициентов показывает, что вариация сменной выработки рабочих на втором заводе в два раза меньше, чем на первом. Подобный же результат дало, как мы видели выше, и сравнение средних квадратических отклонений по этим заводам (1,1 шт. – на первом и 0,55 шт. – на втором). Одинаковость результатов сравнения обусловлена тем, что на обоих заводах величина средней сменной выработки на одного рабочего одинакова (42 шт.). В том же случае, если бы величина средней сменной выработки на одного рабочего на этих заводах была бы различна, непосредственное сравнение величин средних квадратических отклонений могло бы привести к неправильному представлению о сравнительной интенсивности вариации, колеблемости на этих заводах. Поясним это примером.

Допустим, что среднее квадратическое отклонение на первом заводе составляет по-прежнему 1,1 шт. и на втором заводе по-прежнему 55 шт., но средняя сменная выработка одного рабочего на первом заводе – 42 шт., а на втором заводе – 17 шт. В таком случае коэффициент вариации смежной выработки рабочих на первом заводе составляет по-прежнему 2,6%, а на втором заводе

$$v = \frac{0{,}55}{17} = 0{,}032, \quad \text{или } 3{,}2\%.$$

Коэффициент вариации на втором заводе оказался выше, чем на первом заводе, что ясно показывает на ***более*** значительную интенсивность вариации на втором заводе по сравнению с первым.

Таким образом, преимущество коэффициента вариации перед абсолютными показателями, характеризующими меру колеблемости признака, заключаются в том, что он дает относительную характеристику колеблемости, позволяющую проводить сравнение и при разных средних величинах. При этом различия средних величин могут заключаться не только в различиях абсолютных величин средних значений одного и того же признака, но и в том, что абсолютные значения средних величин относятся к разным признакам.

Коэффициент вариации используется для оценки интенсивности вариации и как относительный показатель интенсивности имеет размерность. Он показывает, сколько единиц среднего квадратического отклонения приходится на единицу среднего значения изучаемого признака. Совокупность считается однородной, если коэффициент вариации не превышает 0,33. В этом случае средняя величина исследуемого признака может считаться типичной, надежной характеристикой статистической совокупности. Если же коэффициент вариации больше 0,33, то это означает, что вариация исследуемого признака велика. Найденная средняя плохо представляет всю статистическую совокупность и не может считаться ее типичной, надежной характеристикой. Статистическая совокупность является неоднородной по рассматриваемому признаку.

6.7. Правило «трех сигм»

Дисперсию и среднее квадратическое отклонение используют при расчетах, связанных с организацией выборочного наблюдения, оценке полученных на основе выборки статистических показателей, построении показателей тесноты корреляционной связи, дисперсионном анализе. В условиях нормального распределения существует следующая зависимость между величиной среднего квадратического отклонения и количеством наблюдений: в пределах $\bar{x} \pm 1\sigma$ располагается 0,683 количества наблюдений; $\bar{x} \pm 2\sigma$ – 0,954; $\bar{x} \pm 3\sigma$ – 0,997.

Таблица 6.5

Количество счетов физических лиц, обслуживаемых филиалами коммерческих банков (тыс. ед.)

Филиалы банков	Банк 1	Банк 2
1	6,0	12,4
2	2,5	1,5
3	5,2	3,2
4	5,6	2,0
5	9,3	9,5
Итого	28,6	28,6
Средняя	5,7	5,7

Отклонение $\pm 3\sigma$ можно считать максимально возможным. Это положение называют правилом «трех сигм».

В симметричных распределениях среднее квадратическое отклонение составляет приблизительно 1,25 среднего линейного отклонения. Это соотношение может быть использовано для приближенного вычисления среднего квадратического отклонения, исходя из уже найденного значения среднего линейного отклонения. При таких расчетах следует учитывать и полученные согласно правилу «трех сигм» следующие соотношения:

$$\sigma \approx \frac{1}{6}(x_{\max} - x_{\min}),$$

так как в нормальном распределении в размахе вариации «укладываются» 6σ (±3σ).

Если распределение заведомо асимметричное, то

$$\sigma \approx \frac{1}{5}(x_{\max} - x_{\min}),$$

Рассмотрим порядок вычисления среднего линейного отклонения, дисперсии и среднего квадратического отклонения по приведенным ниже несгруппированным данным о работе филиалов двух банков с клиентами.

Данные, приведенные в табл. 6.5, свидетельствуют о том, что среднее число обслуживаемых одним филиалом счетов физических лиц в каждом банке одинаковое – 5,7 тыс. ед. Вместе с тем различия (вариация) по количеству обслуживаемых счетов клиентов более резко выражены у филиалов банка 2, чем у филиалов банка 1.

Об этом свидетельствуют следующие промежуточные расчеты (табл. 6.6).

Таблица 6.6

Данные расчета показателей

Филиалы	**Банк 1**		**Банк 2**	
	$\lvert x_i - \bar{x}\rvert$	$(x_i - \bar{x})^2$	$\lvert x_i - \bar{x}\rvert$	$(x_i - \bar{x})^2$
1	0,28	0,03	6,68	44,62
2	3,22	10,37	4,22	17,81
3	0,52	0,27	2,52	6,34
4	0,12	0,01	3,72	13,84
5	3,58	12.82	3,78	14,29
Итого	7,72	23,55	20,92	96,90

Используя приведенные ранее формулы, получаем соответствующие абсолютные и относительные характеристики размера вариации (табл. 6.7).

Таким образом, можно сделать вывод, что совокупность филиалов банка 1 имеет более низкий уровень их вариации по числу обслуживаемых счетов клиентов, чем совокупность филиалов банка 2, так как среднее квадратическое отклонение числа обслуживаемых клиентов по филиалам банка 1 составляет 2,17 тыс. ед., а по филиалам банка 2 – 4,40 тыс. ед. Совокупность считается количественно однородной, если коэффициент вариации не превышает 33%. В нашем примере коэффициент вариации у банка 1 составил 38%, а у банка 2 – 76%. Так как у банка 1 коэффициент вариации незначительно превышает 33%, можно предполагать, что среднее число обслуживаемых клиентов является типичным для филиалов этого банка и более точно отражает индивидуальные значения этого показателя у отдельных филиалов», чем у филиалов банка 2.

Таблица 6.7

Характеристики размера вариации

Показатели	Банк 1	Банк 2
Размах вариации	3,3	10,9
Среднее линейное отклонение	1,54	4,18
Дисперсия	4,71	19,38
Среднее квадратическое отклонение	2,17	4,40
Коэффициент осциляции, %	58	191
Относительное линейное отклонение, %	27	73
Коэффициент вариации, %	38	76

6.8. Дисперсия альтернативного признака

В ряде случаев возникает необходимость в измерении дисперсии так называемых **альтернативных признаков, т.е. признаков, которыми каждая отдельная единица совокупности или обладает, или не обладает.** Примером таких признаков могут служить различного рода звания, которые присваиваются за те или иные заслуги.

Наличие альтернативного признака у отдельного элемента совокупности приравнивается единице, его отсутствие – нулю. Доля единиц в совокупности, обладающих альтернативным признаком, обозначается через **p**, а доля единиц не обладающих им, – через **q**. Следовательно, распределение совокупности по альтернативному признаку имеет следующий вид:

Варианты льтернативного признака (x)	**Их удельный вес (f)**
0	**q**
1	**p**

Найдем для этого распределения среднюю дисперсию. Учитывая, что p + q = 1, а q = 1–p, будем иметь:

$$\bar{x} = 0 \cdot q + 1 \cdot q = p$$

$$\sigma^2 = (\mathbf{0} - p)^2 q + (\mathbf{1} - p)^2 p = p^2 q + q^2 p = pq\,(p + q) = pq.$$

Дисперсия альтернативного признака равна произведению доли единиц, обладающих признаком, и доли единиц, не обладающих им ($\sigma_p{}^2 = pq$):

$$\sigma_p^2 = \frac{\Sigma(x-\bar{x})^2 f}{\Sigma f} = \frac{(1-p)^2 p + (0-p)^2 q}{p+q} = pq$$

Например, 10 000 населения: 4000 мужчин, 6000 женщин.

$$p = \frac{4000}{10000} = 0{,}4, \quad p = \frac{6000}{10000} = 0{,}6.$$

$$\sigma_p^2 = pq = 0{,}4 \cdot 0{,}6 = 0{,}24.$$

p+q не может быть больше 1, а pq не может быть больше 0,25.

$$\sigma = \sqrt{\sigma_p^2} = \sqrt{0{,}24} = 0{,}49.$$

Рассмотрим следующий *пример*. Допустим, что при обследовании 1000 коммерческих банков 800 из них являются универсальными. Определите дисперсию и среднее квадратическое отклонение доли универсальных банков.

Решение. В данном случае доля единиц, обладающих изучаемым признаком, т.е. доля универсальных банков p = 800 : 1000 = 0,8, или 80%. Следовательно, 20% банков не обладали изучаемым признаком. Эту величину можно получить двояко:

а) q = (1000 – 800)/1000 = 200/1000 = 0,20, или 20%;

б) q = 1 – 0,80 = 0,20.

Следовательно, дисперсия доли универсальных банков

$$\sigma_p^2 = p \cdot q = 0{,}8 \cdot 0{,}2 = 0{,}16.$$

Среднее квадратическое отклонение –

$$\sigma_p = \sqrt{\sigma_p^2} = \sqrt{0{,}16} = 0{,}4.$$

6.9. Правило сложения дисперсий

Изучая дисперсию интересующего нас признака в пределах исследуемой совокупности и опираясь на общую среднюю в своих расчетах, мы не можем определить влияние отдельных факторов, характеризующих колеблемость индивидуальных значений (вариант) признака.

Это можно сделать при помощи группировок, подразделив изучаемую совокупность на группы, однородные по признаку-фактору. При этом можно определить три показателя колеблемости признака в совокупности: общую дисперсию, межгрупповую дисперсию и среднюю из внутригрупповых дисперсий.

Правило сложения дисперсий гласит, что общая дисперсия (σ_0^2) может быть разложена на две составные части: 1) межгрупповую ($\delta_{мг}^2$) и 2) среднюю из внутригрупповых дисперсий ($\delta_{вг}^2$).

$$\sigma_0^2 = \delta_{мг}^2 + \delta_{вг}^2.$$

Общая дисперсия отражает вариацию результативного признака, сложившуюся под воздействием всей совокупности причин и условий, определяющих его изменение. Она равна среднему квадрату отклонений отдельных значений результативного признака от его средней величины и может быть рассчитана по следующей формуле:

$$\sigma_y^2 = \frac{\Sigma(y_i - \bar{y})^2}{n}$$

Межгрупповая дисперсия характеризует ту часть общей дисперсии, которая обусловлена делением совокупности на группы, т.е. отражает различия в величине изучаемого признака, которые возникают под влиянием одного условия, т.е. это вариация признака, положенного в основу группировки.

Она равна среднему квадрату отклонений групповых средних $\bar{y}_k$ от общей средней $\bar{y}$:

$$\delta^2_{мг} = \frac{\Sigma(\bar{y}_k - \bar{y})^2 f}{\Sigma f},$$

где $\bar{y}_k$ – среднее значение признака k-й группы.

Средняя из внутригрупповых дисперсий характеризует остаточную вариацию, которая происходит под влиянием других, не связанных с группировкой факторов. Она вычисляется как средняя из внутригрупповых дисперсий ($\sigma^2_{вг}$):

$$\bar{\delta}^2_{гр} = \frac{\Sigma\sigma^2_{гр} \cdot f}{\Sigma f}.$$

где $\sigma^2_{гр}$– дисперсия в отдельных группах; f – численость отдельных групп.

Дисперсия в отдельных группах вычисляется по формуле:

$$\sigma^2_{гр} = \frac{\Sigma(y_i - \bar{y}_k)^2}{f}$$

Правило сложения дисперсий используют при определении степени точности типической выборки и измерении тесноты связи изучаемого результативного признака с признаками-факторами.

Ряды распределения дают возможность выявить структуру исследуемой совокупности, но сами по себе не объясняют причин, определяющих изменение этой структуры.

Для изучения влияния какого-либо фактора на результат необходимо оценить вариацию последнего, обусловленную влиянием исследуемого фактора. В аналитической группировке эту задачу решают на основе межгрупповой дисперсий. Очевидно, что чем больше доля межгрупповой дисперсии в общей, тем сильнее влияние группировочного признака-фактора на изучаемый результативный признак. Показателями тесноты связи служат коэффициент детерминации и эмпирическое корреляционное отношение.

Коэффициент детерминации определяют как долю межгрупповой дисперсии в общей дисперсии признака-результата. Он показывает влияние изучаемого фактора **x** на часть общей вариации признака-результата *y*:

$$\eta^2 = \frac{\delta^2_{мг}}{\sigma^2_y}$$

Эмпирическое корреляционное отношение – это корень квадратный из коэффициента детерминации:

$$\eta = \sqrt{\frac{\delta^2_{мг}}{\sigma^2_y}}.$$

Для оценки тесноты связи на основе показателя эмпирического корреляционного отношения можно воспользоваться соотношениями Чэддока (см. табл. 8.9).

Таблица 6.8

Данные расчета для определения показателей тесноты связи

Срок функционирования банка (x), лет	**Число банков (f)**	**Сумма активов по группе банков (Σy_i) (млн сум)**	**Средняя величина активов по группе банков ($\bar{y}_к$), (млн сум)**	**Сумма квадратов отклонений [$(\bar{y}_k-\bar{y})^2$]**	**Внутригрупповая дисперсия (σ^2_k)**
1-4	10	1138	113,80	8886,36	28,63
4-7	37	4904,7	132,56	4517,79	20,351
7-10	69	10 110,5	146,53	588,32	22,54
10-13	24	3951,6	164,65	10 624,36	24,65
Итого	140	20404,8	–	24 616,83	–

Если связь между признаками отсутствует, то межгрупповая дисперсия равна нулю, а следовательно, и коэффициент корреляции равен нулю.

Таким образом, чем ближе значение показателя к единице, тем сильнее связь между признаками. Рассмотрим пример анализа зависимости объема активов коммерческих банков (у) от срока их функционирования (х). Результаты группировки и промежуточные расчеты для определения показателей тесноты связи представлены в табл. 6.8

Проверка правила сложения дисперсий:

$$\bar{y} = \frac{\Sigma y_i}{n} = \frac{20104{,}8}{140} = 143{,}61;$$

$$\bar{\delta}^2_{мг} = \frac{3023{,}54}{140} = 21{,}60;$$

$$\delta^2_{мг} = \frac{24616{,}83}{140} = 175{,}84;$$

$$\sigma^2_y = \frac{27640{,}678}{140} = 197{,}43;$$

Проверка правила сложения дисперсий:

$$\sigma^2_y = \delta^2_{мг} + \bar{\delta}^2_{вг};$$

$$197{,}43 = 175{,}84 + 21{,}60.$$

Расчет показателей тесноты связи:

$$\eta^2 = \frac{175{,}84}{197{,}43} \cdot 100\% = 89{,}06\%;$$

$$\eta = \sqrt{0{,}8906} = 0{,}94.$$

По изучаемой совокупности банков наблюдается высокая теснота связи между сроком их деятельности и объемом активов. Более 89% различий в объеме активов банков определяется вариацией сроков их функционирования.

Правило сложения дисперсий позволяет определить одну из дисперсий, если известны две другие. Покажем это на примере.

Задание 1

По данным аналитической группировки (табл. 6.9) требуется:

1. Расчет показателей вариации.

2. Рассчитать: общую дисперсию; межгрупповую дисперсию; внутригрупповые дисперсии и среднюю из них.

3. Проверить действие правила сложения дисперсий.

Как видно из табл. 6.9, наиболее многочисленной группой фирм является группа со среднесписочной численностью менеджеров 25–30 человек, которая включает 8 фирм (27%); в самую малочисленную группу со среднесписочной численностью менеджеров 40–45 человек входит всего 1 фирма (3%).

Таблица 6.9

Интервальный ряд распределения фирм по среднесписочной численности менеджеров в одном из регионов РУз в I квартале отчетного года

Вариант признака (x_i). Численность менеджеров, чел.	Частота (f_i). Число фирм, ед.	Частость (w_i). Доля фирм в общем итоге (нарастающий итог)
15-20	3	3/30 = 0,10
20-25	6	0,20
25-30	8	0,27
30-35	7	0,23
35-40	5	6,17
40-45	1	0,03
ИТОГО	30	1,00

Решение

1. Для расчета показателей вариации на основании табл. 6.9 построим вспомогательную табл. 6.10.

Таблица 6.10

Вспомогательная таблица для расчета показателей вариации

x_j	x_{uj}	f_j	$x_{uj}f_j$	$\lvert x_{uj}-\bar{x}\rvert$	$\lvert x_{uj}-\bar{x}\rvert f_j$	$(x_{uj}-\bar{x})^2$	$(x_{uj}-\bar{x})^2 f_j$
15-20	17,5	3	52,5	11,33	33,99	128,3689	385,1067
20-25	22,5	6	135,0	6,33	37,98	40,0689	240,4134
25-30	27,5	8	220,0	1,33	10,64	1,7689	14,1512
30-35	32,5	7	227,5	3,67	25,69	13,4689	94,2823
35-40	37,5	5	187,5	8,67	43,35	75,1689	375,8445
40-45	42,5	1	42,5	13,67	13,67	186,8689	186,8689
ИТОГО	Х	30	865,0	Х	165,32	Х	1296,6670

Результаты расчета:

$$\bar{x}_{ар.\,взв} = \frac{\sum_{j=1}^{k} x_{цj} f_j}{\sum_{j=1}^{k} f_j} = \frac{865}{30} = 28{,}83$$ (чел.).

1. $R = 45 - 15 = 30$ *(чел.)*.

2. $$d = \frac{\sum_{j=1}^{k} \left| x_{цj} - \bar{x} \right| f_j}{\Sigma f_j} = \frac{165{,}32}{30} = 5{,}51$$ *(чел.)*.

3. $$\sigma^2 = \frac{\sum_{j=1}^{k} (x_{цj} - \bar{x})^2 f_j}{\sum_{j=1}^{k} f_j} = \frac{1296{,}667}{30} = 43{,}2222.$$

4. $\sigma = \sqrt{43{,}2222} = 6{,}57$ *(чел.)*.

5. $$v_\sigma = \frac{\sigma}{\bar{x}} = \frac{6{,}7}{28{,}83} = 0{,}23 < 0{,}33.$$

Анализ полученных данных говорит о том, что численность менеджеров в каждой из рассматриваемых фирм отличается от средней численности менеджеров в одной фирме в среднем примерно на 7 человек. Значение коэффициента вариации, равное 0,23, не превышает 0,33 (или 33%). Следовательно, вариация численности менеджеров фирм невелика. Таким образом, найденная средняя численность менеджеров фирмы (29 чел.) может представлять всю исследуемую совокупность, является ее типичной, надежной характеристикой, а вся совокупность фирм может считаться однородной по численности их менеджеров.

2. С целью усвоения теоретического материала введем исходную информацию для общих межгрупповых и внутригрупповых дисперсий и средних из них.

Имеются условные данные по среднесписочной численности менеджеров по продажам, количеству проданного ими однокачественного товара, индивидуальной рыночной цене на этот товар, а также объему продаж 30 фирм в одном из регионов РУз в I кв. отчетного года (табл. 6.11).

Используя исходные данные табл. 6.11, а также интервальный ряд распределения фирм по численности менеджеров (табл. 6.9), требуется построить аналитическую группировку зависимости между численностью менеджеров и объемом продаж фирм и на основании ее сделать вывод о наличии (или отсутствии) связи между указанными признаками.

Таблица 6.11

Исходная информация для сквозной задачи

Номер фирмы	Численность менеджеров, чел.	Количество проданного товара, шт.	Цена, тыс. сум	Объем продаж, млн сум
1	15	18	528	9,50
2	24	20	515	10,30
3	39	22	499	10,98
4	25	20	520	10,40
5	20	19	530	10,07
6	27	20	518	10,36
7	20	19	527	10,01
8	25	20	500	10,00
9	29	20	515	10,30
10	27	20	495	9,90
11	22	19	520	9,88
12	33	21	505	10,61
13	32	21	499	10,48
14	35	22	480	10,56
15	17	18	530	9,54
16	25	20	511	10,22
17	33	21	516	10,84
18	32	21	510	10,71
19	30	21	490	10,29
20	35	22	485	10,67
21	18	18	532	9,58
22	45	23	478	11,00
23	33	21	515	10,82
24	39	23	475	10,93

25	27	20	513	10,26
26	20	19	514	9,77
27	38	22	488	10,74
28	34	21	500	10,50
29	28	20	515	10,30
30	22	19	515	9,79
Итого	849	610	–	309,31

Решение:

Аналитическая группировка строится по факторному признаку. В нашей задаче факторным признаком (х) является численность менеджеров, а результативным признаком *(у)* – объем продаж.

Построим теперь аналитическую группировку (табл. 6.12).

На основании данных построенной аналитической группировки можно сказать, что с увеличением численности менеджеров по продажам средний в группе объем продаж фирмы также увеличивается, что свидетельствует о наличии прямой связи между указанными признаками (табл. 6.12).

Таблица 6.12

Зависимость объемов продаж от численности менеджеров фирм в одном из регионов РУз в I кв. отчетного года

Номер группы	Численность менеджеров, чел., x_j	Число фирм, ед., f_j	Средний объем продаж фирмы, млн сум, y
1	15-20	3	9,54
2	20-25	6	9,97
3	25-30	8	10,22
4	30-35	7	10,61
5	35-40	5	10,78
6	40-45	1	11,00
Итого		30	10,31

Используя исходные данные табл. 6.11, производим группировку менеджеров по фактическому признаку (т.е. по численности рабочих) (табл. 6.13).

Таблица 6.13

Вспомогательная таблица для расчета $\sigma^2,\ \delta^2,\ \overline{\sigma}_j^2$

Номер группы	Численность менеджеров, чел., x	Номер фирмы	Объем продаж, млн руб., y
1	**2**	**3**	**4**
1	15-20	1	9,50
		15	9,54
		21	9,58
Итого		3	$\bar{y}_1 = \frac{28{,}62}{3} = 9{,}54$
2	20-25	5	10,07
		2	10,30
		7	10,01
		11	9,88
		26	9,77
		30	9,79
Итого		6	$\bar{y}_2 = \frac{59{,}82}{6} = 9{,}97$
1	2	3	4
3	*25-30*	*4*	*10,40*
		6	*10,36*
		8	*10,00*
		9	*10,30*
		10	*9,90*
		16	*10,22*
		25 ,	*10,26*
		29	*10,30*
Итого		8	$\bar{y}_3 = \frac{81{,}74}{8} = 10{,}22$
4	*30-35*	*17*	*10,84*
		12	*10,61*
		13	*10,48*
		18	*10,71*
		19	*10,29*
		23	*10,82*
		28	*10,50*
Итого		7	$\bar{y}_4 = \frac{74{,}25}{7} = 10{,}61$
5	*35-40*	*3*	*10,98*
		14	*10,56*
		20	*10,67*

		24	*10,93*
		27	*10,74*
Итого		*5*	$\bar{y}_5 = \frac{53{,}88}{5} = 10{,}78$
6	*40-45*	*22*	*11,00*
Итого		*1*	$\bar{y}_6 = $ ***11,00***
Всего		*30*	$\bar{y} = \frac{309{,}31}{30} = 10{,}31$

Решение:

1. Расчет общей дисперсии (σ^2)

2. $$\sigma^2 = \frac{\sum_{j=1}^{k}\sum_{i=1}^{n}(y_{ij} - \bar{y})^2}{n}.$$

$\sigma^2 = [(9{,}50 - 10{,}31)^2 + (9{,}54 - 10{,}31)^2 + (9{,}58 - 10{,}31)^2 +$
$+ (10{,}07 - 10{,}31)^2 + (10{,}30 - 10{,}31)^2 \cdot 3 + (10{,}01 - 10{,}31)^2 +$
$+ (9{,}88 - 10{,}31) + (9{,}77 - 10{,}31)^2 + (9{,}79 - 10{,}31)^2 +$
$+ (10{,}40 - 10{,}31)^2 + (10{,}36 - 10{,}31)^2 + (10{,}00 - 10{,}31)^2 +$
$+ (9{,}90 - 10{,}31)^2 + (9{,}90 - 10{,}31)^2 + (10{,}22 - 10{,}31)^2 +$
$+ (10{,}26 - 10{,}31)^2 + (10{,}84 - 10{,}31)^2 + (10{,}50 - 10{,}31)^2 +$
$+ (10{,}98 - 10{,}31)^2 + (10{,}56 - 10{,}31)^2 + (10{,}67 - 10{,}31)^2 +$
$+ (10{,}93 - 10{,}31)^2 + (10{,}74 - 10{,}31)^2 + (11{,}00 - 10{,}31)^2]/30 =$
$= 5{,}5036/30 = 0{,}1834.$

1. Расчет межгрупповой дисперсии (δ^2).

Прежде всего построим вспомогательную табл. 6.14.

Таблица 6.14

Вспомогательная таблица для расчета межгрупповой дисперсии

Номер груп-пы	Численность менеджеров, чел., x_j	Число фирм, ед., f_j	Объем продаж, млн сум		$(y_i - \bar{y}_i)^2 f_j$
			y_{ij}	в среднем на 1 фирму, $\bar{y}_i$	
1	15-20	3	9,50; 9,54; 9,58	9,54	1,7787
2	20-25	6	10,07; 10,30; 10,01; 9,88; 9,77; 9,79	9,97	0,6936
3	25-30	8	10,40; 10,36; 10,00; 10,30; 9,90; 10,22; 10,26; 10,50	10,22	0,0648

4	30-35	7	10,84; 10,61; 10,48; 10,71; 10,29; 10,82; 10,50	10,61	0,6300
5	35-40	5	10,98; 10,56; 10,67; 10,93; 10,74	10,78	1,1045
6	40-45	1	11,00	11,00	0,4761
Итого		**30**		$\bar{y}$ **= 10,31**	**4,7477**

$$\delta^2 = \frac{\sum_{j=1}^{k}(\bar{y}_j - \bar{y})f_j}{\sum_{j=1}^{k} f_j} = \frac{4,7477}{30} = 0,1583$$

3. Расчет внутригрупповых дисперсий (σ_j^2) и средней из них $(\bar{\sigma}_j^2)$;

$$\bar{\sigma}_j^2 = \frac{\sum_{i=1}^{f_j}(y_{ij} - \bar{y}_j)^2}{f_j}.$$

$$\sigma_1^2 = \frac{(9,50-9,54)^2 + 9,54-9,54)^2 + (9,58-9,54)^2}{3} = 0,00107\ .$$

$$\sigma_2^2 = [(10,07-9,97)^2 + (10,30-9,97)^2 + (10,01-9,97)^2 +$$
$$+ (9,88-9,97)^2 + (9,77-9,97)^2 + (9,79-9,97)^2]/6 = 0,01595.$$

$$\sigma_3^2 = [(10,4 - 10,22)2 + (10,36 + 10,22)2 + (10 - 10,22)2 +$$
$$+ (10,3- 10,22)2 + (9,9 - 10,22)2 + (10,22 - 10,22)2 +$$
$$+ (10,26 - 10,22)2 + (10,5 - 10,22)2]/8 = 0,03615.$$

$$\sigma_4^2 = [(10,84 - 10,61)2 + (10,61 - 10,61)2 + (10,48 - 10,61)2 +$$
$$+ (10,29 - 10,61)2 + (10,82 - 10,61)2 + (10,5 - 10,61)2]/7 = 0,03406.$$

$$\sigma_5^2 = [(10,98 - 10,78)2 + (10,56 - 10,78)2 + (10,67 - 10,78)2 +$$
$$+ (10,93 - 10,78)2+ (10,74 - 10,78)2]/5 = 0,02492.$$

$$\sigma_6^2 = \frac{(11-11)^2}{1} = 0.$$

$$\bar{\sigma}_j^2 = \frac{\sum_{i=1}^{k}\sigma_j^2 f_j}{\sum_{j=1}^{k} f_j} = (0,00107\cdot 3 + 0,01595\cdot 6 + 0,03615\cdot 8 +$$

$$+ 0,03406\cdot 7 + 0,02492\cdot 5 + 0\cdot 1)/30 = 0,0251.$$

4. *Проверка действия правила сложения дисперсий.*

Правило сложения дисперсий: $\sigma^2 = \delta^2 + \sigma_j^2$

0,1834 = 0,1583 + 0,0251;

0,1834 = 0,1834.

6.10. Основные свойства дисперсии

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

1. Если все значения признака уменьшить или увеличить на какое-то постоянное число **а**, то дисперсия от этого не изменится. Следовательно, дисперсию можно исчислить не только по вариантам, но и по их отклонениям от какого-то постоянного числа **а**.

2. Если все значения признака уменьшить или увеличить в К раз, то дисперсия от этого, соответственно, изменится в $К^2$ раз. Следовательно, при исчислении дисперсии можно все значения признака уменьшить в К раз, исчислить дисперсию, а затем умножить ее на это постоянное число в квадрате ($К^2$).

3. Сумма квадратов отклонений индивидуальных значений признака x от их средней $\bar{x}$ меньше суммы квадратов отклонений индивидуальных значений признака от любого данного числа **а** при условии, что $a \neq \bar{x}$, т.е. $\Sigma(x-\bar{x})^2 < \Sigma(x-a)^2$.

Доказано, что эти две суммы отличаются на квадрат разности между a и $\bar{x}$:

$$\Sigma(x-a)^2 < \Sigma(x-\bar{x})^2 = (\bar{x}-a)^2.$$

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

4. Дисперсия признака равна разности между средним квадратом значений признака и квадратом их средней, т.е. $\sigma^2 = \overline{x^2} - \bar{x}^2$.

Рассмотрим вычисление дисперсии с применением ее свойств. Один из упрощенных способов вычисления дисперсии основан на следующем равенстве:

$$\sigma^2 = \frac{\Sigma(x-a)^2 \cdot f}{\Sigma f} - (\bar{x}-a)^2.$$

Этот способ исчисления σ^2 называется способом моментов или способом отсчета от условного нуля.

6.11. Использование показателей вариации в анализе взаимосвязей социально-экономических явлений

Знание о различных видах дисперсии позволяет построить показатели взаимосвязи факторных и результативных признаков и дать им оценку.

Для оценки тесноты корреляционной связи между факторным и результативным признаками на базе эмпирического материала строят следующие показатели.

1. Эмпирический коэффициент детерминации, который определяется как доля межгрупповой (факторной) дисперсии в общей дисперсии, характеризует силу влияния факторного (группировочного) признака на результативный:

$$\eta^2 = \frac{\delta^2}{\sigma^2}.$$

Напомним, что на величину **межгрупповой** дисперсии оказывает влияние только факторный признак, а на величину общей дисперсии помимо факторного признака – все остальные признаки.

Поэтому частное от деления межгрупповой дисперсии на общую дисперсию покажет, насколько вариация результативного признака объясняется вариацией факторного признака (остальная часть вариации результативного признака объясняется вариацией прочих, неучтенных факторов).

1. ***Эмпирическое корреляционное отношение*** – корень квадратный из эмпирического коэффициента детерминации, характеризует тесноту связи:

$$\eta = \sqrt{\eta^2} = \sqrt{\frac{\delta^2}{\sigma^2}}.$$

Оба показателя находятся в пределах от 0 до 1. При этом чем ближе показатели к единице, тем связь между изучаемыми признаками теснее:

- 0 – связь отсутствует;
- 1 – связь функциональная.

Для оценки тесноты связи с помощью корреляционного отношения η можно воспользоваться шкалой американского ученого Чеддока.

Задание 2

На базе расчетных данных, полученных при решении задания 1 требуется:

1. Рассчитать:

а) эмпирический коэффициент детерминации;

б) эмпирическое корреляционное отношение.

2. Сделать выводы.

Решение:

$$\eta^2 = \frac{\delta^2}{\sigma^2} = \frac{0{,}1583}{0{,}1834} = 0{,}8631$$ или 86,31%.

Таким образом, 86,31% вариации объема продаж фирм обусловлено изменением численности менеджеров по продажам, а 13,69% (100,0% – 86,31%) – влиянием прочих факторов.

$$\eta = \sqrt{\eta^2} = \sqrt{0{,}8631} = 0{,}93.$$

По шкале Чеддока можно сделать вывод, что связь между численностью менеджеров и объемом продаж фирм весьма тесная.

6.12. Изучение формы распределения признака

Форма распределения признака в вариационных рядах распределения отражает закономерность изменения частот с ростом значений варьируемого признака.

Обобщающие характеристики формы распределения получают, используя кривые распределения – эмпирические и теоретические.

Эмпирическая кривая распределения – это фактическая кривая распределения, построенная по данным наблюдения, которая отражает как общие, так и случайные условия, определяющие распределение признака.

Теоретическая кривая распределения – это кривая распределения, выражающая связь между варьирующим признаком и частотами.

Анализировать эмпирическую кривую, на изменение которой повлияли, в том числе и случайные условия, – довольно трудно. Поэтому исследователи переходят к теоретической кривой распределения, по характеру распределения наиболее приближенной к первой. Теоретическая кривая призвана отражать основную закономерность распределения признака при полном погашении случайных причин.

При выравнивании (аппроксимации) эмпирических вариационных рядов перед исследователем стоят три задачи:

1. Выяснение общего характера распределения.

2. Непосредственное выравнивание эмпирического распределения – замена эмпирической кривой на теоретическую.

3. Проверка соответствия найденного теоретического распределения эмпирическому.

При анализе вариационных рядов распределения социально- экономических явлений было замечено, что характер изменения частот при изменении варьирующего признака в вариационных рядах распределения обычно обладает определенной особенностью: при росте варьирующего признака в вариационных рядах частоты сначала растут, а затем, достигнув максимального значения, начинают снижаться. Такое распределение характерно для нормального. Поэтому в качестве идеального распределения наиболее часто используют ***нормальное распределение.*** Таким образом, кривая нормального распределения является наиболее распространенной формой распределения и рассматривается в качестве теоретической кривой, по которой выравнивают эмпирическую кривую (известны еще, например, распределение Пуассона и, соответственно, кривая Пуассона).

Непрерывная случайная величина (**x**) подчиняется нормальному закону распределения, если ее функция плотности вероятности имеет вид:

$$f(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{\frac{(x-\bar{x})^2}{2\sigma^2}},$$

где π и e – математические константы (π = 3,14; e = 2,718).

Таким образом, для построения кривой нормального распределения надо знать два параметра: $\bar{x}$ и σ.

Для удобств расчета переходят от случайных величин к нормируемым (стандартизированным), когда средняя равна нулю, а дисперсия и среднеквадратическое отклонение равны единице.

Обозначим: $i=\dfrac{x_i-\bar{x}}{\sigma}$ – нормированные отклонения.

Тогда выражение $\sigma^2_{взв}=\dfrac{\Sigma(x_i-\bar{x})^2 f_i}{\Sigma f_i}$ может быть записано следующим образом:

$$f(t)=\frac{1}{\sqrt{2\pi}}e^{\frac{1}{2}t^2},$$

Это уравнение называется *стандартным уравнением нормальной* **кривой.** ***Нормируемая функция табулирована для различных значений t.***

Особенности кривой нормального распределения (рис. 6.1):

1. Кривая имеет форму колокола.

2. Так как функция нормального распределения – четная, т.е. $f(-t)=f(t)$, то кривая нормального распределения симметрична относительно максимальной ординаты, равной $\dfrac{1}{\sqrt{2\pi}}$ (при $t=0$).

Абсцисса этой точки является центром распределения: $\bar{x}=Mo=Me$.

3. Ветви кривой, приближаясь к оси абсцисс, уходят в $\pm\infty$, так как функция нормального распределения принимает бесконечно малые значения при $t = \pm\infty$.

4. Кривая имеет две точки перегиба при $t = \pm 1$, находящиеся на расстоянии $\pm\sigma$ от $\bar{x}$.

5. При $\bar{x}$ = const с увеличением σ кривая становится более пологой:

При σ = const с изменением $\bar{x}$ кривая не меняет свою форму, а лишь сдвигается вправо или влево по оси абсцисс.

6. В промежутке $\bar{x} \pm \sigma$ находится 68,3% всех значений признака; в промежутке $\bar{x} \pm 2\sigma$ находится 95,4% всех значений признака; в промежутке $\bar{x} \pm 3\sigma$ находится 99,7% всех значений признака (правило трех σ) (см. рис. 6.1 и 6.2).

При сопоставлении эмпирической кривой с кривой нормального распределения **необходимо проверить эмпирическую кривую на:**

• симметричность;

• наличие одной нормальной вершины (не острей и не плоской).

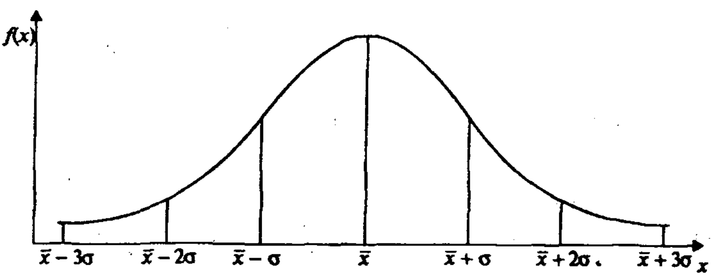


Рис. 6.1. **Кривая нормального распределения**

Эмпирические кривые распределения бывают **симметричные и асимметричные.**

Для симметричных распределений частоты двух вариантов, равностоящих в обе стороны от центра, равны между собой.

Рассчитанные для симметричных рядов характеристики:
$\bar{x} = Mo = Me$. σ = 1,25d; d = 0,8σ; R = 6σ.

Если вышепредставленные характеристики нарушены, то это свидетельствует о наличии асимметрии распределения.

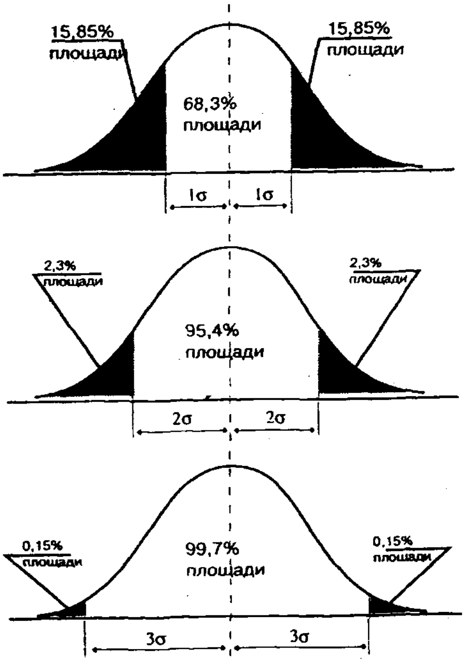


Рис. 6.2. **Соотношение площади под кривой нормального распределения в зависимости от расстояния от средней арифметической**

Асимметричные кривые имеют **правостороннюю** или **левостороннюю** (рис. 6.3) **асимметрию** – в зависимости от того, какая ветвь кривой вытянута: в первом случае – правая, во втором– левая.

Для того чтобы измерить асимметрию, рассчитывают показатели асимметрии.

Наиболее известный среди них – структурный коэффициент асимметрии Пирсона:

$$As_{Mo} = \frac{\bar{x} - Mo}{\sigma}.$$

Если $As_{Mo} > 0$, то асимметрия правосторонняя. В этом случае $Mo < Me < \bar{x}$. Если $As_{Mo} < 0$, то асимметрия левосторонняя. В этом случае $Mo > Me > \bar{x}$.

В симметричных рядах $As_{Mo} = 0$.

Расчетные данные представлены на рис 6.3.

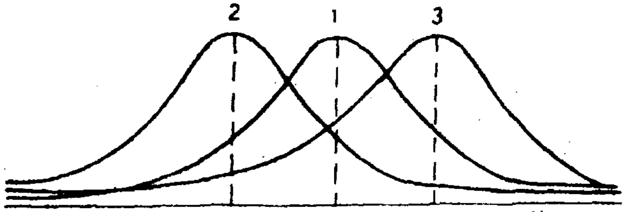


Рис. 6.3. **Кривые нормального распределения:**
1 – нормальное распределение;
2 – левосторонняя асимметрия;
3 – правосторонняя асимметрия

Другой показатель, предложенный Линдбергом, рассчитывают по формуле

$$A_s = n - 50,$$

где n – процент тех значений признака, которые превосходят по величине среднюю арифметическую.

Наиболее точным и распространенным является показатель, основанный на определении центрального момента третьего порядка (в симметричном распределении его величина равна нулю):

$$As = \frac{M_3}{\sigma^3}$$

Применение этого показателя дает возможность определить не только величину асимметрии, но и ответить на вопрос о налйчии или отсутствии асимметрии в распределении признака в генеральной совокупности. Оценка степени существенности этого показателя дается с помощью средней квадратической ошибки, которая зависит от объема наблюдений n и рассчитывается по формуле

$$\sigma_{A_s} = \sqrt{\frac{6(n-1)}{(n+1)(n+3)}}.$$

Если отношение $\frac{|A_s|}{\sigma_{A_s}} < 3,$ асимметрия существенна, а если $\frac{A_s}{\sigma_{A_s}} < 3,$ несущественна, ее наличие может быть объяснено влиянием различных обстоятельств.

6.13. Оценка эксцесса

После оценки симметричности ряда распределения переходят к оценке эксцесса.

Эксцесс (в переводе с англ. – излишество) – несовпадение вершины эмпирического распределения с вершиной кривой нормального распределения, когда она находится выше или ниже первой.

Следует помнить, что симметричные ряды могут иметь не одну вершину, а, например, три или пять, но такое распределение не будет однородным и его нельзя считать нормальным.

Показатели эксцесса рассчитывают только для симметричных распределений, имеющих одну вершину. Для симметричных распределений рассчитывается показатель эксцесса (островершинности). Линденбергом предложен следующий показатель:

$$E_x = n - 38{,}9$$

где n *– доля (%) количества вариантов, лежащих в интервале, равном половине среднего квадратического отклонения в ту и другую сторону от* $\bar{x}$.

Следовательно, наиболее точным показателем эксцесса является коэффициент, рассчитанный с использованием центрального момента четвертого порядка:

$$Ek = \frac{M_4}{\sigma_4} - 3.$$

Если Ek = 0 – это нормальное распределение (при нормальном распределении $\frac{M_4}{\sigma_4} = 3$); Ek > 0 – распределение островершинное; Ек < 0 – распределение плосковершинное (рис. 6.4).

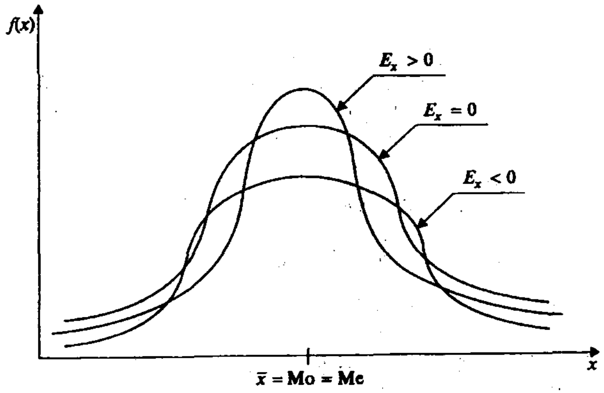


Рис. 6.4. **Эксцесс распределения**

Эксцесс представляет собой выпад вершины эмпирического распределения вверх или вниз от вершины кривой нормального распределения, где отношение $M_4/\sigma_4=3$. Средняя квадратическая ошибка эксцесса рассчитывается по формуле

$$\sigma_{Es} = \sqrt{\frac{24n(n-2)(n-3)}{(n-1)^2(n+3)(n+5)}},$$

где n – число наблюдений.

Оценка существенности показателей асимметрии и эксцесса позволяет сделать вывод о том, можно ли отнести данное эмпирическое распределение к типу кривых нормального распределения.

Если непрерывная случайная величина, имеет плотность распределения, как было отмечено выше, т.е.

$$f(x)=\frac{1}{\sqrt{2\pi\sigma}}e^{\frac{(x-\bar{x})^2}{2\sigma^2}},$$

то она подчиняется закону нормального распределения.

Если отношение |Ex| /Σ_{Ex} имеет значение больше трех, то это свидетельствует о существенном характере эксцесса.

Продолжим рассмотрение примера о распределении коммерческих банков по объему активов (табл. 6.15), исходный ряд которого представлен в первых 2-х графах, произведем расчет данных, которые необходимы для определения показателей степени вариации и характеристик формы распределения.

Таблица 6.15

Данные расчета показателей

Активы, млн	f	x	xf	$\lvert x-\bar{x}\rvert$	$\lvert x-\bar{x}\rvert_f$	$(x-\bar{x})^2 f$	$(x-\bar{x})^3 f$	$(x-\bar{x})^4 f$
105-115	4	110	440	33,3	133,2	4435,58	147 704,1	4 918570
115-125	9	120	1080	23,3	209,7	4886,01	113 844,0	2 652 566
125-135	21	130	2730	13,3	279,3	3714,69	49 405,4	657 091
135–145	49	140	6860	3,3	161,7	533,61	1760,9	5811
145-155	28	150	4200	6,7	187,6	1256,92	8421,4	56423
155-165	18	160	2880	16,7	300,6	5020,02	83 834,3	1400033
165-175	11	170	1870	26,7	293,7	7841,79	209375,8	5 590334
Итого	140	–	20080	–	1565,8	27688,60	614 345,9	15280828

Из формул следует, что для расчета показателей вариации на основе интервального ряда необходимо использовать середину интервала и предварительно определить среднюю величину изучаемого признака:

$$\bar{x}=\frac{\Sigma xf}{\Sigma f}=\frac{20080}{140}=143{,}3$$ млн сум;

$$\bar{d}=\frac{\Sigma\lvert x_i-\bar{x}\rvert f}{\Sigma f}=\frac{1565{,}8}{140}=11{,}15$$ млн сум;

$$\sigma^2=\frac{\Sigma(x_i-\bar{x})^2 f_i}{\Sigma f_i};\quad \sigma^2=\frac{27688{,}5}{140}=197{,}77;$$

ы $$\sigma=\sqrt{\frac{\Sigma(x_i-\bar{x})^2 f_i}{\Sigma f_i}}=14{,}06$$ млн сум;

$$\varphi_3 = \frac{\Sigma(x_i - \bar{x})^3 f}{\Sigma f} = \frac{6143460}{140} = 4{,}39;$$

$$\varphi_3 = \frac{\Sigma(x_i - \bar{x})^4 f}{\Sigma f} = \frac{15280829}{140} = 109{,}15;$$

$$A_s = \frac{\mu^3}{\sigma^3} = \frac{4{,}39}{2780{,}65} = 0{,}0016$$

$$\sigma_{A_s} = \sqrt{\frac{6(140-1)}{(140+1)(140+3)}} = 0{,}2;$$

$$\frac{|A_s|}{\sigma_{A_s}} = 0{,}008;$$

$$E_x = \frac{\varphi_4}{\sigma^4} - 3 = -2{,}997;$$

$$\sigma_{E_x} = \sqrt{\frac{24 \cdot 140(140-2)(140-3)}{(140-1)^2(140+3)(140+5)}} = 0{,}4;$$

$$|E_x| : \sigma_{E_x} = 7{,}493.$$

На основе рассчитанных обобщающих характеристик статистической совокупности коммерческих банков можно сделать следующие выводы. Средний объем активов кредитной организации составляет 143,3 млн сум, а показатели вариации: среднее линейное отклонение – 11,15 млн сум, среднее квадратическое отклонение – 14,06 млн сум, коэффициент вариации равен 9,81%. Отсюда следует, что изучаемая совокупность банков однородна по объему активов; асимметрия имеет несущественный характер, распределение более плосковершинно, чем нормальное, а отклонение от нормального распределения по показателю эксцесса является существенным.

5.14. Критерий Колмагорова

Для проверки, насколько фактическое распределение признака соответствует нормальному, нужно частоты фактического распределения сравнить с теоретическими частотами, которые характерны для нормального распределения. Для этого по фактическим данным нужно вычислить теоретические частоты, которые характерны для нормального распределения. Математическая статистика предлагает несколько показателей, называемых критериями согласия, по которым можно судить, насколько фактическое распределение согласуется с нормальным. Например, критерий Колмогорова:

$$\lambda = \frac{D}{\sqrt{n}},$$

где D – максимальная разность (по модулю) между кумулятивными фактическими и теоретическими частотами.

Рассчитав λ, определяют по его значению в специальной таблице соответствующее ему значение вероятности, с которой можно утверждать, что отклонения фактических частот от теоретических являются случайными.

При нормальном распределении средние показатели наиболее точно отражают характер изучаемого явления. Чем ближе фактическое распределение к нормальному, тем его средние характеристики становятся достовернее.

Задание 3

Для интервального ряда распределения фирм по численности менеджеров (табл. 6.9) используя уже имеющиеся расчетные данные (см. решения заданий 1 и 2), требуется рассчитать структурный коэффициент асимметрии Пирсона и сделать вывод.

Решение:

Используем в расчете структурного коэффициента асимметрии уже найденные значения:

• средней численности менеджеров фирмы: $\bar{x}$ = 28,83 чел. (см. решение задания 1);

• среднего квадратического отклонения: σ = 6,57 чел. (см. решение задания 1);

• моды: Мо = 28,33 чел. (расчет произведен по данным табл. 6.12) т.е.

$$Mo = x_0 + (x_1 - x_0)\frac{f_{Mo} - f_{Mo-1}}{(f_{Mo} - f_{Mo-1}) + (f_{Mo} - f_{Mo+1})} =$$

$$= 25 + 5\frac{8-6}{(8-6)+(8-7)} = 25 - 5 \cdot \frac{2}{3} = 28{,}33 \text{ (чел).}$$

Вывод. Чаще всего встречаются фирмы с численностью менеджеров примерно 28 человек.

Тогда в соответствии с формулой:

$$As_{Mo} = \frac{\bar{x} - Mo}{\sigma} = \frac{28{,}83 - 28{,}33}{6{,}57} = 0{,}0761.$$

Вывод. $As_{Mo} > 0$, т.е. асимметрия правосторонняя, незначительная (близка к нулю).

Интеллектуальный тренинг

1. Что такое вариация признака?
2. Чем объясняется необходимость изучения вариации признака?
3. Что такое размах вариации и в чем его особенности как показателя вариации?
4. Назовите показатели, являющиеся абсолютными характеристиками степени вариации.
5. Какие показатели являются относительными показателями характеристики вариации?
6. Назовите виды дисперсий и укажите способы их расчета.
7. Какова суть правила сложения дисперсий?
8. Как определяется дисперсия признака, обладающего альтернативной изменчивостью?
9. В чем состоят особенности расчета показателей вариации по сгруппированным данным?
10. Какое аналитическое значение имеет коэффициент вариации?
11. Как определяется внутригрупповая дисперсия?
12. Что характеризует межгрупповая дисперсия, формула ее расчета?
13. Что называется эмпирическим корреляционным отношением, и как оно интерпретируется?

14. Как изучается форма распределения признака?

15. Чем отличается эмпирическая кривая распределения от теоретической кривой распределения?

16. Какая кривая считается стандартным уравнением нормальной кривой?

17. Что такое эксцесс?

18. В чем суть выравнивания вариационных рядов по кривой нормального распределения?

Использованная и рекомендуемая специальная литература

1. Кривенкова Л., Юзбашев М. Область существования показателей вариации и ее применение // Вестник статистики. 1991. №6.

2. Коррель Ф.Ф. Анализ вариации и корреляции связи между экономическими явлениями. Т.: ТИНХ, 1972.

Глава VII
ВЫБОРОЧНОЕ НАБЛЮДЕНИЕ

Дорожная карта

7.1. Теоретические основы выборочного наблюдения

В связи с тем, что статистика имеет дело с массовыми совокупностями, статистические исследования весьма трудоемки и дороги. Поэтому давно возникла мысль о замене сплошного наблюдения выборочным.

Система правил отбора единиц и способов характеристики изучаемой совокупности исследуемых единиц составляет теоретическую основу выборочного наблюдения.

Важная роль в формировании выборочного метода наблюдения принадлежит работам Якоба Бернулли (1654-1705). Весомый вклад в разработку теоретических основ выборочного метода внесли русские математики – П.Л. Чебышев, А.М. Ляпунов, А.А. Марков. Российская статистика имеет немалые заслуги в практическом применении выборочного мстода. Так уже во второй половине XIX в. выборочные обследования проводились земскими статистиками и отличались определенной новизной в решении вопросов организации отбора единиц. При проведении подворной переписи крестьянских хозяйств Пензенской губернии 1909-1911 гг. был использован метод сочетания звеньев выборочного обследования разной подробности, впоследствии названный *многофазным отбором*. В ряде обследований использовалась *гнездовая выборка*, в других – *сплошное обследование* сочеталось *с механическим отбором единиц для контроля* полученных данных, например, при Всероссийской переписи населения 1916 г. теория выборочного метода получила развитие в трудах известного русского статистика А.А. Чупрова и в работе А.Г. Ковалевского «Основы теории выборочного метода», вышедшей в 1924 г. Классификацию форм выборочного наблюдения дали в 1930 г. известные русские статистики А.Я. Боярский и Б.С. Ястремский.

В последние годы выборочные обследования стали широко применяться в работе органов государственной статистики. Крупные и средние предприятия охватываются сплошным наблюдением за их деятельностью, а наблюдение за деятельностью малых предприятий производится с помощью выборочных обследований. В ряде случаев выборочные наблюдения применяются в сочетании со сплошными переписями и учетами. Например, программа Всесоюзной переписи населения 1989 г. содержала как вопросы сплошного наблюдения, относящиеся ко всему населению, так и вопросы выборочного наблюдения 25% населения для характеристики основного занятия, положения в занятии, места работы, а также вопросы 5% выборочного обследования с целью изучения брачности и рождаемости.

Обычно, при выборочном наблюдении обследованию подвергается определённая, заранее обусловленная часть совокупности, например, 1/10, 1/20, 1/50 и т.д., а результаты обследования распространяются на всю совокупность.

Выборочное наблюдение – это способ не сплошного наблюдения, при котором обследуется не вся совокупность, а лишь часть ее, отобранная по определенным правилам выборки и обеспечивающая получение данных, характеризующих всю совокупность в целом. Реализация выборочного метода базируется на понятиях генеральной и выборочной совокупностей.

Та статистическая совокупность, из которой делают отбор, называется *генеральной статистической совокупностью.*

В свою очередь, та статистическая совокупность, которая получилась в результате отбора единиц для наблюдения, называется *выборочной статистической совокупностью.*

Введем обозначения для характеристик генеральной и выборочной совокупностей (табл. 7.1).

Совокупность, из которой производится отбор, называют *генеральной,* а ее численность обозначают через N. Часть совокупности, подвергающуюся наблюдению, называют *выборочной,* а ее численность обозначают как n. Средние величины признаков, относительные величины, характеризующие генеральную совокупность, принято называть *генеральной средней* ($\bar{x}$), *генеральной долей (p)* и т.п.

Для выборочной совокупности, соответственно, эти величины называют *выборочной средней* ($\tilde{x}$), *выборочной долей (w)* и т.д. *Численность единиц, отбираемых в выборочную совокупность (n), называют объемом выборки.*

По мере увеличения численности (объема) выборки показатели выборочной совокупности все более приближаются к показателям генеральной совокупности, вероятность значительных расхождений между ними становится все меньше.

Числовые характеристики генеральной совокупности (средняя, дисперсия и др.) называют *параметрами* генеральной совокупности. *Оценка параметра* – это числовая характеристика, полученная на основе выборки.

Таблица 7.1

Основные характеристики параметров генеральной совокупности и оценок выборочной совокупности

Характеристики	Генеральная совокупность	Выборочная совокупность
Объем совокупности (численность единиц)	N	n
Численность единиц, обладающих обследуемым качеством (признаком)	M	m
Доля единиц, обладающих обследуемым качеством (признаком), выборочная доля	$p = \frac{M}{N}$	$w = \frac{m}{n}$
Среднее значение признака	$\bar{x} = \frac{\Sigma x}{N}$	$\tilde{x} = \frac{\Sigma \tilde{x}}{n}$
Дисперсия количественного признака	$\sigma_{\bar{x}^2} = \frac{\Sigma (x - \bar{x})^2}{N}$	$\sigma_{\tilde{x}^2} = \frac{\Sigma (x - \bar{x})^2}{n}$
Дисперсия альтернативного признака (доли)	$\sigma_p^2 = pq$	$\sigma_w^2 = w(1-w)$

Например, урожайность пшеницы в случайно отобранном из генеральной совокупности может оказаться намного выше, или, наоборот, ниже генеральной средней.

Но при продолжении случайного отбора хозяйств маловероятно чтобы во всех тридцати или сорока отобранных хозяйствах урожайность оказалась ниже или выше генеральной средней.

Напротив, весьма велика вероятность того, что в некоторых отобранных хозяйствах она выше генеральной средней, а в других – ниже, ввиду чего при определении выборочной средней величины урожайности произойдет более или менее полное взаимопогашение отклонений урожайности, и выборочная средняя окажется близка к генеральной. В этом проявляется действие закона больших чисел по отношению к выборке: *средняя (или доля, или иной показатель), полученная при обобщении достаточно большого количества случайных величин, практически перестает быть случайной и отражает уровень признака, сложившийся в генеральной совокупности.*

На основании теоремы, доказанной А.А. Ляпуновым, можно определить вероятность того или иного отклонения выборочных характеристик от характеристик генеральной совокупности при достаточно большом объеме выборки. Другими математиками и статистиками разработаны методы определения подобных вероятностей для разных видов выборки, в том числе и для выборок малого объема (>30).

Выборочный метод отличается от других видов несплошного наблюдения (монографических описаний, анкетного метода и метода основного массива, особенности которых изложены в гл. 3) двумя признаками:

1) заранее устанавливается, сколько единиц или какая часть единиц генеральной совокупности будет обследована, и

2) заранее определяется порядок отбора единиц, при котором выборочная совокупность в достаточной мере представляла бы (репрезентовала) генеральную совокупность.

7.2. Причины применения выборочного наблюдения

Имеется ряд причин, в силу которых во многих случаях выборочному наблюдению отдается предпочтение перед сплошным.

Из них наиболее существенны следующие:

Во-первых, экономия времени, материальных, трудовых и финансовых ресурсов в результате сокращения объема работы. Действительно, подвергнув обследованию один или два процента общего числа единиц совокупности, сокращают объем работы в первом варианте примерно в 100 раз, а во втором – в 50. Этим достигается сокращение времени исследования и одновременно снижение его стоимости. Например, для составления баланса, денежных доходов и расходов населения, для изучения денежного обращения, выявления дифференциации населения по уровню жизни, определения черты бедности и т.д. необходимы данные о бюджетах домохозяйств. Сбор этих данных осуществляется государственной статистикой, но один статистик в состоянии курировать ежедневные записи доходов, расходов, потребления не более чем в 20–25 домохозяйствах. Если бы решили собирать данные о бюджетах всех домохозяйств, то только для этой цели (не учитывая потребности последующей обработки) потребовалось бы примерно два миллиона статистиков.

Так что использование выборочного наблюдения является единственным экономически выгодным решением, тем более что по результатам изучения сравнительно небольшой части можно получить с достаточно высокой степенью уверенности данные о всей совокупности. Подобная ситуация возникает и при аудиторских проверках крупных фирм, когда вместо детального изучения каждого платежного документа ограничиваются анализом выборки документов, и в других областях применения статистики.

При громадном объеме статистических работ в нашей стране удешевление их стоимости и ускорение сроков их проведения имеют большое народнохозяйственное значение. Вот почему выборочный метод получил такое широкое распространение в статистических исследованиях.

Во-вторых, без выборки не обойтись, когда наблюдение связано с *порчей наблюдаемых объектов*. Это относится, прежде всего, к изучению качества продукции, которое основано на испытаниях образцов на вибрацию, упругость, разрыв и т.д. Всю продукцию, конечно же, таким испытаниям не подвергают, только отобранные образцы.

То же можно сказать об исследовании молока на жирность, зерна на содержание белка, влажность, чистоту и всхожесть семян, электрических лампочек – на длительность горения и т.д. На выборках основаны маркетинговые исследования, оценки качества поставок.

Естественно, что при подобного рода исследованиях применяется только выборочное наблюдение, так как сплошное наблюдение привело бы к бессмысленному уничтожению плодов человеческой деятельности.

В-третьих, при выборочном наблюдении объем работы в несколько раз меньше объема работы при сплошном наблюдении и вполне понятно, что применяя выборочное обследование, можно получить необходимые данные **значительно быстрее, оперативно**.

В-четвертых, при выборке обследованию подвергают сравнительно небольшую часть всей совокупности, поэтому создается возможность расширения **программы наблюдения**. Выборочное наблюдение часто применяется в тех случаях, когда нужно произвести обследование по весьма широкой программе, по которой проведение сплошного наблюдения чрезвычайно затруднено или вовсе невозможно. Например, для изучения бюджетов населения (доходов – по их источникам и расходов – по видам затрат), необходимо, чтобы в отдельных семьях велся точный ежедневный учет доходов и расходов. Невозможно, конечно, да и нет необходимости, чтобы каждая семья вела подобного рода подробные бюджетные записи.

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

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

Таким образом, выборочный метод обладает следующими достоинствами:

1. Экономия средств (материальных, трудовых, денежных).
2. Возможность проверки качества продукции, которая при этом уничтожается (проверка сплошным наблюдением здесь бессмысленна).
3. Оперативность получения результатов.
4. Возможность расширения программы наблюдения.
5. Высокая достоверность результатов.

7.3. Задачи, решаемые при применении выборочного метода

Применение выборочного наблюдения требует решения следующих основных задач:

❖ определение объема выборки, необходимого для получения требуемой точности результатов с заданной вероятностью (см. параграф 7.13).

❖ определение возможного предела ошибки репрезентативности, - гарантированного с заданной вероятностью, и сравнение его с величиной допустимой погрешности (см. параграф 7.7, решение задания 3 и 4; в параграфе 7.8 решение задания 5; в параграфе 7.10 решение задания 7);

❖ определение вероятности того, что ошибка выборки не превысит допустимой погрешности (см. решение заданий 3, 4, 5 и 7);

❖ оценка результатов выборочного наблюдения (см. параграф 7.14);

❖ распространение результатов выборочного наблюдения на генеральную совокупность (см. параграф 7.15)

7.4. Условия правильного применения выборочного наблюдения

Все достоинства выборочного наблюдения, которые рассматривали выше, проявляются лишь при условии правильного решения проблем выборочного обследования. К ним относятся:

1) определение границ генеральной совокупности;
2) разработка программы наблюдения и инструкций;
3) определение основы для проведения выборки – списка единиц генеральной совокупности, сведений об их размещении и т.д.;
4) установление допустимого размера погрешности и определение объема выборки;
5) определение вида выборочного наблюдения;
6) установление сроков проведения наблюдения;
7) определение потребности в кадрах для проведения выборочного наблюдения, их подготовка;
8) оценка точности и достоверности данных выборки, определение порядка их распространения на генеральную совокупность.

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

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

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

В выводной статистике принято строго различать *параметры* и свойства генеральной совокупности и их *оценки* по данным выборки.

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

Выборочная оценка	=	***Генеральный параметр***	±	***Ошибка наблюдения***	±	***Ошибка выборки***

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

7.5. Определение границ генеральной совокупности

Выборочный метод наблюдения согласно рекомендациям методологических положений по статистике включает следующие этапы:

- определение генеральной совокупности и единиц наблюдения, обладающих первичной информацией, необходимой для решения задач обследования;
- создание основы выборки (списка элементов совокупности);
- формирование выборочной совокупности путем отбора элементов основы;
- распространение собранных по выборке данных на генеральную совокупность.

Последний этап зависит от примененного способа отбора элементов в выборку и формулы оценивания характеристик генеральной совокупности по данным выборки.

Для использования единых программных средств и координации конкретных выборочных наблюдений независимо от отрасли экономической деятельности в системе государственной статистики, предусмотрен единый порядок формирования перечня подлежащих наблюдению объектов на основе информационного массива «База данных Генеральной совокупности».

База данных Генеральной совокупности – информационная система, созданная на основе Единого государственного реестра предприятий и организаций (ЕГРПО), банка данных

«Бухгалтерская отчетность организаций», данных статистического наблюдения и построенная с учетом единых методологических, программно-технологических и технических решений по всем уровням органов государственной статистики.

Генеральная совокупность объектов статистического наблюдения – перечень юридических лиц, их филиалов, представительств, других объектов статистического наблюдения, осуществляющих деятельность на территории РУз, характеризующийся установленным набором индивидуальных признаков, необходимых для организации конкретных статистических наблюдений.

Объект наблюдения в Генеральной совокупности – юридическое лицо, филиал, представительство, иной хозяйствующий субъект без права юридического лица, осуществляющий деятельность на территории РУз (субъекта РУз), подлежащий статистическому наблюдению и представляющий в органы государственной статистики статистическую и (или) бухгалтерскую отчетность.

Территориальный раздел Генеральной совокупности объектов статистического наблюдения – фрагмент Генеральной совокупности, содержащий перечень объектов статистического наблюдения, осуществляющих деятельность на территории субъекта *РУз.*

Отраслевой раздел Генеральной совокупности объектов статистического наблюдения – фрагмент Генеральной совокупности, представляющий совокупность объектов, систематизированных по основному виду деятельности (см. рис. 7.1).

Создание Генеральной совокупности связано с решением следующих задач:

✓обеспечение методологической основы для перехода к интегрированному принципу сбора отчетной информации и проведение статистического анализа по сопоставимому кругу объектов;

✓определение единого круга хозяйствующих субъектов, подлежащих статистическому наблюдению, по отраслям экономики;

✓установление стандартного описания предприятий и организаций, подлежащих статистическому наблюдению (для всех типов предприятий).

Рис. 7.1. **Границы и объекты генеральной совокупности** (ГС)

Цель организации выборочного наблюдения

Цель организации выборочного наблюдения заключается в том, чтобы обеспечить в пределах имеющихся средств и установленной степени точности результатов соблюдение баланса между затратами и надежностью полученных оценок. Надежность результатов выборочного наблюдения зависит, как и при любом другом наблюдении, от точности регистрации фактов, а также от того, насколько учтенные явления отличаются от неучтенных.

В связи с этим при использовании метода выборочного наблюдения необходима оценка **репрезентативности** (представительности) собранных выборочных данных относительно всей генеральной совокупности.

7.6. Способы отбора из ГС, обеспечивающие репрезентативность выборки. Виды выборки

Для того чтобы можно было по выборке делать вывод о свойствах генеральной совокупности, выборка должна быть *репрезентативной* (представительной), т.е. она должна полно и адекватно представлять свойства генеральной совокупности. Репрезентативность выборки может быть обеспечена только при *объективности отбора данных*.

Основным условием проведения выборочного обследования является предупреждение возникновения систематических (тенденциозных) ошибок, возникающих вследствие нарушения принципа равных возможностей попадания в выборку каждой единицы генеральной совокупности.

Предупреждение систематических ошибок достигается в результате применения научно обоснованных способов формирования выборочной совокупности.

Выборочная совокупность формируется по принципу массовых вероятностных процессов, без каких бы то ни было исключений от принятой схемы отбора; необходимо обеспечить относительную однородность выборочной совокупности или ее разделение на однородные группы единиц. При формировании выборочной совокупности должно быть дано четкое определение единицы отбора. Желателен приблизительно одинаковый размер единиц отбора, причем результаты будут тем точнее, чем меньше единица.

Практика применения выборочного метода в экономике статистических исследований использует следующие способы отбора единиц из генеральной совокупности:

1) *индивидуальный отбор* – в выборку отбираются отдельные единицы;

2) *групповой отбор* – в выборку попадают качественно однородные группы или серии изучаемых единиц;

3) *комбинированный отбор* как комбинация индивидуального и группового отбора.

Способы отбора определяются правилами формирования выборочной совокупности.

Различают следующие виды отбора совокупности единиц наблюдения:

- случайный отбор (жеребьевка, таблица случайных чисел);

- отбор единиц по какой-либо схеме (единицы упорядочивают таким образом, чтобы это было не связано с изучаемыми свойствами, далее проводят механический отбор единиц с шагом, равным *N: n)*. Обычно отбор начинают не с первой единицы, а отступив пол шага, чтобы уменьшить возможность смещения выборки;
- типическая;
- серийная;
- комбинированная.

Приведем более подробно характеристику отделных видов выборочного наблюдения. При этом на условных данных (табл. 7.2) покажем решение сквозных задач.

7.7. Случайный отбор

Собственно-случайная выборка (рис. 7.2) состоит в том, что выборочная совокупность образуется в результате случайного (непреднамеренного) отбора отдельных единиц из генеральной совокупности. При этом количество отобранных в выборочную совокупность единиц обычно определяется исходя из принятой доли выборки.

Доля выборки есть отношение числа единиц выборочной совокупности ***n*** к численности единиц генеральной совокупности ***N***, т.е.

$$\frac{n}{N} = K_в.$$

Так, при 5% выборке из партии товара в 2000 ед. численность выборки *n* составляет 100 ед. $\left(\frac{5 \cdot 2000}{100}\right)$, а при 20% выборке она составит 400 ед. $\left(\frac{20 \cdot 2000}{100}\right)$ и т.д.

Важным условием репрезентативности собственно-случайной выборки является то, что каждой единице генеральной совокупности предоставляется равная возможность попасть в выборочную совокупность. Именно принцип случайности попадания любой единицы генеральной совокупности в выборку предупреждает возникновение систематических (тенденциозных) ошибок выборки. Это представлено на рис. 7.2. Одним из примеров использования собственно-случайной выборки является проведение тиражей выигрышей денежно-вещевой лотереи, при которых обеспечивается равная возможность попадания в тираж любого номера лотерейного билета.

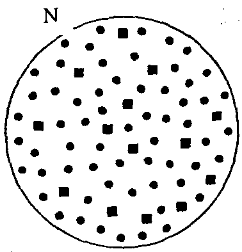


Рис. 7.1. **Собственно-случайная выборка:**
N – генеральная совокупность;
■ *– единицы, отобранные в выборку (в случайном порядке)*

Формирование собственно-случайной выборки обычно осуществляется с помощью специальных фишек. При этом все единицы генеральной совокупности нумеруются и каждый номер записывается на фишку (жребий) одинаковой формы. Фишки тщательно перемешиваются и отбираются в выборку по одной.

Порядок применения этих видов отбора покажем на примере решения сквозных задач.

Задание 1

Пусть, например, в сборочном цехе машиностроительного завода работают 300 рабочих, объединенных в 30 бригад, по 10 человек в каждой бригаде (см. табл. 7.2, сведения о разрядах и порядковых номерах рабочих при группировке по разрядам будут использованы при рассмотрении типического отбора).

Поставлена задача определить средний стаж рабочих и долю (процент) рабочих со стажем более 10 лет.

Общая сумма проработанных всеми рабочими человеко-лет равна 2 700, а средний стаж работы составляет x = 2 700 : 300 = 9 (лет).

Рабочих со стажем свыше 10 лет – 93 человека, а доля их в общей численности рабочих p = 93 : 300 = 0,31, или 31 %.

Теперь посмотрим, какие бы мы получили результаты, если бы, не имея полных данных, взялись за решение поставленной задачи на основе десятипроцентной собственно-случайной повторной выборки. Для этого необходимо было бы проделать следующее.

Таблица 7.2

Производственный стаж и квалификация рабочих сборочного цеха

Табельный номер	Стаж работы	Разряд	№ п/п	Табельный номер	Стаж работы	Разряд	№ п/п	Табельный номер	Стаж работы	Разряд	№ п/п
А	I	2	3	А	1	2	3	А	1	2	3
1	30	6	241	51	2	2	24	101	3	2	49
2	J3	4	151	52	2	1	25	102	32	6	264
3	6	2	1	53	18	5	253	103	2	1	50
4	5	2	2	54	7	3	166	104	1	1	51
5	11	3	15?	55	19	4	167	105	16	5	265
6	19	6	242	56	6	2	26	106	1	1	52
7	11	3	153	57	28	5	254	107	2	1	53
8	1	1	3	58	6	2	27	108	3	2	54
9	12	4	154	59	3	2	28	109	7	3	180
10	21	5	243	60	11	6	255	110	12	4	181
11	3	1	4	61	1	1	29	111	30	6	266
12	11	3	155	62	6	3	168	112	5	4	182
13	9	2	5	63	20	4	169	113	13	3	183
14	21	5	244	64	5	2	30	1Г4	1	1	55
15	21	5	245	65	33	6	256	115	4	2	56
16	2	1	6	66	10	3	170	116	5	1	57
17	10	3	156	67	2	1	31	117	4	3	184
18	19	4	157	68	2	2	32	118	7	2	58
19	6	2	7	69	6	3	171	119	6	4	185
20	18	5	246	70	1	1	33	120	8	2	59

21	1	1	8	71	7	3	172	121	15	4	186
25	7	2	9	72	11	4	173	122	24	5	267/
23	21	5	247	73	28	6	257	123	2	1	60
24	2	1	10	74	19	5	258	124	7	2	61
25	14	3	158	75	17	4	174	125	1	1	62
26	9	3	159	76	10	3	175	126	3	2	63
27	3	1	11	77	7	2	34	127	3	1	64
28	4	2	12	78	14	6	259	128	5	3	187
29	19	5	248	79	9'	2	35	129	20	4	188
30	1	1	13	80	8	2	36	130	2	1	65
31	10	3	160	81	22	5	260	131	4	2	66
32	8	2	14	82	1	1	37	132	5	2	67
33	17	6	249	83	9	2	38	133	4	3	189
34	5	2	15	84	10	3	176	134	3	1	68
35	1	1	16	85	2	1	39	135	1	1	69
36	25	5	250	86	5	1	40	136	21	5	268
37	1	1	17	87	14	4	177	137	19	4	190
38	1	2	18	88	1	1	41	138	4	1	70
39	6	3	161	89	19	5	261	139	4	3	191
40	4	1	19	90	10	2	42	140	10	4	192
41	28	6	251	91	9	2	43	141	5	2	71
42	18	5	252	92	1	1	44	142	6	3	193
43	2	2	20	93	3	2	45	143	7	2	72
44	1	1	21	94	1	1	46	144	1	1	73
45	3	3	162	95	4	3	178	145	18	4	194
46	17	4	163	96	1	1	47	146	9	2	74
47	1	1	22	97	9	6	262	147	2S	6	269
48	10	3	164	98	3	4	179	148	12	5	270
49	13	4	165	99	2	1	48	149	1	1	75
50	1	2	23	100	6	5	263	150	5	4	195
151	7	2	76	201	13	4	211	251	4	2	119
152	9	3	196	202	5	2	101	252	5	2'	120
153	14	6	271	203	12	4	212	253	8	4	234
154	9	3	197	204	1	1	102	254	11	6	290
155	6	2	77	205	6	3	213	255	21	5	291
156	2	1	78	206	18	4	214	256	10	4	235
157	2	1	79	207	4	2	103	257	11	4	236
158	5	3	198	208	23	6	281	258	3	2	121
159	8	2	80	209	2	1	104	259	3	1	122
160	10	4	199	210	2	1	105	260	2	1	123
161	10	2	81	211	1	1	106	261	3	2	124
162	3	1	82	212	1	1	107	262	1	1	125

163	32	5	272	213	8	3	215	263	7	2	126
164	9	6	273	214	7	3	216	264	10	4	237
165	19	4	200	215	9	4	217	265	4	2	127
166	1	1	83	216	16	4	218	266	17	6	292
167	4	3	201	217	3	1	108	267	23	6	293
168	18	4	202	218	22	5	282	268	5	2	128
169	7	2	84	219	5	2	109	269	3	1	129
170	2	1	85	220	6	3	219	270	4	2	130
171	3	2	86	221	2	1	110	271	22	5	294
172	19	4	203	222	3	2	111	272	23	5	295
173	11	5	274	223	8	3	220	273	8	2	131
174	2	1	87	224	15	4	221	274	7	2	132
175	30	6	275	225	24	6	283	275	18	6	296
176	7	3	204	226	13	4	222	276	4	1	133
177	9	4	205	227	5	2	112	277	11	3	238
178	4	2	88	228	1	1	113	278	6	2	134
179	5	2	89	229	8	3	223	279	6	2	135
180	1	1	90	230	5	2	114	280	3	1	136
181	7	2	91	231	5	2	115	281	5	2	137
182	11	5	276	232	6	3	324	282	8	2	138
183	21	5	277	233	15	4	225	283	4	1	139
184	3	1	92	234	7	3	226	284	17	5	297
185	6	2	93	235	19	5	284	285	7	2	140
186	8	4	206	236	21	5	285	286	5	2	141
187	3	1	94	237	6	5	227	287	8	2	142
188	10	5	278	238	8	3	228	288	21	6	298
189	1	1	95	239	10	4	229	289	7	2	143
190	6	2	96	240	14	5	286	290	3	1	144
191	5	3	207	241	15	6	287	291	6	2	145
192	1	1	97	242	5	2	116	292	17	5	299
193	11	5	279	243	8	3	230	293	5	2	146
194	21	6	280	244	9	4	231	294	5	2	147
195	18	4	208	245	9	4	232	295	15	4	239
196	2	1	98	246	22	6	288	296	20	6	300
197	4	2	99	247	12	5	288	297	13	3	240
198	8	3	209	248	5	2	117	298	4	1	148
199	14	4	210	249	3	1	118	299	3	1	149
200	4	1	100	250	9	3	233	300	8	2	150

1. Составить список рабочих цеха по порядку табельных номеров (от 1 до 300 номера) с проставлением стажа.

2. Изготовить из бумаги одинакового размера билеты (жребии), надписать их номерами от 1 до 300, проставить на них соответствующий стаж рабочего, свернуть билеты в трубочки и вложить в специальные патроны.

3. Поместить патроны с билетами в ящик (урну) и тщательно перемешать их.

4. Вынуть из урны один билет, записать проставленный на нем стаж, после чего положить его снова в урну, опять перемешать билеты, после чего вынуть второй билет и т.д., а всего вынуть таким образом 30 билетов (10% от 300).

Рабочие, чьи табельные номера будут значиться на вынутых билетах, **составят выборочную совокупность**. Как видно из этого, при случайном отборе для каждой единице создаются условия равной возможности, попасть в выборку в каждом отдельном акте отбора. Благодаря этому обобщающие показатели выборочной совокупности должны довольно точно воспроизводить обобщающие показатели генеральной совокупности.

Задание 2

Решение:

Допустим, что в выборку попали рабочие, имеющие следующие табельные номера: 116, 54, 40, 177, 12, 37, 145, 42, 127, 84, 205, 271, 226, 256, 297, 147, 93, 103, 85, 154, 1-2, 200, 81, 186, 1, 103, 275, 300, 230, 222.

Стаж рабочих, попавших в выборку (лег): 5, 7, 4, 9, 11, 1, 18, 18, 3, 10, 6, 22, 13, 10, 13, 23, 3, 2, 2, 9, 11, 4, 20, 8, 30, 2, 18, 8, 5, 3.

Сумма этих показателей составляет 300 лет, средний стаж рабочих в выборочной совокупности равен $\bar{x}$ = 300: 30 = 10 (лет).

Так как в выборку попало одиннадцать рабочих со стажем свыше 1,0 лет, то выборочная доля составляет: w – 11 :30 = 0,367, или 36,7%.

Как видим, выборочная средняя отличается от генеральной средней, а выборочная доля – от генеральной доли.

При собственно-случайном отборе выборочная средняя, равно как и выборочная доля, являются переменными величинами. Они могут принимать различные значения при том или ином исходе выборки с той или иной вероятностью, колеблясь, соответственно, около значений генеральной средней или доли. Мерой колеблемости возможных значений выборочной средней около генеральной средней, а выборочной доли около генеральной доли является *дисперсия,* т.е. *средний квадрат отклонений.* Обозначив эту величину через μ^2 (греческая буква «ми» в квадрате), имеем:

$$\mu_{\bar{x}}^2 = \frac{\Sigma(\bar{x}_i - \bar{x})^2}{n} \quad \text{и} \quad \mu_p^2 = \frac{\Sigma(w_i - p)^2}{n}.$$

В математической статистике доказывается, что для случайной повторной выборки между дисперсией выборочной средней (доли) и генеральной дисперсией существует следующее соотношение: *дисперсия выборочной средней (доли) равна дисперсии признака в генеральной совокупности, деленной на число отобранных единиц либо,* что то же, на объем выборки:

$$\mu_{\bar{x}}^2 = \frac{\sigma^2}{n} \quad \text{и} \quad \mu_p^2 = \frac{p(1-p)}{n},$$

где σ^2 – генеральная дисперсия признака; p – доля признака в совокупности с альтернативными признаками; $p(1-p)$ – дисперсия альтернативного признака.

Корень квадратный из этих выражений носит название средней ошибки выборки. Так, средняя ошибка при выборочном определении средней $\mu_{\bar{x}} = \sqrt{\frac{\sigma^2}{n}}$ (1). Соответственно, средняя ошибка при выборочном определении доли:

$$\mu_p = \sqrt{\frac{p(1-p)}{n}} \quad (2).$$

Из формул средней ошибки выборки видно, что она прямо пропорциональна среднему квадратическому отклонению признака. Поэтому, чем больше колеблемость значений признака в генеральной совокупности, тем больше средняя ошибка выборки и, наоборот, с уменьшением колеблемости значений признака уменьшается и размер возможной ошибки выборки. Эти же формулы показывают, что *средняя ошибка выборки обратно пропорциональна корню квадратному из числа наблюдений* (объем выборки). Поэтому по мере возрастания объема выборки размер средней ошибки выборки уменьшается. Если, например, требуется уменьшить среднюю ошибку выборки в 2 раза, то для этого при прочих равных условиях необходимо увеличить объем выборки в 4 раза, чтобы уменьшить среднюю ошибку выборки в 3 раза, объем выборки следует увеличить в 9 раз и т.д.

Применение приведенных формул средней ошибки выборки предполагает, что известна генеральная дисперсия и, следовательно, генеральная средняя и доля. Однако в действительности эти показатели неизвестны, так как выборка для того и проводится, чтобы на основе выборочных показателей судить о неизвестных величинах соответствующих генеральных показателей. Математической статистикой доказано, что выборочная дисперсия в среднем несколько меньше генеральной и что их соотношение можно выразить следующим образом:

$$\sigma^2 = \sigma_0^2 \frac{n}{n-1}.$$

где σ_0^2 – выборочная дисперсия; n – объем выборки.

Нетрудно заметить, что по мере возрастания объема выборки поправочный коэффициент $\frac{n}{n-1}$ стремится к единице и расхождения между генеральной и выборочной дисперсиями становятся все меньшими. Например, если $n=10$, то $\frac{n}{n-1} = \frac{10}{9} = 1{,}11$; $n=100$ $\frac{n}{n-1} = \frac{100}{99} = 1{,}01$ и т.д.

Следовательно, если для малой выборки величина – имеет существенное значение и должна быть учтена (см. п. 3 данной главы), то для большого объема выборки можно полагать, что σ_0^2 приблизительно равна σ^2.

Заменив в приведенных формулах показатели генеральной совокупности показателями выборочной совокупности, получим следующие приближенные значения средней ошибки выборки:

$$\mu_{\bar{x}} = \sqrt{\frac{\sigma^2}{n}} \quad (1) \text{ и } \quad \mu_p = \sqrt{\frac{w(1-w)}{n}} \quad .(3)$$

Средняя ошибка выборки характеризует меру отклонений выборочной средней (или доли) от генеральной средней (или доли). При этом, как доказывается в математической статистике, с определенной вероятностью можно утверждать, что эти отклонения не превысят некоторую величину, которую можно назвать *предельной ошибкой выборки.* Обозначив предельную ошибку греческой буквой Δ (дельта), а буквой t – *коэффициент доверия,* зависящий от вероятности, с которой можно гарантировать, что предельная ошибка выборки не превысит t – кратную среднюю ошибку, можно написать следующее равенство:

$$\Delta = t\mu.$$

При t =1 предельная ошибка (Δ) обращается в среднюю ошибку (μ). Формулы расчета *предельной ошибки выборки* представлены в табл. 7.3.

Таблица 7.3

Формулы расчета предельной ошибки (Δ) выборки

Порядок выборки	**Допустимая ошибка** при определении среднего размера изучаемого признака	**Допустимая ошибка** при определении доли данного признака
Повторная выборка	$\Delta\tilde{x} = t\sqrt{\frac{\bar{\sigma}^2}{n}}$ (5)	$\Delta\omega = t\sqrt{\frac{\omega(1-\omega)}{n}}$ (6)
Бесповторная выборка	$\Delta\tilde{x} = t\sqrt{\frac{\bar{\sigma}^2}{n}\left(1-\frac{n}{N}\right)}$ (7)	$\Delta\omega = t\sqrt{\frac{\omega(1-\omega)}{n}\left(1-\frac{n}{N}\right)}$ (8)

Положение о том, что по своему абсолютному значению разность между выборочными и генеральными обобщающими показателями с определенной вероятностью не превысит предельной ошибки выборки, вытекает из существа закона больших чисел. Выдающемуся русскому математику П.Л. Чебышеву принадлежит следующая обобщающая формулировка закона больших чисел (применительно к выборке): с вероятностью, сколь угодно близкой к единице, можно утверждать, что при достаточно большом объеме выборки в ограниченной генеральной дисперсии выборочные обобщающие показатели (средняя, доля) «будут сколь угодно мало отличаться от соответствующих» генеральных обобщающих показателей.

Применительно к нахождению среднего значения признака теорема П.Л. Чебышева (с уточнениями А. М. Ляпунова) может быть записана так:

$$P\left[\left|\tilde{x}-\bar{x}\right| \le \Delta\right] = F(t),$$

а для доли признака, соответственно,

$$P\left[\left|\omega - p\right| \le \Delta\right] = F(t)$$

где

$$F(t) = \frac{1}{2\pi}\int_{-t}^{+t} e^{-\frac{Z_2}{2}} d(z).$$

Значения функции *F(t)* и *t* определяются на основе специально составленных математических таблиц. Приведем некоторые из них, применяемые наиболее часто (табл. 7.4).

Таблица 7.4

Уровни вероятности и соответствующие им значения коэффициента кратности средней ошибки выборки

Уровни вероятности, Ф(t)	0,683	0,954	0,997
Коэффициент кратности средней ошибки выборки, t	1,00	2,00	3,00

Таким образом, предельная ошибка выборки отвечает на вопрос о точности выборки с определенной вероятностью, величина которой зависит от значения коэффициента доверия

t. Так, при t = 1 вероятность *F(t)* отклонения выборочных характеристик от генеральных на величину однократной средней ошибки равна 0,683. Следовательно, в среднем из каждой 1000 выборок 683 дадут обобщающие показатели (среднюю, долю), которые будут отличаться от генеральных не более чем на величину однократной средней ошибки. При $t = 2$ вероятность *F(t)* равна 0,954, это означает, что из каждой 1000 выборок 954 дадут обобщающие показатели, которые будут отличаться от генеральных не более чем на двукратную среднюю ошибку выборки и если при расчете предельной ошибки выборки мы используем значение $t = 3$, то с вероятностью 0,997 можно утверждать, что расхождение между средней выборочной совокупности и средней генеральной совокупности не превысит трехкратной величины средней ошибки выборки.

• для средней $\Delta_{\tilde{x}}\% = \dfrac{\Delta\tilde{x}}{\bar{x}} \cdot 100\%;$

• для доли $\Delta_{\omega}\% = \dfrac{\Delta\omega}{\omega} \cdot 100\%.$

На практике принято задавать величину Δ, как правило, в пределах 10% предполагаемого среднего уровня признака.

Расчет средней и предельной ошибок выборки позволяет определить пределы, в которых будут находиться характеристики генеральной совокупности:

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

$$w - \Delta_{w} \le P \le w + \Delta_{w}$$

Пределы, в которых с данной степенью вероятности будет заключена неизвестная величина изучаемого показателя в генеральной совокупности, называют *доверительным интервалом,* а вероятность *F(t) – доверительной вероятностью.* Чем выше значение Δ, тем больше величина доверительного интервала и, следовательно, ниже точность оценки.

На основании теоремы Чебышева решается ряд задач выборочного наблюдения, в частности, определение предельной ошибки выборки при заданной вероятности и определение численности выборки, необходимой для обеспечения заданной ее точности с определенной вероятностью. Рассмотрим это на примере 10% собственно-случайной повторной выборки.

Задание 3

Решение:

По приведенным выше данным о стаже 30 рабочих выборочная дисперсия $\hat{\sigma}^2$ составляет 54,87 и средняя ошибка выборки по формуле (1) равна $\mu_{\tilde{x}} = \sqrt{\dfrac{54,87}{30}} = 1,35$ (года).

Из формулы (5) следует, что предельные значения генеральной средней можно определить по формуле: $\bar{x} = \tilde{x} \pm \Delta$. Отсюда с вероятностью 0,683 (t=1) можно утверждать, что средний стаж всех рабочих цеха находится в пределах 10± 1,35, т.е. от 8,65 до 11,35 лет. С вероятностью 0,954 (t=2) можно утверждать, что средний стаж всех рабочих цеха находится в пределах 10±2,7 т.е. от 7,3 до 12,7 лет и т.д.

Средняя ошибка выборочной доли по формуле $\mu_p = \sqrt{\dfrac{w(1-w)}{n}}$ составляет:

$$\mu_p = \sqrt{\frac{0,367 \times 0,633}{30}} = 0,088$$

(0,367 – выборочная доля рабочих со стажем свыше 10 лет).

Следовательно, с вероятностью 0,683 (t=1) можно утверждать, что для рабочих цеха со стажем свыше 10 лет находится в пределах 0,367±0,088, т.е. от 0,279 до 0,455, или от 27,9 до 45,5%.

Определение необходимой численности выборки производился на основе алгебраического преобразования формулы Δ=tμ. Из этой формулы следует, что при определении средней

$$\Delta = t\mu_{\tilde{x}} = t\sqrt{\frac{\sigma_0^2}{\bar{n}}}.$$

Отсюда

$$n = \frac{t^2\sigma_0^2}{\Delta^2} \quad (9)$$

Если выборка еще не производилась, а лишь планируется, то выборочная дисперсия, конечно, неизвестна. В таких случаях при использовании вышеуказанной формулы берутся ориентировочные значения σ_0^2, полученные при ранее проведенных аналогичных обследованиях или на основе пробной выборки.

Задание 4

Решение:

Обращаясь к рассматриваемому примеру, определим, какой должен быть объем выборки, чтобы с вероятностью 0,954 предельная ошибка выборки не превышала двух лет. По формуле (9) получаем:

$$n = \frac{2^2 \times 54,87}{2^2} = 55$$ человек.

При определении доли имеем $\Delta = t\mu_p = t\sqrt{\frac{w(1-w)}{n}}$, откуда необходимая численность выборки равна:

$$n = \frac{t^2 w(1-w)}{\Delta^2}.$$

Вычислим по этой формуле численность выборки, которая необходима для того, чтобы с вероятностью 0,954 предельная ошибка выборки не превышала 0,15.

Подставляя в формулу условия задания, получаем величину необходимой численности выборки:

$$n = \frac{2^2 \times 0{,}367 \times 0{,}0633}{0{,}15^2} = 41$$ человек.

Из двух полученных значений численности выборки следует взять максимальную величину, т.е. 55 человек, ибо только в этом случае будет обеспечена заданная точность выборки как для средней, так и для доли, т.е. по обоим интересующим нас показателям. Образуем теперь выборочную совокупность способом собственно-случайного *бесповторного отбора* притом же объеме выборки (30 единиц, или 10% к общей численности единиц). В этом случае, так же как и при повторном отборе, можно применить жеребьевку, но каждый билет после записи проставленного на нем стажа в урну не возвращается.

Допустим, что в выборку попали рабочие, имеющие следующие табельные номера:[1] 101, 90, 69, 194, 267, 270, 61, 287, 86, 70, 114, 276, 9, 32, 1, 30, 72, 74, 85, 237, 66, 12, 291, 258, 281, 148, 95, 15, 112, 27.

Эти рабочие имеют следующий стаж работм (лет): 3, 10, 6, 21, 23, 4, 1, 8, 5, 1, 1, 4, 12, 8, 30, 1, 11, 19, 2, 6, 10, 11, 6, 3, 5, 12, 4, 21, 5, 3.

Сумма этих показателей составляет 256 лет, а средний стаж рабочих в выборочной совокупности равен: $\bar{x}$ = 256:30 = 8,53 (года). Так как в выборку попали девять рабочих со стажем свыше 10 лет, то выборочная доля составляет: w = 9:30 = 0,30, или 30%. В математической статистике доказывается, что для бесповторного отбора формулы средней ошибки собственно-случайной выборки имеют следующий вид:

$$\mu_{\bar{x}} = \sqrt{\frac{\sigma^2}{n}\left(1-\frac{n}{N}\right)}, \quad (10)$$

$$\mu_p = \sqrt{\frac{w(1-w)}{n}\left(1-\frac{n}{N}\right)}, \quad (11)$$

где N – объем генеральной совокупности.

Сравнение формул (10) и (11) с формулами (3) и (4) показывает, что они отличаются лишь множителем $\left(1-\frac{n}{N}\right)$, содержащимся под знаком корня. Так как этот множитель всегда меньше единицы, то ошибка выборки при бесповторном отборе меньше ошибки выборки повторного отбора. По мере увеличения объема выборки множитель $1-\frac{n}{N}$ стремится к нулю, а для *n* = *N*, т.е. когда будет подвергнута обследованию

[1] Для отбора необходимой численности единиц совокупности могут быть использованы и таблицы случайных чисел. См. Венецкий И.Г., Кильдишев Г.С. Основы математической статистики. М.: Госстатиздат, 1963.; Романовский В.И. Применение математической статистики в опытном деле. 1947. и др.

вся совокупность единиц, множитель $1-\frac{n}{N}$ превращается в нуль и вместе с ним случайная ошибка репрезентативности становится равной нулю. Следует отметить, что при бесповторном отборе часто применяются формулы ошибок повторного отбора. Эта замена находит свое оправдание для случаев, когда выборка составляет малую долю генеральной совокупности. Тогда отношение $n:N$ становится очень малой величиной и, следовательно, разность $1-\frac{n}{N}$ будет мало отличаться от единицы.

Для рассматриваемого примера бесповторного отбора выборочная дисперсия составляет 53,77, откуда средняя ошибка выборки при определении стажа равна:

$$\mu_{\bar{x}} = \sqrt{\frac{53,77}{30}\left(1-\frac{30}{300}\right)} = 1,27 \text{ (лет).}$$

Соответственно, средняя ошибка доли равна:

$$\mu_p = \sqrt{\frac{0,3\times 0,7}{30}\left(1-\frac{30}{300}\right)} = 0,079.$$

Определение предельной ошибки выборки, т.е. отклонения среднего размера признака и доли в выборочной совокупности в ту или другую сторону от генеральной средней и доли с заданной вероятностью, а также определение необходимой численности выборки производится аналогично тому, как это было показано применительно к повторному отбору.

Поставив, например, в формулу (10) значение средней ошибки $\mu_{\bar{x}}$ получим:

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

Откуда необходимая численность выборки при определении средней равна:

$$n = \frac{t^2\sigma_0^2 N}{\Delta^2 N + t^2\sigma^2}.$$

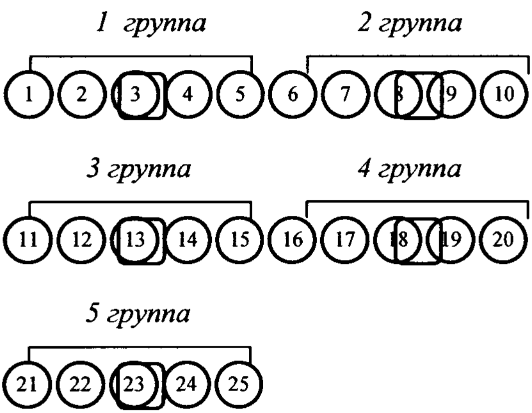
При определении доли численность бесповторной выборки составляет:

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

7.8. Механический отбор

Механическая выборка состоит в том, что отбор единиц в выборочную совокупность производится из генеральной совокупности, разбитой на равные интервалы (группы). При этом размер интервала в генеральной, совокупности равен обратной величине доли выборки. Так, при 2% выборке отбирается каждая 50 единица (1:0,02), при 5% выборке – каждая 20 единица (1 : 0,05) и т.д.

Таким образом, в соответствии с принятой долей отбора генеральная совокупность как бы механически разбивается на равновеликие группы. Из каждой такой группы в выборку отбирается лишь одна единица (рис. 7.3).



- единицы генеральной совокупности, отобранные в выборку через равный интервал (20% выборка).

Размер интервала равен обратной величине объема выборки (1: 0,2 = 5).

Рис. 7.3. **Механическая выборка**

Для обеспечения репрезентативности (представительности) выборки все единицы генеральной совокупности должны располагаться в определенном порядке. При этом по отношению к изучаемому показателю единицы генеральной совокупности могут быть упорядочены по существенному, второстепенному или нейтральному признаку. Это важно для установления порядка отбора единиц в выборку.

При упорядочении генеральной совокупности по существенному признаку, т.е. по признаку, который всецело определяет поведение изучаемого показателя, в выборочную совокупность должна отбираться та единица, которая находится в середине каждой группы (рис. 7.3). Это позволяет избежать появления систематической ошибки выборки.

Задача 5

Решение:

Пусть, например, необходимо произвести 10% механическую выборку рабочих сборочного цеха (см. табл. 7.2) для определения среднего стажа и доли рабочих со стажем более 10 лет. Объем выборки равен $\frac{300\times10}{100}=30$ человек. Величина интервала составляет 10 человек (300:30).

Следовательно, в нашем примере, где имеется 30 бригад по 10 человек в каждой, необходимо отобрать по одному рабочему из каждой бригады. Допустим, путем жеребьевки определили, что это будет седьмой рабочий. Тогда в выборочную совокупность войдут 30 рабочих, табельные номера которых будут: 7, 17, 27, 37, 47, 57, 67, 77, 87, 97, 107,117, 127, 137,147, 157, 167, 177, 187, 197, 207, 217, 227, 237, 247, 257, 267, 277, 287 и 297.

Эти рабочие имеют следующий общий производственный стаж (лет): 11, 10, 3, 1, 1, 28, 2, 7, 14, 19, 2, 4, 3, 19, 23, 2, 4, 9, 3, 4, 4, 3, 5, 6, 12, 11, 23, 11, 8 и 13. Средний стаж рабочих в выборочной совокупности составляет 8,83 года, дисперсия стажа – 48,9 и доля рабочих со стажем свыше 10 лет – 36,7%.

Широкое распространение в практике статистических работ имеет механическая выборка на основе предварительного расположения единиц генеральной совокупности по возрастающему (или убывающему) значению изучаемого или связанного с ним признака. В этих случаях величина интервала обычно определяется путем деления общего итога по какому-либо связанному с изучаемым показателю на число единиц, подлежащих отбору. Механический отбор имеет определенные преимущества перед собственно-случайной выборкой, так как дает обычно более близкое распределение отобранных единиц к распределению единиц в генеральной совокупности по изучаемым признакам, и выборка становится более репрезентативной. Кроме того, при механическом отборе проще организовать и легче проверить правильность отбора единиц.

Оценку точности выборки при механическом отборе производят по формулам собственно-случайной выборки. Это объясняется тем, что средняя ошибка выборки при механическом отборе меньше либо, в крайнем-случае, равна средней ошибке собственно-случайной выборки. При этом более вероятно, что ошибка репрезентативности механической выборки будет не больше, чем ошибка собственно-случайной бесповторной выборки. Но для большей уверенности в достоверности полученных результатов часто применяют формулу ошибки собственно-случайной повторной выборки.

Для рассматриваемого примера средняя ошибка выборки составляет: а) для среднего стажа работы:

$$\mu_{\bar{x}} = \sqrt{\frac{48{,}9}{30}} \sqrt{1{,}63} = 1{,}28 \quad \text{(года);}$$

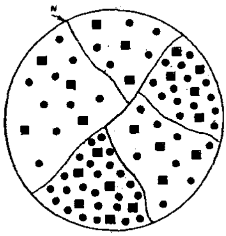
б) для доли рабочих со стажем свыше 10 лет:

$$\mu_p = \sqrt{\frac{0{,}367 \times 0{,}633}{30}} = 0{,}088$$

Следовательно, с вероятностью 0,683 (t=1) можно утверждать, что средний стаж всех рабочих цеха находится в пределах 8,8±1,28, т.е. от 7,52 до 10,08 года. В отношении доли рабочих со стажем свыше 10 лет с вероятностью 0,683 (t=1) можно утверждать, что она находится в пределах 0, 367 = 0,088, т.е. от 0,279 до 0, 455, или от 27,9 до 45,5%.

7.9. Типическая выборка

В случае, когда генеральная совокупность не однотипна и это влияет на размер научаемого признака, применяется предварительное деление ее на *типически однородные группы*. Группировка производится по таким признакам, которые связаны с изучаемыми признаками. Затем из каждой типической группы собственно-случайной или механической выборкой производится индивидуальный отбор единиц в выборочную совокупность. Это представлено на рис. 7.4.



N – генеральная совокупность расчленена на однородные группы.

Внутри групп отбираются единицы

■ в случайном порядке (или механическим способом отбора)

Рис. 7.4. **Типическая выборка** (при пропорциональном отборе)

Образованные таким образом группы иногда неравны между собой по объему, поэтому отбор обычно производится пропорционально объему групп. При типическом отборе в выборку попадают представители всех типических групп, поэтому достигается большая репрезентативность и большая точность выборки, чем при случайном или механическом отборе. При этом точность выборки, зависит от того, насколько хорошо отобранные единицы репрезентируют среднюю (или долю) каждой типической группы.

Задание 6

Решение:

Если, например, из 300 рабочих сборочного цеха, по которому ранее определялся средний стаж работы, 150 человек имеют низкую квалификацию (1 и 2 разряды), 90 человек – среднюю квалификацию (3 и 4 разряды) и 60 человек – высокую квалификацию (5 и 6 разряды), и если установлено, что квалификация рабочих тесно связана со стажем, то выбрав 30 человек путем случайного отбора или механически, мы рискуем получить недостаточно точную характеристику среднего стажа. Это может произойти в том случае, когда, допустим, в выборку попадут главным образом рабочие низкой и средней квалификации и очень мало рабочих высокой квалификации или, наоборот, когда в выборку попадет много рабочих высокой и средней квалификации и мало рабочих низкой квалификации.

Чтобы добиться в данном случае более точных результатов, поделим предварительно всех рабочих цеха (генеральная совокупность) на группы по квалификации, а затем уже из каждой такой группы будем производить выборку необходимого числа рабочих способом механического отбора. При такой организации отбора выборочная совокупность будет лучше репрезентировать генеральную совокупность, так как отбор из отдельных типических групп обеспечивает пропорциональное попадание в выборку рабочих разной квалификации.

По каждой типической группе составим список рабочих в порядке увеличения табельных номеров (порядковые номера рабочих в этих списках приведены в гр. 3 табл. 7.2). Внутри каждой типической группы произведем механическую группировку по десяткам и в каждом десятке отберем второго по счету рабочего.

Тогда из первой типической группы численностью в 150 человек в выборку попадут 15 рабочих, имеющих порядковые номера: 2, 12, 22, 32, 42, 52, 62, 72, 82, 92, 102, 112, 122, 132 и 142.

Найдем для этой группы средний стаж и дисперсию стажа:

$\Sigma x_1 = 5 + 4 + 1 + 2 + 10 + 1 + 1 + 7 + 3 + 3 + 1 + 1 + 3 + 7 + 8 = 57;$

$\Sigma x_1^2 = 25 + 16 + 1 + 4 + 100 + 1 + 1 + 49 + 9 + 9 + 1 + 1 + 9 + 49 + 64 = 339.$

Средний стаж $\bar{x}_1 = 57:15 = 3{,}8$ года;

дисперсия $\sigma_{01}^2 = \frac{339}{15} - 3{,}8^2 = 22{,}6 - 14{,}4 = 8{,}2.$

Так как в выборке нет ни одного рабочего со стажем свыше 10 лет, то выборочная доля по этому признаку равна нулю, точно так же нулю равна и дисперсия по этому признаку.

Из второй типической группы численностью в 90 человек в выборку попадут 9 рабочих, имеющих порядковые номера: 152, 162, 172, 182, 192, 202, 212, 222 и 232.

Для этой группы средний стаж ($\bar{x}_2$) составит 9,78 года, а дисперсия (σ_{02}^2) – 19,9.

Так как в выборку попало 4 рабочих со стажем свыше 10 лет, то их доля в выборке равна:

$w_2 = 4:9 = 0{,}444$, или 44,4%, а дисперсия – соответственно, $w_2(1 - w_2) = 0{,}444 \times (1 - 0{,}444) = = 0{,}247 \approx 0{,}25$.

Наконец, из третьей типической группы численностью 60 человек в выборку попадут 6 рабочих, имеющих порядковые номера: 242, 252, 262, 272, 282 и 292.

Для этой группы, соответственно, получим:

- средний стаж $\bar{x}_3 = 21{,}17$ года;
- дисперсия стажа $\sigma_{03}^2 = 25{,}66$, $w_3 = 1$ и $w_3(1 - w_3) = 0$.

В целом средний стаж рабочих, вошедших в выборочную совокупность, составит:

$$\tilde{x} = \frac{57 + 88 + 127}{15 + 9 + 6} = \frac{272}{30} = 9{,}07 \text{ (года)},$$

а доля рабочих со стажем свыше 10 лет будет:

$$w = \frac{0 + 4 + 6}{15 + 9 + 6} = \frac{10}{30} = 0{,}33, \text{ или } 33\%.$$

Расчет средней ошибки выборки при пропорциональном отборе численности единиц в типических группах, производится с помощью следующих формул (табл. 7.5).

Таблица 7.5

Формулы расчета средней ошибки при типическом отборе

	Повторная выборка	Бесповторная выборка
При определении среднего размера изучаемого признака	$\mu_{\bar{x}} = \sqrt{\frac{\overline{\sigma_i^2}}{n}}$	$\mu_x = \sqrt{\frac{\overline{\sigma_i^2}}{n}\left(1-\frac{n}{N}\right)}$
При определении доли признака	$\mu_p = \sqrt{\frac{\overline{w_i(1-w_i)}}{n}}$	$\mu_p = \sqrt{\frac{\overline{w_i(1-w_i)}}{n}\left(1-\frac{n}{N}\right)}$

Сравнив формулы средней ошибки выборки типического отбора с соответствующими формулами собственно-случайного отбора, увидим, что они отличаются тем, что вместо дисперсий по всей выборочной совокупности [σ_0^2 и $w_i(1 - w_i)$] взяты средние величины из соответствующих показателей вариации по группам: $\overline{\sigma_i^2}$ и $\overline{w_i(1-w_i)}$.

Согласно закону сложения дисперсии (см. гл. 6) общая дисперсия (σ_0^2) равна сумме межгрупповой дисперсии ($\sigma_{мг}^2$) и средней величины из внутригрупповых дисперсий ($\overline{\sigma_{вг}^2}$):

$$\sigma_0^2 = \delta_{мг}^2 + \overline{\sigma_{вг}^2}.$$

Поэтому средняя ошибка типической выборки менее или, в крайнем случае (если межгрупповая дисперсия равна нулю), равна средней ошибке собственно-случайной выборки. Отсюда вытекает, что выборку из генеральной совокупности, состоящей из резко различающихся между собой групп, целесообразно производить путем типического отбора и нецелесообразно производить путем случайного отбора, при котором средняя ошибка выборки зависит от общей вариации признака.

Рассчитаем предельную ошибку выборки и найдем значение генеральной средней по данным типической выборки по материалам табл. 7.2. Выборочная средняя, как подсчитано выше, составляет 9,07 года. Среднюю ошибку типической выборки при бесповторном отборе определим по формуле $\delta_{\bar{x}} = \sqrt{\frac{\tilde{\sigma}_i^2}{n}\left(1-\frac{n}{N}\right)}$. Используя данные, приведенные в табл. 7.2, найдем среднюю взвешенную из дисперсий отдельных групп рабочих:

$$\overline{\sigma_0^2} = \frac{8{,}2\times15+19{,}9\times9+25{,}66\times6}{15+9+6} = \frac{456{,}06}{30} = 15{,}2.$$

Тогда средняя ошибка типической выборки будет:

$$\mu_{\bar{x}} = \sqrt{\frac{15{,}2}{30}\left(1-\frac{30}{300}\right)} = \sqrt{\frac{13{,}68}{30}} = \sqrt{0{,}456} = 0{,}67$$ (года).

Средняя ошибка при выборочном изучении величины стажа рабочих составила при собственно-случайном повторном отборе 1,35 года, при собственно-случайном бесповторном отборе – 1,27 и при механическом – 1,28. Следовательно, применительно к данному примеру типическая выборка при прочих равных условиях дает более точные результаты по сравнению с собственно-случайной выборкой и механическим отбором.

Уменьшение средней ошибки выборки влечет за собой уменьшение и предельной ошибки выборки. Так, с вероятностью 0,683 (t=1) можно утверждать, что средний стаж всех рабочих цеха находится в пределах 9,07±0,67, или от 8,4 до 9,74 года. С вероятностью 0,954 (t=2) можно утверждать, что средний стаж всех рабочих цеха находится в пределах 9,07±1,34, или от 7,73 до 10,41 года и т.д.

Аналогично обстоит дело и с предельной ошибкой доли. Выборочная доля рабочих со стажем свыше 10 лет, как подсчитано выше, составляет 33%. Среднюю ошибку доли определим по среднеарифметической взвешанной. Используя данные, приведенные в табл. 7.2, найдем среднюю взвешенную из дисперсий отдельных групп рабочих:

$$\overline{\sigma_0^2} = \frac{0\times15+0{,}25\times9+0\times6}{15+9+6} = \frac{2{,}25}{30} = 0{,}075.$$

Тогда средняя ошибка типической выборки будет:

$$\mu_{\bar{x}} = \sqrt{\frac{0{,}075}{30}\left(1-\frac{30}{300}\right)} = \sqrt{\frac{0{,}0675}{30}} = \sqrt{0{,}00225} = 0{,}047,$$

или 4,7%.

Средняя ошибка при выборочном изучении доли рабочих со стажем свыше 10 лет составила при собственно-случайном повторном отборе 8,8%, при собственно-случайном бесповторном отборе – 7,9% и при механическом – 8,8%. Следовательно, как видно из рассматриваемого примера, типическая выборка при прочих равных условиях дает более точные результаты и при изучении доли признака.

Типическая выборка обычно применяется при изучении сложных статистических совокупностей. Например, при выборочном обследовании производительности труда работников торговли, состоящих из отдельных групп по квалификации.

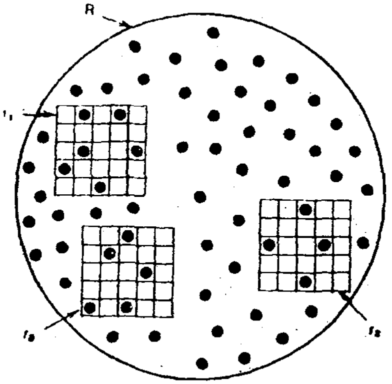
Важной особенностью типической выборки является то, что она дает более точные результаты по сравнению с другими способами отбора единиц в выборочную совокупность.

7.10. Серийный (либо кластерный или гнездовой) отбор

В рассмотренных видах выборки (собственно-случайный, механический и типический отбор) речь шла об отборе отдельных единиц из генеральной совокупности.

При серийной выборке из генеральной совокупности отбираются не отдельные единицы, а целые их серии (гнезда). Внутри же каждой из попавшей в выборку серии обследуются все без исключения единицы, т.е. применяется сплошное наблюдение (рис. 7.5).

Широкое применение гнездового отбора единиц наблюдения в статистической практике (особенно в статистике сельского хозяйства и статистике населения) обусловлено двумя основными причинами. Первая из них вызвана тем, что при проведении обследования отсутствует основа выборки (список элементов совокупности), а ее составление или невозможно, или требует слишком больших материальных затрат.



R - генеральная совокупность разбивается на серии.
В случайном порядке отбираются целые серии (г), в которых проводится сплошное обследование

Рис. 7.5. **Серийная выборка**

Например, такая ситуация имеет место, когда для обследований населения нет полных и не устаревших его списков. Однако по картам подлежащие обследованию районы могут быть разделены на территориальные участки с легко идентифицируемыми границами. Считая такие участки гнездами, можно решить задачу построения списка единиц отбора. Вторая причина состоит в том, что, даже если имеется списочная основа элементов, согласно экономическим соображениям может быть необходим выбор более крупных единиц отбора.

Серийный отбор значительно проще в организационном отношении и дешевле, чем другие способы. Однако получающаяся в процессе этого отбора случайная ошибка выборки в подавляющем большинстве случаев больше, чем при любом другом способе.

Серийную выборку применяют в двух случаях:

1) когда все серии имеют одинаковое количество единиц;

2) серии различны по объему.

Отбор серий (групп) производят указанными выше способами: собственно-случайным (жеребьевка) и по определенной схеме (механический отбор). В каждой серии (группе) единицы подвергаются сплошному учету.

Серии состоят из единиц, связанных между собой различным образом:

- территориально (районы, поселки и т.п.);
- организационно (предприятия, цеха, бригады);
- во времени (совокупность единиц продукции, выработанной за конкретный отрезок времени).

Серийная выборка обеспечивает экономию в расходах, если обследования распространяются на обширную территорию и гнездами являются территориальные единицы.

Применение серийной выборки в торговле обусловлено тем, что многие товары для их транспортировки, хранения и продажи упаковываются в пачки, коробки, ящики и т.п. Поэтому при контроле качества поступившего в упаковке товара рациональнее проверить несколько отдельных упаковок (серий), чем из всех упаковок отобрать необходимое количество единиц товара (рис 7.5).

Ошибки серийной выборки (с равными сериями) определяются по формулам, сходным с формулами ошибок собственно-случайной выборки. Допустим, что генеральная совокупность поделена на некоторое количество (S) равных по численности серий. Производится отбор определенного числа (s) серий в порядке случайной или механической выборки. При этом точность выборки будет зависеть оттого, насколько хорошо средние показатели серий ($\bar{x}_i$) будут репрезентировать генеральную среднюю. Чем меньше средние показатели серий будут отклоняться от генеральной средней, тем точнее будут результаты выборки.

При серийном отборе мерой колеблемости является межсерийная выборочная дисперсия, представляющая средний квадрат отклонений серийных средних от общей выборочной средней ($\bar{x}_s$):

$$\delta_0^2 = \frac{\Sigma(\tilde{x}_i - \tilde{x})^2}{S},$$

либо при изучении доли, произведение долей признака

$$\delta_0^2 = w_S(1 - w_S),$$

где s – число серий, отобранных для обследования.

Средняя ошибка серийной выборки (μ) определяется по таким формулам (табл. 7.6).

Покажем расчет предельной ошибки выборки и определение границ генеральной средней по данным серийной выборки на ранее приводившемся примере. Рабочие сборочного цеха поделены на 30 бригад по 10 человек в каждой. Для определения среднего производственного стажа рабочих цеха путем 10% выборки нужно на основе случайной жеребьевки выбрать 3 бригады. Пусть это оказались 10 (табельный № 91–100), 24 (табельный № 231–240) и 28 (табельный № 271–280). Тогда определение границ генеральной средней достигается путем следующих расчетов.

Таблица 7.6

Формула расчета средней выборки при серийном отбере

	Повторная выборка	Бесповторная выборка
При определении среднего размера изучаемого признака	$\mu_{\tilde{x}} = \sqrt{\frac{\delta^2}{S}}$ (16)	$\mu_{\tilde{x}} = \sqrt{\frac{\delta^2}{S}\left(\frac{S-s}{S-1}\right)}$ (17)
При определении доли данного признака	$\mu_p = \sqrt{\frac{w_S(1-w_S)}{S}}$ (18)	$\mu_p = \sqrt{\frac{w_S(1-w_S)}{S}\left(\frac{S-s}{S-1}\right)}$ (19)

Средний стаж в 10 бригаде равен $\tilde{x}_1$ = 4,9 года, в 24 $\tilde{x}_2$ = 11,1 года, в 28 – $\tilde{x}_3$ = 10,8 года и в целом по выборке $\tilde{x}$ = (4,9 + 11,1 + +10,8) : 3 = 8,9 (года).

Межсерийная дисперсия составляет:

$$\delta_0^2 = \frac{(4{,}9 - 8{,}9)^2 + (11{,}1 - 8{,}9)^2 + (10{,}8 - 8{,}9)^2}{3} = \frac{24{,}45}{3} = 8{,}15.$$

Для определения величины среднего стажа работы всех рабочих цеха нужно найти среднюю ошибку выборки. Так как отбор производился по схеме бесповторной выборки, то эта ошибка определяется по формуле (46):

$$\mu_{\tilde{x}} = \sqrt{\frac{\delta^2}{S}\left(\frac{S-s}{S-1}\right)} = \sqrt{\frac{8{,}15}{3}\left(\frac{30-3}{30-1}\right)} = 1{,}59 \quad \text{(года)}.$$

Отсюда с вероятностью 0,683 (t = 1) можно утверждать, что средний стаж всех рабочих цеха находится в пределах 8,9 ± 1,59, или от 7,31 до 10,49 лет.

С вероятностью 0,954 (t = 2) можно утверждать, что средний стаж рабочих цеха находится в пределах 8,9 ± 3,18, или от 5,72 до 12,08 года и т.д.

Применение серийного отбора целесообразно и тогда, когда серийные средние мало отличаются друг от друга и межсерийная дисперсия мала.

7.11. Комбинированный отбор

Рассмотренные способы выборки на практике обычно применяются не в «чистом» их виде, а комбинируются в различных сочетаниях и с различной последовательностью.

Это вызвано тем, что отбор единиц из генеральной совокупности для их обследования представляет порой сложный процесс, который затрагивает различные стороны образования выборки и в каждом конкретном случае может быть осуществлен по различным схемам.

В статистике торговли можно, например, комбинировать серийный отбор со случайной выборкой. При этом генеральная совокупность вначале разбивается на серии и отбирается нужное число серий. Далее в отобранных сериях производится случайный отбор единиц в выборочную совокупность.

Средняя ошибка комбинированной выборки определяется по формулам:

а) при повторном отборе

$$\mu = \sqrt{\frac{\overline{\sigma^2}}{n} + \frac{\delta^2}{r}},$$

б) при бесповторном отборе

$$\mu = \sqrt{\frac{\overline{\sigma^2}}{n}\left(1 - \frac{n}{N}\right) + \frac{\delta^2}{r}\left(\frac{R-r}{R-1}\right)},$$

при этом n – число единиц, взятое в выборку из серий.

В статистике различают также одноступенчатый и многоступенчатый способы отбора единиц в выборочную совокупность.

При одноступенчатой выборке каждая отобранная единица сразу же подвергается изучению по заданному признаку. Так обстоит дело при собственно-случайной и серийной выборке.

При многоступенчатой выборке производят отбор из генеральной совокупности отдельных групп, а из групп выбираются отдельные единицы. Так производится типическая выборка с механическим способом отбора единиц в выборочную совокупность.

Комбинированная выборка может быть двухступенчатой. При этом генеральная совокупность сначала разбивается на группы. Затем производят отбор групп, а внутри последний осуществляется отбор отдельных единиц.

Выборка может быть, многоступенчатой, если сначала производит отбор крупных групп. Затем из крупных групп отбираются средние, потом мелкие и внутри последних отбираются отдельные единицы. Например, трехступенчатый отбор осуществляется при бюджетных обследованиях семей дехканских хозяйств. Вначале выбирают районы, затем внутри каждого района отбираются хозяйства и внутри последних выбираются личные хозяйства крестьян. При этом на отдельных ступенях могут изменяться и виды выборки. Так, при отборе районов обычно применяется случайная выборка, при отборе крестьян – механическая выборка, а при отборе хозяйств – снова применяется собственно-случайный отбор.

В отличие от типической выборки, где формирование выборочной совокупности производится из всех групп, при многоступенчатой выборке производится отбор самих групп. Поэтому не все они попадают в выборочную совокупность.

Средняя ошибка выборки при многоступенчатом отборе определяется по формуле

$$\mu = \sqrt{\mu_1^2 + \frac{\mu_2^2}{n_1} + \frac{\mu_3^2}{n_1 n_2} + \ldots + \frac{\mu_n^2}{n_1 n_2, \ldots, n_n}}, \quad (22)$$

где μ_1, μ_2, ..., μ_n – средние ошибки выборки на отдельных ступенях отбора; n_1, n_2, ..., n_n – численность выборки на соответствующих ступенях отбора.

На практике известны случаи, когда выборочное обследование организуется так, что одни сведения получают от всех единиц, а другие – только по некоторым из них. Такая выборка называется многофазной.

Отличие многофазной выборки от многоступенчатого отбора заключается в том, что при многофазной выборке на каждой фазе сохраняется одна и та же единица отбора. В многоступенчатых выборках единица отбора на каждой ступени выборки различная.

Важной особенностью многофазной выборки является возможность использовать данные первой фазы наблюдения для дополнительной характеристики и уточнения результатов, полученных на второй фазе, а эти данные, в свою очередь, – для третьей фазы и т.д.

Например, на первой фазе по краткой программе (пять вопросов) обследуется 25% генеральной совокупности, на второй фазе по более широкой программе (включающей еще десять вопросов) – 15% генеральной совокупности, на третьей фазе по расширенной программе (включающей еще десять вопросов) – 5% и т.д.

Особым видом выборочного наблюдения является *моментное наблюдение,* т.е. выборочное во времени. Все элементы изучаемой совокупности единиц подлежат сплошному учету.

Объектами выборки являются отрезки времени. В связи с этим понятия генеральной и выборочной совокупностей относятся не к совокупности единиц, а ко времени наблюдения.

7.12. Ошибки выборки

Чтобы критически оценить качество материалов, полученных выборочным наблюдением, необходимо выяснить, какие вообще могут быть расхождения между выборочными и генеральными характеристиками.

При выборочном наблюдении могут возникать двоякого рода погрешности или, как их называют иначе, ошибки: *ошибки регистрации* и *ошибки репрезентативности* (рис. 7.6).

Ошибки выборки

Ошибки регистрации	Ошибки репрезентативности
• случайные ошибки • систематические ошибки	• случайные ошибки • систематические ошибки

Рис. 7.6. **Погрешности выборочного наблюдения**

Ошибки регистрации

Ошибками регистрации называются такие погрешности, которые возникают вследствие неправильной регистрации данных об отдельных обследуемых единицах. Этого рода ошибки рассмотрены в III главе и поэтому нет необходимости разъяснять их сущность.

Ошибки репрезентативности

Ошибками репрезентативности называются расхождения между обобщающими (сводными) показателями отобранной части совокупности и всей совокупности в целом в условиях правильно произведенной регистрации данных. Ошибки репрезентативности так же как и ошибки регистрации, могут быть систематическими и случайными.

Систематические ошибки репрезентативности

Систематические ошибки репрезентативности возникают в силу нарушения принципов проведения выборочного наблюдения (неправильное производство отбора, произвольная замена попавших в выборку единиц другими единицами совокупности и т.д.).

Если при производстве отбора нарушаются требования теории выборки, то это неизбежно приводит к ошибкам. Основное требование научно-правильного производства отбора заключается в том, чтобы отбор каждой в отдельности единицы производился объективно, без всякого предвзятого мнения. Ошибки отбора приводят к неслучайным ошибкам. Так бывает, если объективный отбор подменяется «удобной» выборкой. Например, когда появляются добровольные респонденты – те, кто сами предлагают, чтобы их опросили. Очевидно, что характеристики таких добровольцев и не добровольцев могут быть отличны и это приведет к ошибочному заключению о генеральной совокупности.

Такая же опасность возникает при замене по какой-либо причине единиц, попавших в выборку, другими единицами (например, вместо отобранного домохозяйства, где в момент прихода интервьюера никто не открыл дверь, был проведен опрос в соседней квартире; или интервьюер встретил решительный отказ участвовать в опросе и был вынужден пойти на замену домохозяйства).

Неслучайные ошибки могут возникнуть из-за методов сбора данных: вопросов, слишком болезненных для опрашиваемых (об отношении к властям, если опрашиваются беженцы или пострадавшие от стихийных бедствий и т.д.) или формы задания вопроса (очень трудно, чтобы всем было все понятно), или времени опроса (например, на вопрос молодым родителям, не жалеют ли они о том, что у них есть дети, можно получить разное, распределение ответов в зависимости от того, проводился ли опрос долгим зимним вечером, когда все утомлены приготовлением уроков, простудами и т.д., или прекрасным летним днем, когда дети находятся на даче, в оздоровительном лагере).

Плохо, когда ошибка выборки превышает допустимый размер погрешности, но слишком высокая точность также подозрительна и, как правило, свидетельствует об ошибках отбора.

При правильном проведении выборки систематических ошибок репрезентативности не будет.

Однако все же и в этом случае возможны расхождения между сводными характеристиками выборочной и генеральной совокупности. Расхождения эти возникают в силу того, что выборочная совокупность не воспроизводит точно генеральную совокупность.

Расхождения между характеристиками отобранной части совокупности и характеристиками всей совокупности в целом, в условиях правильно произведенного отбора и правильной регистрации, носят название случайных ошибок репрезентативности.

Случайные ошибки – те, которые изменяются по вероятностным законам. К случайным относится ошибка выборки.

Ошибки выборки

Ошибка выборки или, иначе говоря, *ошибка репрезентативности* – это разница между значением показателя, полученного по выборке, и генеральным параметром. Так, ошибка репрезентативности:

• выборочной средней	$\mu_{\tilde{x}} = \tilde{x} - \bar{x}$
• выборочной доли	$\mu_w = w - p$
• дисперсии	$\mu_{\sigma^2} = \tilde{\sigma}^2 - \bar{\sigma}^2$
• коэффициент корреляции	$\mu_r = \tilde{r} - \bar{r}$

Различают также среднюю (стандартную) и предельную ошибки выборки (рис. 7.7).

Рис. 7.7. **Ошибки выборки**

Средняя ошибка

Средняя ошибка характеризует меру отклонений выборочных показателей от аналогичных показателей генеральной совокупности (см табл. 7.5; 7.6).

Предельная ошибка

Предельной ошибкой принято считать максимально возможное расхождение выборочной и генеральной характеристик, т.е. максимум ошибки при заданной вероятности ее появления.

Величина случайной ошибки репрезентативности зависит:

- от объема выборки;
- от степени вариации изучаемого признака в генеральной совокупности;
- от принятого способа формирования выборочной совокупности.

7.13. Определение необходимого объема выборочной совокупности

Основным принципом применения выборочного метода является обеспечение равной возможности для всех единиц генеральной совокупности быть отобранными в выборочную совокупность. При таком подходе соблюдается требование случайного, объективного отбора и, следовательно, ошибка выборки определяется, прежде всего, ее объемом *(n)*. С увеличением последнего величина средней ошибки уменьшается, характеристики выборочной совокупности приближаются к характеристикам генеральной совокупности. То есть, с увеличением объема выборочной совокупности повышается точность выборочных данных. Поэтому вопрос о необходимом объеме выборки является одним из принципиальных вопросов организации выборочного наблюдения. Поскольку статистическое исследование является довольно трудоемким процессом, то перед статистиками встает вопрос об определении минимальной численности выборки.

Формулы для расчета объема выборки (n) можно получить путем преобразования соответствующих формул предельной ошибки выборки.

$t \cdot \mu \leq \Delta_0$ или $t \cdot \sqrt{\frac{\sigma^2}{n}} = \Delta_0$ отсюда $n = \left(t \frac{\sigma}{\Delta}\right)^2$. (23)

В зависимости от способа и схемы отбора объем «n» рассчитывается по следующим формулам (табл. 7.7).

Таблица 7.7

Формулы расчета необходимой численности выборки

Способ отбора	Необходимый объем выборки	
	для расчета средней	для расчета доли
Повторная выборка	$n = \frac{t^2\sigma^2}{\Delta_{\bar{x}}^2}$ (24)	$n = \frac{t_w^2(1-w)}{\Delta_w^2}$ ()25
Бесповторная выборка	$n = \frac{t^2\sigma^2 N}{\Delta_{\bar{x}}^2 N + t^2\sigma^2}$ (26)	$n = \frac{t_w^2(1-w)N}{\Delta_w^2 N + t_w^2(1-w)}$ (27)

Следовательно, для определения необходимого объема выборочной совокупности нужно иметь следующие исходные данные:

- σ^2 или (1-w) – дисперсия и доля;
- Δ – допустимая ошибка;
- t и p – коэффициент доверия и уровень вероятности – F(t).

При определении необходимого объема выборки по приведенным формулам возникает трудность в нахождении значений σ^2 и то, так как эти величины можно получить только после проведения выборочного обследования. В связи с этим вместо фактических значений данных показателей подставляют приближенные, которые могли быть определены на основе каких-либо пробных выборочных наблюдений или из аналитических предыдущих обследований.

В тех случаях, когда статистик знает среднее значение изучаемых признаков (например, из инструкций, законодательных актов и т.п.) или пределы, в которых этот признак варьируется, можно применить следующий расчет по приближенным формулам:

$$\sigma \cong \frac{1}{3}\bar{x}, \quad (28)$$

а произведение w(1 – w) заменить значением 0,25 (w = 0,5).

Чтобы получить более точный результат, принимают максимально возможное значение этих показателей. Если распределение признака в генеральной совокупности подчиняется нормальному закону, то размах вариации примерно равен 6σ (крайние значения отстоят в ту и другую сторону от средней на расстоянии 3σ). Отсюда $\sigma = \frac{1}{6}R$, но если распределение заведомо асимметрично, то $\sigma = \frac{1}{5}R$.

При любом виде выборки ее объем начинают рассчитывать по формуле повторного отбора

$$n = \frac{t^2\sigma^2}{\Delta_{\bar{x}}^2}.$$

Если в результате расчета доля отбора (n) превысит 5%, то проводят расчет по формуле бесповторного отбора.

Объем выборки определяется на подготовительном этапе обследования.

Несколько слов о **степени вариации изучаемого признака** в генеральной совокупности. Надо сказать, что при одинаковой численности выборочных совокупностей и прочих равных условиях ошибка выборки будет меньше в той из них, которая отобрана из генеральной совокупности **с меньшей вариацией изучаемого признака**. Уменьшение вариации признака означает снижение величины дисперсии (σ^2 – для количественного признака или [p(1– p)] – для альтернативного признака). Если например, для изучения общественного мнения жителей города в архитектурном управлении получить сведения о жилом фонде и из всех имеющихся в городе квартир отобрать случайным образом 400 квартир, а затем предложить интервьюерам опросить всех, кого они застанут в момент посещения в этих квартирах, то полученные данные не будут репрезентативны. Допущена систематическая ошибка: более подвижная часть населения попадает в выборку в меньшей пропорции, а менее подвижная – в большей пропорции, чем в генеральной совокупности. Пенсионеров, например, можно чаще застать дома, чем студентов-вечерников. При увеличении выборки эта ошибка не устраняется: если мы проведем опрос в 800 квартирах или даже во всех квартирах города (сплошной опрос), то полученные данные будут репрезентативны для населения, находящегося дома в момент прихода интервьюера, а не для всех жителей города.

Наконец, о зависимости величины ошибки выборки от способов формирования выборочной совокупности подробно рассмотрено в параграфах 7.7, 7.8, 7.9 и 7.10.

7.14. Оценка результатов выборочного наблюдения

Распространению результатов выборочного наблюдения на генеральную совокупность, которое является заключительным этапом выборочного наблюдения, предшествует оценка этих результатов.

Оценка результатов выборочного наблюдения включает:

➢анализ качества отобранных данных и, в первую очередь, полноты информации (все ли группы, все ли типы единиц генеральной совокупности представлены в выборочной совокупности);

➢расчет относительной ошибки.

Относительная ошибка, %:

• для средней $\Delta_{\bar{x}}^{\%} = \frac{\Delta_{\bar{x}}}{\bar{x}} \cdot 100;$ (29)

• для доли $\Delta_{w}^{\%} = \frac{\Delta_{w}}{p} \cdot 100$ (30)

Если величина относительной ошибки не превышает заранее установленного для данного обследования предельного значения (в %), то данные выборочного наблюдения являются представительными и могут быть распространены на генеральную совокупность.

Как было отмечено выше объем выборки (n) рассчитывается на стадии проектирования выборочного обследования. Так как

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

то

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

где Δ – допустимая погрешность, которая задается исследователем исходя из требуемой точности результатов проектируемой выборки;

t – табличная величина, соответствующая заданной доверительной вероятности *F(t)*, с которой будут гарантированы оценки генеральной совокупности по данным выборочного обследования;

σ^2 – генеральная дисперсия.

Последняя величина, как правило, неизвестна. Используются какие-либо ее оценки: результаты прошлых обследований той же совокупности, если ее структура и условия развития достаточно стабильны, или же зная примерную величину средней, находят дисперсию из соотношения

$$\sigma = \frac{1}{3}\bar{x};$$

если известны x_{max} и x_{min}, то можно определить среднее квадратическое отклонение в соответствии с правилом «трех сигм».

$$\sigma = \frac{1}{6}(x_{max} - x_{min}),$$

так как в нормальном распределении в размахе вариации «укладывается» $6\sigma(\pm 3\sigma)$. Если распределение заведомо асимметричное, то

$$\sigma \approx \frac{1}{5}(x_{max} - x_{min}).$$

Для относительной величины принимают максимальную величину дисперсии $\sigma^2_{max} = 0{,}5 \cdot 0{,}5 = 0{,}25$.

При расчете и не следует гнаться за большими значениями t и малыми значениями Δ, так как это приведет к увеличению объема выборки, а следовательно, к увеличению затрат средств, труда и времени, вовсе не являющемуся необходимым.

После проведения выборки рассчитывают *возможные ошибки выборочных показателей* (ошибки репрезентативности), которые используются для оценки результатов выборки и для получения характеристик генеральной совокупности.

Пример. На электроламповом заводе взято для проверки 100 ламп. Средняя продолжительность их горения оказалась 1420 ч со средним квадратическим отклонением 61,03 ч. Поскольку приемщика продукции интересует качество всей партии (50 тыс. электроламп), оценивают точность полученной средней.

Средняя возможная ошибка вычисленной выборочной средней

$$S_{\bar{x}} = \frac{61{,}03}{\sqrt{100}} = \pm 6{,}1.$$

С вероятностью 0,954 предел возможной ошибки ДЛ = 2 • 6,1 = ± 12,2 ч.

$$\Delta_x = 2 \cdot 6{,}1 = \pm 12{,}2$$

С вероятностью 0,954 можно утверждать, что средняя продолжительность горения 1 электролампы во всей партии будет находиться в пределах от 1408 до 1432 ч; 46 электроламп из 1000 могут иметь срок горения, выходящий за эти пределы. Приемщика продукции интересуют отклонения от вычисленных пределов только в сторону сокращения продолжительности горения. Меньше чем 1408 ч могут гореть 23 лампы из 1000. На основании этого приемщик продукции решает вопрос о годности всей партии электроламп.

Решение вопроса может быть уточнено: определим, у какой доли ламп срок службы окажется меньше установленного лимита. Для потребителя продукции таким лимитом являются 1410 ч, продукция с меньшим сроком горения неприемлема.

При контрольной проверке 1000 ламп 100 ламп горели менее 1410ч, их удельный вес p = 0,1, или 10%. Средняя возможная ошибка этой доли

$$Sp = \sqrt{\frac{0{,}1 \cdot 0{,}9}{100}} = \pm 0{,}03,$$ или ±3%.

С вероятностью 0,954 предел ошибки доли Δp = 2·0,03 = ± 0,06, или ±6%. Следовательно, во всей партии можно ожидать от 4 до 16% некачественных электроламп.

Чаще всего делают заключение об удовлетворительности выборки, сопоставляя получившиеся пределы ошибок выборочных показателей с величинами допустимых погрешностей. Может получиться, что предел ошибки, рассчитанный с заданной вероятностью, окажется выше допустимого размера погрешности. В этих случаях определяют *вероятность того, что ошибка выборки не превзойдет допускаемую погрешность*. Решение этой задачи и заключается в отыскании $F(t)$ на основе формулы предела ошибки выборки:

$$\Delta = t\sqrt{\frac{s^2}{n}},$$

где Δ – допустимый размер погрешности оцениваемого показателя;

s^2 – дисперсия показателя, рассчитанная по данным выборочного наблюдения;

n – объем проведенной выборки.

Продолжим пример с оценкой качества электроламп. Если при приемке партии электроламп ставится условие, что минимальный срок горения электроламп 1410 ч, то, учитывая среднюю продолжительность горения по выборке ($\bar{x}$ = 1420 ч), допустимая погрешность равна 10 ч: 1410 – 1420 = – 10 ч.

Как было установлено выше, с вероятностью 0,954 предел возможной ошибки выборочной средней составил 12,2 ч, что превосходит допустимую погрешность. Является ли это основанием для браковки всей партии? Для ответа на этот вопрос определяют вероятность риска при приемке продукции:

$$\Delta = t\sqrt{\frac{61{,}03^2}{100}},$$ отсюда t = 1,64.

Соответствующая доверительная вероятность 0,899. Вероятность того, что средний срок горения лампы меньше 1410 ч, равна:

$$1 - \frac{1+F(t)}{2} = 1 - \left(0{,}5 + \frac{0{,}899}{2}\right) = 0{,}05.$$

Следовательно, из 100 ламп 5 могут гореть менее 1410 ч – риск появления некачественной продукции достаточно высок.

Аналогично можно определить вероятность того, что предел ошибки доли не превысит допускаемую погрешность доли.

Оценки надежности выборочных показателей, как показано на примере, позволяют принять обоснованные решения в отношении генеральной совокупности.

В качестве показателя репрезентативности выборки используется отношение соответствующего показателя выборочной совокупности к показателю генеральной совокупности, выраженное в процентах. При сопоставлении средних этот показатель будет иметь следующий вид:

$$i_p = \frac{\tilde{x}}{\bar{x}} \cdot 100, \quad (33)$$

где i_p – показатель репрезентативности.

Если отношение i_p близко к 100%, то репрезентативность отбора считается удовлетворительной. На практике часто принимают границы удовлетворительной репрезентативности в ±5%, т.е. от 95 до 105%. Это значит, что если отношение $\tilde{x}$ к $\bar{x}$ не меньше 95% и не больше 105%, совокупность отобранных единиц считается репрезентативной.

Если окажется, что отобранная совокупность нерепрезентативна, то отбор производится заново. Если же и вторичный отбор не дал удовлетворительной репрезентативности, то увеличивают численность выборки.

7.15. Распространение данных выборочного наблюдения на генеральную совокупность

Конечной целью выборочного наблюдения является характеристика генеральной совокупности на основе данных, полученных по выборке. Для этого результаты всякого выборочного обследования должны быть распространены на всю генеральную совокупность.

Расчет объемных показателей генеральной совокупности на основании данных выборочного наблюдения называется в статистике распространением выборочных характеристик на всю совокупность. Существуют два способа такого распространения: 1) *способ прямого пересчета* и 2) *способ коэффициентов.*

В первом случае средний размер признака, найденный в результате выборочного обследования ($\tilde{x}$,w):

$$\tilde{x} - \Delta_x \le \bar{x} \le \tilde{x} + \Delta_x \quad (34)$$

$$w - \Delta_w \le P \le w + \Delta_w \quad (35)$$

умножается на число единиц генеральной совокупности (N)

$$\tilde{x}N - \Delta_x N \le \bar{x}N \le \bar{x}N + \Delta_{\tilde{x}} N; \quad (36)$$

$$wN-\Delta_w N \le PN \le wN+\Delta_w N. \quad (37)$$

Способ коэффициентов целесообразно применять тогда, когда выборочное наблюдение проводится с целью проверки и уточнения данных сплошного наблюдения. Например, численности учтенных единиц наблюдения. В этом случае используется формула:

$$N_1 - N_0 \frac{n_1}{n_0},$$

где N_0 – численность совокупности без поправки на недоучет;

N_1 – численность совокупности с поправкой на недоучет;

n_0 – численность совокупности в контрольных точках по первоначальным данным;

n_1 – численность совокупности в контрольных точках по данным проверки.

Задание 8

Для анализа объемов продаж однотипных предприятий и фирм, реализующих однокачественный условный товар в одном из регионов РУз в I кв. исследуемого года, была проведена 20% механическая выборка, в результате которой было *А* отобрано 30 фирм.

Требуется, используя способ прямого пересчета, определить общий объем продаж предприятий генеральной совокупности если известно, что средний объем продаж фирм выборочной совокупности составил 10,31 млн сум.

Решение:

По условию: $n = 0{,}2N$, отсюда

$$N = \frac{n}{0{,}2} = \frac{30}{0{,}2} = 150 \quad \text{(ед.)}.$$

Тогда общий объем продаж всех однотипных предприятий и фирм, реализующих однокачественный условный товар в данном регионе в I кв. исследуемого года, составит:

10,31 · 150 = 1546,5 (млн сум).

Малая выборка

Распространение выборочных характеристик (средняя, доля) на генеральную совокупность, основанное на действии закона больших чисел, предполагает достаточно большой объем выборки. Из теоремы П.Л. Чебышева вытекает, что, чем больше объем выборки, тем выше репрезентативность выборочных характеристик. Наоборот, уменьшение объема выборки ухудшает репрезентативность, приводит к значительным ошибкам при определении генеральных характеристик.

В то же время не всегда возможен и целесообразен большой объем выборки. Изучение качества продукции, производительности труда, урожайности и других показателей в ряде случаев не позволяет производить наблюдение в больших масштабах. Возникает задача отыскания путей получения генеральных характеристик при *малой выборке, объемом до 30 единиц.* В математической статистике доказывается, что и при малых выборках возможно распространение выборочных характеристик на генеральную совокупность.

Начало разработке теории малой выборки положил английский математик-статистик Вильгельм Госсет, публиковавший свои работы под псевдонимом Стьюдент. Затем эти работы продолжил английский математик-статистик Р.А. Фишер. Большой вклад в теорию малой выборки сделан советскими математиками В.И. Романовским и А.Н. Колмогоровым.

Интеллектуальный тренинг

1. Что такое выборочное наблюдение, каковы его особенности?

2. Каковы причины применения выборочного наблюдения?

3. Какие задачи решаются при применении выборочного наблюдения?

4. Что означает ошибка репрезентативности, какие факторы определяют ее величину?

5. Чем отличается распределение ошибок простой случайной выборки при проведении больших и малых выборок?

6. От чего зависит точность оценки параметров генеральной совокупности (генеральный средней и генеральной доли)?

7. Запишите доверительные интервалы генеральной средней с вероятностью 0,95 и 0,99.

8. Чем отличается величина средней квадратической ошибки простой случайной выборки при повторном и бесповториом отборе? Какая из этих ошибок больше?

9. Как определяется предельная ошибка при проведении большой и малой выборок?

10. Какой вид выборочного наблюдения следует использовать, если генеральная совокупность не является однородной?

11. В чем состоят преимущества серийной выборки перед простой случайной выборкой?

12. В чем преимущества механической выборки и как определяется величина ее стандартной ошибки?

13. Как нужно изменить объем механической выборки, если среднюю квадратическую ошибку следует уменьшить в 1,5 раза?

14. Назовите важнейшие области применения выборочного метода в практике государственной статистики.

15. Как оцениваются результаты выборочного наблюдения?

16. Какие методы применяются для распространения данных выборочного наблюдения в генеральную совокупность?

Использованная и рекомендуемая специальаня литература

1. Бокун Н.И., Чернышова Н.М. Методы выборочных обследований. Минск, 1997.

2. Кокрен У. Методы выборочного исследования. М.: Статистика, 1997.

Глава VIII
КОРРЕЛЯЦИОННО-РЕГРЕССИОННЫЙ АНАЛИЗ

Дорожная карта

8.1. Понятие о корреляционной связи

Общая постановка задачи

В научном исследовании общественных явлений первостепенное значение имеет выявление и количественное выражение связей между ними.

Связи между явлениями, отдельными их признаками весьма многообразны. Но в любой конкретной связи одни признаки выступают в качестве факторов, воздействующих на другие и обусловливающих их изменение, другие же – в качестве результатов действия этих факторов. Иными словами, одни представляют собой причину, другие – следствие. **Первые** принято в статистике называть признаками-факторами или факторными признаками, **вторые** – результативными признаками.

Многообразие связей, в которых находятся явления, порождает необходимость в их классификации, в сведении связей к определенным типам, формам по их существенным чертам, свойствам. В основе наиболее общей классификации связей в статистике рассматривают различия и сходства связей по таким их особенностям, как степень тесноты, направление, аналитическое выражение, единичность или множественность факторов и следствий (результатов). В соответствии с этим различают следующие виды связи (рис. 8.1).

Виды связей

По типам связи	**По направлению связи**	**По формам зависимости**	**По числу факторных признаков**
• функциональные • корреляционные	• прямые • обратные	• прямолинейные • криволинейные	• однофакторные •многофакторные

Рис. 8.1. **Классификация связей**

Функциональные и корреляционные связи

Функциональной называют такую связь, при которой значения результативного признака (**величина его вариантов**) целиком и полностью определяются значениями признаков-факторов (**аргументов**), **так что определенной системе значений факторных признаков всегда соответствует строго определенное одно или несколько (в случае многозначной функции) значений результативного признака.**

Функциональная связь является жестко детерминированной, проявляется в каждом отдельном случае. Функциональная связь – полная связь, когда результативный признак (следствие) всецело зависит от факторного признака (причины) и не зависит от окружающих условий.

То есть, здесь нет места отклоняющему действию случайных причин. Если варианты факторного признака обозначить через **x**, а результативного через у то функциональную связь между ними можно выразить, например, в виде следующего уравнения $\mathbf{y = x^2}$. Это может быть, например, связь между площадью квадрата и его стороной.

Существуют функциональные связи и иного рода, в которых результативный признак является функцией нескольких факторных признаков, каждый из которых в той или иной мере влияет на формирование результативного признака, причём степень влияния известна. Примером может служить связь между площадью треугольника, его основанием и высотой. Значит, установив на основе единичного исследования эту зависимость, мы можем использовать его в любых аналогичных случаях.

При **функциональной связи** изменение результативного признакака **у** всецело обусловлено действием факторного признака **x**:

$$y = f(x).$$

Примером функциональной связи является зависимость длины окружности от радиуса (**R**):

$$l = 2\pi r.$$

*Т*акая форма связи дает возможность перечислить взаимодействующие факторы и установить определенную их пропорциональность. Например, если грузоподъемность автомобиля, используемого на перевозке зерна к элеватору, равна 2,5 т, то общее количество перевезенного зерна находится в функциональной зависимости от числа рейсов.

Но в экономической практике функциональные связи практически не встречаются. Такой «чистый» вид связи между признаками обычно используется в так называемых «точных» науках (физике, математике).

Статистическая (стохастическая) связь – это связь, при которой изменение одной переменной приводит к изменению распределения другой переменной. В частности, статистическая за***висимость*** проявляется в том, что при изменении одной из величин изменяется среднее значение другой. В этом случае речь идет о корреляционной зависимости.

Корреляционная связь

При **корреляционной связи** изменение результативного признака **у** обусловлено влиянием факторного признака **х** не всецело, а лишь частично, так как возможно влияние прочих факторов ε

$$y = \psi(x) + e.$$

Корреляционная связь[1] – это связь, при которой определенному значению факторного признака соответствует **среднее** значение результативного признака. Она выражается в том, что с изменением значения факторного признака меняется средняя величина результативного. Здесь между значениями признаков нет строгого соответствия в каждом отдельном случае.

Корреляционные связи проявляются, как правило, при большом числе наблюдений. Степень, полнота проявления этих связей и число случаев наблюдения, необходимое для их проявления, зависят от многих причин и, прежде всего, от степени тесноты этой связи. Чем теснее связь, чем ближе она к функциональной, тем яснее, точнее она проявляется и тем меньше число наблюдений, необходимое для ее выявления.

В качестве примера корреляционной связи можно назвать связь между себестоимостью продукции и производительно-

---

[1] Основоположниками теории корреляции считаются английские биометрики Ф. Гальтон (1822-1911) и К.Пирсон (1857-1936). Само слово **корреляция** ввел в употребление в статистику английский биолог и статистик Френсис Гальтон в конце XIX в. Тогда оно писалось как «corelation» (соответствие), но не просто «связь» (relation), а «как бы связь», т.е. связь, но не в привычной в то время функциональной форме. В науке вообще, а именно, в палеонтологии термин «корреляция» применил еще раньше, в конце XVIII в., знаменитый французский палеонтолог (специалист по ископаемым останкам животных и растений прошлых эпох) Жорж Кювье. Он ввел даже «закон корреляции» частей и органов животных. «Закон корреляции» помогает восстановить по найденным в раскопках черепу, костям и т.д. облик всего животного и его место в системе: если череп с рогами, то это было травоядное животное, а его конечности имели копыта; если же лапа с когтями – то хищное животное без рогов, но с крупными клыками.

стью труда рабочих, ***между*** размерами предприятия по объему их основных производственных фондов и объемом вырабатываемой продукции, между квалификацией рабочего и его заработком. Здесь, при одном и том же значении учтенного факторного признака возможны различные значения результативного признака. Это обусловлено наличием других факторов, которые могут быть различными по составу, направлению и силе действия на отдельные (индивидуальные) единицы статистической совокупности. Поэтому для изучаемой статистической совокупности в целом здесь устанавливается такое соотношение, в котором определенному изменению факторного признака соответствует среднее изменение признака результативного. Следовательно, характерной особенностью корреляционных связей является то, что они проявляются не в единичных случаях, а в массе. Поэтому изучаются корреляционные связи по так называемым эмпирическим данным, полученным в статистическом наблюдении. В таких данных отображается совокупное действие всех причин и условий на изучаемый показатель.

При статистическом изучении корреляционной связи определяется влияние учтенных факторных признаков при отвлечении (абстрагировании) от прочих аргументов. Применяемый таким образом способ научной абстракции хотя и ведет к некоторому упрощению (аппроксимации) реального механизма связи, но делает возможным установление закономерностей взаимодействия изучаемых показателей, что позволяет, не прибегая к экспериментированию, получать количественные характеристики корреляционной связи.

При изучении корреляционной связи перед статистикой ставятся следующие основные задачи:

✓ проверка положений экономической теории о возможности связи между изучаемыми показателями и придание выявленной связи аналитической формы зависимости;

✓ установление количественных оценок тесноте связи, характеризующих силу влияния факторных признаков на результативные.

Связь прямая и обратная

По направлению, как сказано выше, обычно различают два вида связи: прямую и ***обратную***. *Прямой называют такую связь,* **при которой с увеличением или уменьшением значений факторного признака, соответственно, увеличиваются или уменьшаются и средние значения результативного признака.** В приведенных примерах корреляционной связи две из них относятся к прямой. Это – связь между размерами предприятия по объему основных производственных фондов и объемом их продукции и связь между квалификацией рабочего и его заработками.

Чем крупнее предприятие по величине основных производственных фондов, тем больше, при прочих равных условиях, оно производит продукции, и чем выше квалификация рабочего (чем больше его тарифный разряд), тем выше, опять-таки при прочих равных условиях, его заработок.

Обратной называют такую связь, при которой значения результативного признака изменяются в противоположном направлении по отношению к изменениям значений факторного признака. При увеличении значений факторного признака значения результативного признака уменьшаются и наоборот.

Следовательно, по направлению связи **различают:**

- прямую регрессию, когда с увеличением факторного признака результативный признак увеличивается (и наоборот);
- обратную регрессию, когда с увеличением факторного признака результативный признак уменьшается (и наоборот).

Примером такой формы связи может служить упомянутая выше, связь между производительностью труда рабочих и себестоимостью единицы продукции. Чем выше производительность труда рабочих, тем ниже, при прочих равных условиях, себестоимость единицы продукции.

Необходимо иметь в виду, что прямая и обратная связи могут переходить одна в другую. Так, прямая связь по достижении факторным признаком определенной величины в некоторых случаях исчерпывает себя и затем становится обратной, т.е. с дальнейшим увеличением значений факторного признака, значения признака результативного уменьшаются. Такой характер имеет, например, связь между возрастом рабочего и его производительностью, между количеством удобрений, вносимых на единицу площади пашни, и величиной урожайности, получаемой с нее. Рассматривая производительность труда рабочего в отдельные периоды работы в течение рабочего дня, не трудно заметить, что в начале она не так высока, но постепенно повышается, однако к концу рабочего дня начинает сказываться усталость и производительность труда постепенно понижается.

Связи прямолинейные и криволинейные

По аналитическому выражению в общем виде обычно выделяют связи прямолинейные и криволинейные. Если данная связь может быть выражена точно или приближенно **уравнением какой-либо прямой линии,** ее называют **прямолинейной,** а если **уравнением какой-либо кривой,** то она называется **криволинейной связью.**

Аналитическими уравнениями точно могут быть выражены только функциональные связи, связи же корреляционные могут быть выражены ими лишь приближенно. Приближенное выражение при помощи аналитических уравнений корреляционных связей общественных явлений во многих случаях дает возможность получить результаты, вполне пригодные для научных и практических целей. Этим и объясняется широкое применение в статистике корреляционного метода исследования связей, основанного на аналитическом их выражении.

Таким образом по форме зависимости различают:

• линейную регрессию, которая выражается уравнением прямой:

$$\hat{y} = a + bx \ ;$$

• нелинейную регрессию, которая выражается уравнениями вида:

$$\hat{y} = a + bx + cx^2 \text{ – для параболы;}$$

$$\hat{y} = a + \frac{b}{x} \text{ – для гиперболы и др.}$$

Связи однофакторные и многофакторные

Изменение среднего значения результативного признака в зависимости только от одного фактора называют парной корреляцией, а в зависимости от нескольких факторов – **множественной (однофакторной или многофакторной)**. Несмотря на то, что в массовых общественных явлениях действует множество факторов, можно исследовать зависимость результативного признака только от одного или нескольких факторов, абстрагируясь или, как говорят в статистике, элиминируя действие на него всех других факторов, называемых при этом обычно прочими факторами.

Таким образом, по числу факторных признаков **различают:**

❖ парную корреляцию, когда один факторный признак;

❖ множественную корреляцию, когда много факторных признаков.

Общий вид уравнений парной регрессии:

$$\hat{y} = f(x).$$

Общий вид многофакторных уравнений регрессии:

$$\hat{y} = f(x_1 x_2, \ldots, x_n).$$

8.2. Методы выявления корреляционной связи

На первом этапе статистического изучения взаимосвязей социально-экономических явлений определяют наличие связи и ее направление (если связь существует) при помощи следующих элементарных методов.

Важнейшие приемы изучения взаимосвязи общественных явлений

В установлении взаимосвязи между признаками общественных явлений определяющая роль принадлежит теоретическому анализу. Так, теоретически бесспорно, например, что уровень себестоимости продукции зависит от производительности труда, урожайности – от качества обработки почвы, количества внесенных удобрений и т.д.

Теоретически установленные зависимости одних признаков от других могут иметь различное количественное выражение в конкретных условиях места и времени.

Эти выражения взаимосвязи, имеющие большое практическое значение, могут быть статистически охарактеризованы при помощи различных методов, главнейшими из которых являются: метод приведения параллельных данных, аналитические группировки, балансовый метод. Важное место в изучении взаимосвязи общественных явлений занимает корреляционный анализ, которому посвящена настоящая глава.

Метод приведения параллельных данных

Наиболее простым приемом статистической характеристики взаимосвязи является приведение в табличной форме рядов статистических данных, отображающих развитие взаимосвязанных между собой явлений (см. гл. 9, табл. 9.12). Сопоставление параллельных рядов – метод, когда ряд значений факторного признака (x), построенный в порядке возрастания, сопоставляют с рядом соответствующих значений результативного признака (y) и таким образом прослеживают их взаимосвязь (см. табл. 8.1).

Метод аналитических группировок

Для выявления и характеристики взаимосвязи между признаками общественных явлений широко и плодотворно используется метод аналитических группировок.

Как уже показано в главе 3, этот метод позволяет не только констатировать связь между признаками, но и выявить причины, приведшие к тем или иным конкретным результатам (см. табл. 3.4).

Таблица 8.1

Сопоставление радов численности менеджеров и объемов продаж однотипных фирм в одном из регионов РУз в I кв. исследуемого года

Номер фирмы	Численность менеджеров, чел. x_i	Объем продаж, млн сум y_i
1	15	9,50
15	17	9,54
21	8	9,58 '
7	20	10,01
5	20	10,07
26	20	9,77
11	22	9,88
30	22	9,79
2	24	10,30
4	25	10,40
8	25	10,00
16	25	10,22
6	27	10,36
10	27	9,90
25	27	10,26
29	28	10,30
9	29	10,30
19	30	10,29
13	32	10,48
18	32	10,71
17	33	10,84
12	33	10,61
23	33	10,82
28	34	10,50
14	35	10,56
20	35	10,67
27	38	10,74
3	39	10,98
24	39	10,93
22	45	11,00

Данные табл. 8.1 свидетельствуют о наличии прямой корреляционной связи между численностью менеджеров и объемом продаж фирм. По мере возрастания количества менеджеров, увеличивается и объем продаж.

При аналитических группировках обычно в основу образования групп кладется факторный признак. Рассчитав затем для каждой группы обобщающие показатели результативного признака в виде средних или относительных величин, прослеживают, какое влияние оказывает изменение фактического признака на результативный. Так, например, данные приведенные в табл. 8.2 разбиваются на группы в зависимости от величины признака-фактора, и по каждой группе вычисляются средние значения результативного признака.

Таблица 8.2

Группировка туристических фирм по затратам на рекламу

Группы туристических фирм по затратам на рекламу, усл. ден. ед.	Число фирм в группе	Среднее число туристов, воспользовавшихся услугами данной группы фирм, человек
1	2	3
8	3	790
9	5	860
10	5	966
11	4	1063
12	3	1100
Итого	20	

Примечание. *Различие в величине среднего числа туристов каждой группы фирм, вычисленных в корреляционной (см. графу 8 табл. 8.4) и групповой таблицах* (см. *графу 3 табл. 8.2) объясняется тем, что при расчете средних в корреляционной таблице действительные значения результативного признака заменяются центральными значениями интервалов группировки.*

Сравнив средние значения результативного признака по группам, можно сделать вывод, что рост затрат туристических фирм на рекламу влечет за собой увеличение числа клиентов, пользующихся услугами фирмы, т.е. в рассматриваемом примере можно предположить наличие прямой корреляционной зависимости между признаками. Корреляционная зависимость отчетливо обнаруживается только при рассмотрении средних значений результативного признака, соответствующих определенным значениям факторного признака, так как при достаточно большом числе наблюдений в каждой группе влияние прочих случайных факторов при расчете групповой средней будет взаимопогашаться, и четче выступит зависимость.

Балансовый метод

Взаимосвязь между многими статистическими показателями вполне может быть охарактеризована при помощи построения балансов в виде соотношений между наличием и распределением тех или иных ресурсов. Например, *балансовая связь* показателей коммерческой деятельности характеризует зависимость между источниками формирования ресурсов (средств) и их использованием. Свое проявление она получает, например, в формуле товарного баланса:

$$О_н + П = В + О_к,$$

где $О_н$ – остаток товаров на начало изучаемого периода; П – поступление товаров за период; В – выбытие товаров в изучаемом периоде; $О_н$ – остаток товаров на конец периода.

Левая часть формулы баланса характеризует предложение товаров ($О_к$ + П), а правая часть – использование товарных ресурсов (В + $О_н$). Важное практическое значение формулы товарного баланса состоит в том, что при отсутствии количественного учета продажи товаров на основе формулы баланса определяют величину розничной реализации отдельных товаров.

Компонентный метод

Компонентные связи характеризуются тем, что изменение статистического показателя определяется изменением компонентов, входящих в этот показатель, как множители:

$$a = b \cdot c.$$

Например, в статистике коммерческой деятельности компонентные связи используются в индексном методе выявления роли отдельных факторов в совокупном изменении сложного показателя. Так, в гл. 10 показано, что индекс товарооборота в фактических ценах I_{qp} представляет произведение двух компонентов – индекса товарооборота в сопоставимых ценах I_q и индекса цен I_p т.е.

$$I_{qp} = I_q \cdot I_p.$$

Важная практическая значимость показателей, состоящих в компонентной связи, в том, что она позволяет определять величину одного из неизвестных компонентов: $I_q = I_{qp} : I_p$ или $I_p = I_{qp} : I_q$.

Графический метод Данный метод позволяет выявить наличие и направление связи между двумя признаками корреляции.

Используя данные об индивидуальных значениях признака факторных соответствующих ему значениях результативного признака можно построить в прямоугольных координатах точечный график, который называют «полем корреляции». Для нашего примера «поле корреляции» имеет следующий вид (см. рис. 8.2).

Положение каждой точки на графике определяется величиной двух признаков – уровнем затрат на рекламу и соответствующим ему числом туристов, пользующихся услугами данной фирмы. Точки корреляционного поля не лежат на одной линии, они вытянуты определенной полосой слева направо. Имеющийся в нашем распоряжении статистический материал был сгруппирован (табл. 8.2) и по каждому значению затрат фирм на рекламу определены значения среднего числа туристов в группе (см. гр. 3 табл. 8.2). Нанеся эти средние на график и соединяя последовательно отрезками прямых соответствующие им точки, получим так называемую эмпирическую линию связи.

Если эмпирическая линия связи по своему виду приближается к прямой линии, то можно предположить наличие прямолинейной корреляционной связи между признаками. Если же имеется тенденция неравномерного изменения значений результативного признака, и эмпирическая линия связи будет приближаться к какой-либо кривой, то это может быть связано с наличием криволинейной корреляционной связи (см. гл. 3).

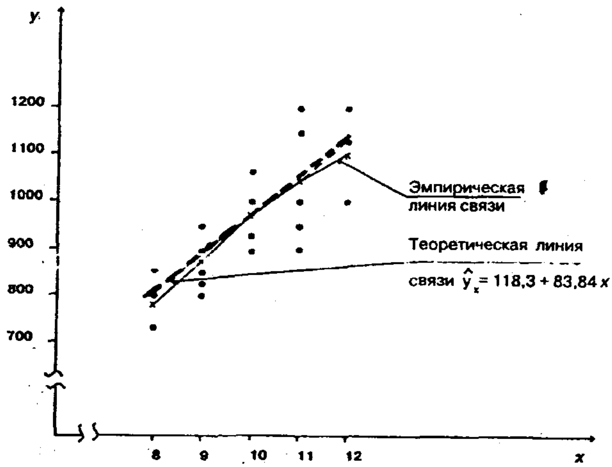


Рис. 8.2. **Зависимость числа клиентов фирмы (у) от ее затрат на рекламу** *(х)*

Метод корреляционной таблицы

Наличие большого числа различных значений результативного признака (у), сооответствующих одному и тому же значению признака и фактора (х) затрудняет восприятие таких параллельных рядов особенно при большом числе единиц, составляющих изучаемую совокупность.

В таких случаях целесообразнее воспользоваться для установления факта наличия связи статистическими таблицами – корреляционными или групповыми. Например, по 20 туристическим фирмам были установлены затраты на рекламу (факторный признак) и количество туристов воспользовавшихся услугами каждой фирмы (результативный признак).[1] В табл. 8.2 фирмы ранжированы по величине затрат на рекламу.

[1] Поскольку речь идет об изложении методологии изучения взаимосвязей, мы ограничились совокупностью малого объема. Данные в примере условные.

Таблица 8.3

Показатели, характеризующие связь между затратами на рекламу и количество туристов, воспользовавшихся услугами фирмы

Порядковые номера фирм	Затраты на рекламу, усл. ден. ед.	Количество туристов, воспользовавшихся услугами фирмы, человек	Порядковые номера фирм	Затраты на рекламу, усл. ден. ед.	Количество туристов, воспользовавшихся услугами фирмы, чел.
1	8	800	11	10	920
2	8	850	12	10	1060
3	8	720	13	10	950
4	9	850	14	11	900
5	9	800	15	11	1200
6	9	880	16	11	1150
7	9	950	17	11	1000
8	9	820	18	12	1200
9	10	900	19	12	1100
10	10	1000	20	12	1000

Можно увидеть, что в целом для всей совокупности фирм увеличение затрат на рекламу приводит к увеличению количества туристов, пользующихся услугами фирмы, хотя в отдельных случаях наличие такой зависимости может и не усматриваться. Например, сопоставим данные по фирмам с порядковыми номерами 7 и 11. Здесь мы видим даже обратное соотношение: у фирмы 11 количество туристов меньше, чем у фирмы 7 и составляет 920 человек, хотя затраты на рекламу выше, чем у фирмы 7 на 1 усл. ден. ед. В каждом отдельном случае количество туристов, воспользовавшихся услугами фирмы будет зависеть не только от размера затрат фирмы на рекламу, но и от того, как сложатся прочие факторы, определяющие величину результативного признака.

Построение **корреляционной таблицы** начинают с группировки значений факторного и результативного признаков. Так как в приводимом примере факторный признак представлен всего пятью вариантами повторяющихся значений, достаточно в первом столбце табл. 8.2 выписать эти результаты.

Для результативного признака **необходимо** определить величину интервала. Для этого воспользуемся формулой Стэрджесса:

$$h_y = \frac{y_{max} - y_{min}}{1 + 3{,}322 \lg n} = \frac{1200-720}{5} = 96$$ человек.

В корреляционной таблице факторный признак **x**, как правило, располагают в строках, а результативный признак **y** – в столбцах (графах) таблицы. Числа, расположенные на пересечении строк и столбцов таблицы, означают частоту повторения данного сочетания значения **x** и **y** (табл. 8.4).

Данная корреляционная таблица уже при общем знакомстве дает возможность выдвинуть предположение о наличии или отсутствии связи, а также выяснить ее направление. Если частоты в корреляционной таблице расположены на диагонали из левого верхнего угла в правый нижний угол (т.е. большим значениям фактора соответствуют большие значения функции), то можно предположить наличие прямой корреляционной зависимости между признаками. Если же частоты расположены по диагонали справа налево, то предполагают наличие обратной связи между признаками.

Уместно подчеркнуть, что при рассмотрении корреляционной таблицы важно установить расположение основной части частот. Возможны варианты, когда все клетки корреляционной таблицы окажутся заполненными. Однако это обстоятельство еще не означает, что корреляционная связь между признаками отсутствует. Нужно установить, как расположена в таблице основная масса частот. Для того чтобы сделать восприятие корреляционной таблицы более доступным и в целях более четкого выявления основной тенденции связи, можно для каждой строки рассчитать средние значения результативного признака, соответствующие определенному значению признака-фактора (каждая строка таблицы дает условное распределение y при определенном значении x).

Таблица 8.4

Корреляционная таблица

Центральное значение интервала, y	**768**	**865**	**962**	**1059**	**1156**	f_x	$\bar{y}_j$
группы по y / группы по x	**720-816**	**817-913**	**914-1010**	**1011-1107**	**1108-1207**		
1	2	3	4	5	6	7	8
8	2	1				3	800
9	1	3	1			5	865
10		1	3	1		5	962
11		1	1		2	4	1035
12			1	1	1	3	1059
f_y	3	6	6	2	3	20	

Примечание: *$\bar{y}_j$ – среднее значение результативного признака для j - той группы значений факторного признака; f_x – частота повторения данного варианта значения факторного признака во всей совокупности;*

f_y – частота повторения результативного признака во всей совокупности.

Среднее число туристов для первой группы, состоящей из трех фирм, которые тратят на рекламу 8 усл. ден. ед., будет равно 800:

$$\bar{y}_1 = \frac{768 \cdot 2 + 865 \cdot 1}{3} = 800 \text{ человек.}$$

Для следующей группы, состоящей из пяти фирм, у которых затраты на рекламу 9 усл. ден. ед. $\bar{y}_2 = \frac{768 \cdot 1 + 865 \cdot 3 + 962 \cdot 1}{5} =$ 865 человек и т.д. (рассчитанные таким образом средние представлены в графе 8 табл. 8.4).

Итак, увеличение средних значений результативного признака с увеличением значений факторного признака еще раз свидетельствует о возможном наличии прямой корреляционной зависимости числа туристов, воспользовавшихся услугами фирмы, от затрат фирмы на рекламу.

Корреляционная таблица позволяет сжато, компактно изложить материал, поэтому все последующие расчеты (показателей тесноты связи и параметров уравнения регрессии) можно вести по корреляционной таблице.

Корреляционно-регрессионый метод

Корреляционный анализ – это количественное определение тесноты связи между факторным и результативными признаками, осуществляемой с помощью показателей корреляции.

Регрессионый анализ

Регрессионный анализ – определение теоретического выражения связи между признаками, т.е. определение формы связи (построение уравнения регрессии). Корреляционный и регрессионный виды анализа тесно взаимосвязаны между собой. Установив при помощи элементарных статистических методов наличие и направление связи между признаками, на следующем этапе анализа переходят к построению регрессионной модели зависимости результативного признака от факторного и оценке тесноты корреляционной связи, т.е. переходят к корреляционно-регрессионному анализу.

Более подробно эти два вида анализа будут рассмотрены ниже.

Метод корреляции рангов

Этот метод называется коэффициентом корреляции рангов Спирмена и рассчитывается на базе ранжированного ряда изучаемых единиц (объектов) по занимаемому месту. Формула расчета имеет следующий вид:

$$R = 1 - \frac{6(x-y)^2}{n(n^2-1)}.$$

Покажем расчет этого коэффициента на примере связи между ростом населения (x) и ростом валовой внутренней продукции на душу населения (y). табл. 8.5.

Таблица 8.5

Ранжированный ряд областей по удельному весу населения в общей численности

Ранжированный ряд		**Место областей по удельному весу в производстве ВВП республики**	**Разность мест**	**Квадрат разности мест**
областей по удельному весу населения в общей численности населения РУз	**место**			
(К)	(x)	(y)	(x–y)	$(x–y)^2$
Самаркандская	1	6	–5	25
Ферганская	2	4	–2	4
Кашкадарьинская	3	3	0	0
Андижанская	4	5	–1	1
Ташкентская	5	2	3	9
Наманганская	6	9	–3	9
г. Ташкент	7	1	6	36
Сурхандарьинская	8	10	–2	4
Бухарская	9	7	2	4
Республика КК	10	12	–2	4
Хорезмская	11	11	0	0
Джизакская	12	13	–1	1
Навоийская	13	8	5	25
Сырдарьинская	14	14	0	0
Итого	–	–		$\Sigma(x-y)^2=122$

Отсюда

$$R=1-\frac{6\Sigma(x-y)^2}{n(n^2-1)}=1-\frac{6}{14(14^2-1)}=1-\frac{732}{14\cdot195}=$$

$$1-\frac{732}{2730}=1-0{,}268=0{,}732$$ или 73,2%.

Следовательно, связь между численностью и выпуском ВВП существует высокая связь, о чем свидетельствует шкалы Чеддока (см. табл. 8.9).

Следует отметить, что если последовательность рангов обеих переменных совпадают, то коэффициент корреляции рангов Спирмена равняется 1. Это наблюдается в Кашкадарьинской, Хорезмской и Сырдарьинской областях. Так как в этом случае:

$$x_i - y_i = 0$$

для каждого значения:

$$i = 1, 2, 3, \ldots, n,$$

и следовательно, числитель дроби в правой части формулы равняется нулю, поэтому

$$R = 1.$$

Можно также показать, что коэффициент корреляции рангов R = – 1, если последовательность рангов, соответствующих переменным x и y, имеют противоположные тренды (один возрастает от 1 до n, а второй понижается от n до 1).

$$-1 \leq R \leq 1$$

и, следовательно, R – это, как и обычный коэффициент корреляции r – удобная мера связи признаков.

Коэффициент Спирмена можно применять и в тех случаях когда изучаемым признакам нельзя придать количественного выражения, однако есть возможность осмысленного упорядочения эмпирических наблюдений над каждым из признаков.

8.3. Задачи корреляционно-регрессионного анализа

Корреляционный анализ решает две основные задачи

Первая задача заключается в определении **формы связи**, т.е. в установлении математической формы, в которой выражается данная связь. Это очень важно, так как от правильного выбора формы связи зависит конечный результат изучения взаимосвязи между признаками.

Вторая задача состоит в измерении **тесноты, т.е. меры связи** между признаками с целью установить степень влияния данного фактора на результат. Она решается математически, путем определения параметров корреляционного уравнения.

Затем проводятся оценка и анализ полученных результатов при помощи специальных показателей корреляционного метода (коэффициентов детерминации, линейной и множественной корреляции и т.д.), а также проверка существенности связи между изучаемыми признаками.

Выбор формы связи

Определяющая роль в выборе формы связи между явлениями принадлежит теоретическому анализу. Так, например, чем больше размер основного капитала предприятия (факторный признак), тем больше при прочих равных условиях оно выпускает продукции (результативный признак). С ростом факторного признака здесь, как правило, равномерно растет и результативный, поэтому зависимость между ними может быть выражена уравнением прямой $Y = a_0 + a_1x$, которое называется **линейным уравнением регрессии.**

Параметр a_1 называется **коэффициентом регрессии** и показывает, насколько в среднем отклоняется величина результативного признака у при отклонении величины факторного признака x на одну единицу. При $x = 0$ $a_0 = Y$. Увеличение количества внесенных удобрений приводит, при прочих равных условиях, к росту урожайности, но чрезмерное внесение их без изменения других элементов к дальнейшему повышению урожайности не приводит, а, наоборот, снижает ее. Такая зависимость может быть выражена уравнением параболы $Y = a_0 + a_1x + a_2x^2$.

Параметр $\mathbf{a_2}$ характеризует степень ускорения или замедления кривизны параболы, и при $a_2 > 0$ парабола имеет минимум, а при $a_2 < 0$ – максимум. Параметр a_1 характеризует угол наклона кривой, а параметр $\mathbf{a_0}$ – начало кривой.

Однако с помощью теоретического анализа не всегда удается установить форму связи. В таких случаях приходится только предполагать о наличии определенной формы связи. Проверить эти предположения можно при помощи графического анализа, который используется для выбора формы связи между явлениями, хотя графический метод изучения связи применяется и самостоятельно.

Аналитическое выражение связи

Применение методов корреляционного анализа дает возможность выражать связь между признаками аналитически – в виде уравнения – и придавать ей количественное выражение. Рассмотрим применение приемов корреляционного анализа на конкретном примере.

Допустим, что между стоимостью основного капитала и выпуском продукции существует прямолинейная связь, которая выражается **уравнением прямой** $Y = a_0 + a_1x$. Необходимо найти параметры $\mathbf{a_0}$ и $\mathbf{a_1}$, что позволит определить теоретические значения Y для разных значений x. Причем $\mathbf{a_0}$ и $\mathbf{a_1}$ должны быть такими, чтобы было достигнуто максимальное приближение к первоначальным (эмпирическим) значениям y теоретических значений Y. Эта задача решается при помощи способа наименьших квадратов, основное условие которого сводится к определению параметров a_0 и a_1, таким образом, чтобы $\Sigma(y_i - Y)^2 = \min$[1]. Математически доказано, что условие минимума обеспечивается, если параметры $\mathbf{a_0}$ и $\mathbf{a_1}$ определяются при помощи **системы** двух **нормальных уравнений**, отвечающих требованию метода наименьших квадратов:

$$\Sigma y = na_0 + a_1\Sigma x_i;$$
$$\Sigma xy = a_0\Sigma x + a_1\Sigma x^2.$$

Первое уравнение есть сумма всех первоначальных уравнений. Второе получается умножением обеих частей уравнения прямой на один и тот же множитель. Математически доказано, что условие $\Sigma(y_i - Y)^2 = \min$ ***соблюдается***, если в качестве такого множителя принять значение факторного признака, т.е. если уравнение прямой умножить на x.

Кроме рассмотренных функций связи в экономическом анализе часто применяются степенная, показательная и гиперболическая функции.

Степенная функция Это функция имеет вид $Y = a_0x^{a_1}$. Параметр степенного уравнения называется показателем эластичности и указывает, на сколько процентов изменится y при возрастании **x** на 1%. При **x** = 1 a_0 = Y.

[1] Этот метод впервые разработан К.Ф.Гаусом (1777–1865).

Для определения параметров степенной функции вначале ее приводят к линейному виду путем логарифмирования: lg y = lg a_0 + a_1 lg x_1, а затем строят систему нормальных уравнений:

$$\Sigma \lg y = n \lg a_0 + a_1 \Sigma \lg x;$$

$$\Sigma \lg y \lg x = \lg a_0 \Sigma \lg x + a_1 \Sigma \lg x^2.$$

Решив систему двух нормальных уравнений, находят логарифмы параметров логарифмической функции **a_0** и **$a_г$**, а затем и сами параметры **a_0** и **a_1**. При помощи степенной функции определяют, например, зависимость между фондом заработной платы и выпуском продукции, затратами труда и выпуском продукции и т.д.

Если факторный признак **x** растет в арифметической прогрессии, а результативный **y** – в геометрической, то такая зависимость выражается **показательной функцией** $Y = a_0 a_1^x$. Для определения параметров показательной функции ее также вначале приводят к линейному виду путем логарифмирования: lg y = lg a_0 + x lg a_1, а затем строят систему нормальных уравнений:

$$\Sigma \lg y = n \lg a_0 + \lg a_1 \Sigma x;$$

$$\Sigma x \lg y = \lg a_0 \Sigma x + \lg a_1 \Sigma x^2.$$

Вычислив соответствующие данные и решив систему двух нормальных уравнений, находят параметры показательной функции **a_0** и **a_1**.

В ряде случаев обратная связь между факторным и результативным признаками может быть выражена **уравнением гиперболы:**

$$Y = a_0 + \frac{a_1}{x}.$$

И здесь задача заключается в нахождении параметров a_0 и a_1 при помощи системы двух нормальных уравнений:

$$\Sigma y = na + a_1 \Sigma \frac{1}{x};$$

$$\Sigma y\frac{1}{x}=a_0\Sigma\frac{1}{x}+a_1\Sigma\frac{1}{x^2}$$

При помощи гиперболической функции изучают, например, связь между выпуском продукции и себестоимостью, уровнем издержек обращения (в процентах к товарообороту) и товарооборотом в торговле, сроками уборки и урожайностью и т.д.

Таким образом, применение различных функций в качестве уравнения связи сводится к определению параметров уравнения по способу наименьших квадратов при помощи системы нормальных уравнений.

В малых совокупностях значение коэффициента регрессии подвержено случайным колебаниям. Поэтому возникает необходимость в определении-достоверности коэффициента регрессии.

Достоверность коэффициента регрессии определяется так же, как и в выборочном наблюдении, т.е. устанавливаются средняя и предельная ошибки для выборочной средней и доли.

Средняя ошибка коэффициента регрессии определяется по формуле

$$\mu_B=\sqrt{\frac{\sigma_0^2}{\sigma^2(n-2)}},$$

где σ_0^2 – случайная дисперсия; σ^2 – общая дисперсия, а n – число коррелируемых пар.

8.4. Измерение тесноты связи

Для определения тесноты корреляционной связи при проверке адекватности регрессионной модели (т.е. соответствия эмпирическим данным) рассчитывают следующие показатели корреляции.

1. **Линейный коэффициент корреляции** помимо тесноты (или силы) связи показывает и ее направление; определяется по следующей формуле:

$$r=\frac{\overline{xy}-\overline{x}\,\overline{y}}{\sigma_x\sigma_y}=\frac{\overline{xy}-\overline{x}\,\overline{y}}{\sqrt{\overline{x^2}-\overline{x}^2}\sqrt{\overline{y^2}-\overline{y}^2}}.$$

Линейный коэффициент корреляции принимает значение:

$$-1\le r\le +1$$

Со знаком (+) – прямая связь, со знаком (–) – обратная связь, при r = 0 – линейная связь отсутствует, при r = ±1 – связь функциональная (линейная). Чем ближе линейный коэффициент корреляции к ± 1, тем корреляционная связь теснее.

2.**Линейный коэффициент детерминации (R^2)** – квадрат линейного коэффициента корреляции. Его числовые значения всегда заключаются в пределах от нуля до единицы:

$$0<r^2<+1.$$

Линейный коэффициент детерминации – более жесткий показатель тесноты связи, чем линейный коэффициент корреляции.

3. ***Теоретическое корреляционное отношение рассчитывается по формуле:***

$$\eta_T=\sqrt{\frac{\delta_T^2}{\sigma^2}}$$

где δ_T^2 – межгрупповая дисперсия выравненных значений результативного признака, т.е. рассчитанных по уравнению регрессии; σ^2 – общая дисперсия эмпирических значений результативного признака.

Общая дисперсия определяется по уже известной формуле

$$\sigma^2_{общ}=\frac{\Sigma(y-\overline{y})^2f}{\Sigma f}$$

Межгрупповая дисперсия выравненных значений результативного признака определяется по формуле:

$$\delta_T^2=\frac{\sum_{j=1}^{k}(y_j-\overline{y})^2f_j}{\sum_{j=1}^{k}f_j}$$

где $\hat{y}_j$ – теоретическое значение результативного признака в j-й группе.

При расчете теоретического корреляционного отношения можно использовать правило сложения дисперсий, которое в этом случае может быть представлено следующим образом:

$$\sigma^2 = \delta_T^2 + \sigma_{ост}^2,$$

где $\sigma_{ост}^2$ – остаточная дисперсия.

Тогда теоретическое корреляционное отношение можно рассчитать по следующей формуле:

$$\eta_T = \sqrt{\frac{\sigma^2 - \sigma_{ост}^2}{\sigma^2}} = \sqrt{1 - \frac{\sigma_{ост}^2}{\sigma^2}}.$$

Теоретическое корреляционное отношение, исчисляемое по этой формуле часто называют **индексом (коэффициентом) корреляции (R).**

Теоретическое корреляционное отношение η_T является более универсальным показателем тесноты связи, чем линейный коэффициент корреляции r, поскольку может использоваться как для прямолинейных, так и криволинейных зависимостей.

Теоретическое корреляционное отношение не следует путать с эмпирическим корреляционным отношением, которое также используется в корреляционном анализе, но строится непосредственно по фактическим данным. Эмпирическое корреляционное отношение является показателем рассеяния точек корреляционного поля относительно эмпирической линии регрессии, закономерности, изменения которой нарушаются случайными зигзагами ломаной, возникающими вследствие действия неучтенных факторов. Теоретическое корреляционное отношение, в свою очередь, характеризует рассеяние точек корреляционного поля относительно теоретической (выравненной) линии регрессии.

4.Теоретический коэффициент детерминации определяется как квадрат теоретического корреляционного отношения:

$$\eta_T^2 = \frac{\delta_T^2}{\sigma^2}.$$

По аналогии с индексом (коэффициентом) корреляции теоретический коэффициент детерминации может обозначаться как R^2.

При прямолинейной связи коэффициент корреляции по своей абсолютной величине равен индексу корреляции: $|r| = R$.

Если индекс корреляции возвести в квадрат, то получим **коэффициент детерминации**

$$R^2 = \frac{\sigma_T^2}{\sigma^2}$$

Он характеризует роль факторной вариации в общей вариации и по построению аналогичен корреляционному отношению η^2. Как и корреляционное отношение, коэффициент детерминации R^2 может быть исчислен при помощи дисперсионного анализа, так как дисперсионный анализ позволяет расчленить общую дисперсию на факторную и случайную. Однако при дисперсионном анализе для разложения дисперсии пользуются методом группировок, а при корреляционном анализе – корреляционными уравнениями.

Коэффициент детерминации является наиболее конкретным показателем, так как он отвечает на вопрос о том, какая доля в общем результате зависит от фактора, положенного в основание группировки.

При прямолинейной парной связи факторную дисперсию можно определить без вычисления теоретических значений Y по следующей формуле:

$$\sigma^2 = \frac{1}{n}\left(a_0\Sigma y + a_1\Sigma xy\right) - \bar{y}^2.$$

8.5. Проверка адекватности регрессионной модели

Корреляционные связи, как ранее указывалось, проявляется в массовом статистическом материале. Поэтому при проведении корреляционно-регрессионного анализа (особенно, если исследуемая совокупность образована на основе малой выборки) могут возникнуть сомнения в том, что обнаруженная связь носит закономерный, а не случайный характер. Таким образом, необходимо оценить адекватность (соответствия эмпирическим данным) полученной регрессионной модели.

Для малых выборок и парных регрессий наиболее известен метод оценки значимости корреляционных зависимостей на основе использования критерия, предложенного английским статистиком и названного его именем – **критерия Р. Фишера (критерия дисперсионного отношения):**

$$F_{расч} = \frac{\sigma^2_{факт}}{\sigma^2_{ост}},$$

где $\sigma^2_{факт}$ – факторная дисперсия; $\sigma^2_{ост}$ – остаточная дисперсия.

Факторная дисперсия определяется по следующей формуле:

$$\sigma^2_{факт} = \frac{\sum_{j=1}^{k}(y_i - \bar{y})^2 f_i}{k-1},$$

где $(k - 1)$ – число степеней свободы для $\sigma^2_{факт}$.

Остаточную дисперсию, используя правило сложения дисперсий, можно определить по формуле:

$$\sigma^2_{ост} = \frac{\sum_{j=1}^{n}(y_i - \bar{y})^2 f_i - \sum_{j=1}^{k}(y_j - \bar{y})^2 f_j}{n-k}$$

где $(N - K)$ – число степеней свободы для $\sigma^2_{ост}$.

Число степеней свободы для общей суммы квадратов отклонений:

$$(k-1)+(n-k)=n-1.$$

Если расчетное значение критерия **Фишера** больше табличного, то гипотеза о несоответствии заложенных в уравнение регрессии связей отвергается. Табличные значения F-критерия определяются на основании уровней значимости (0,05 и 0,01), а также по числу степеней свободы.

Поясним все сказанное примером

Задание 1

На базе данных талб. 8.1 требуется построить:

1. Линейное уравнение парной регрессии, отражающее взаимосвязь между указанными признаками.

2. График теоретической линии зависимости объемов продаж от численности менеджеров фирм.

Решение:

1. Для построения линейного уравнения регрессии необходимо определить параметры этого уравнения: свободный член уравнения a_0 и коэффициент регрессии a_1 для чего построена табл. 8.6.

Таблица 8.6

Вспомогательная таблица для определения параметров линейного уравнения регрессии

Исходные данные			Расчетные данные	
номера фирм	численность менеджеров, чел., x_i	объем продаж, млн сум, y_i	x_i^2	$x_i y_i$
1	15	9,50	225	142,50
15	17	9,54	289	162,18
21	18	9,58	324	172,44
5	20	10,07	400	201,40
7	20	10,01	400	200,20
26	20	9,77	400	195,40
11	22	9,88	484	217,36
30	22	9,79	484	215,38
2	24	10,30	576	247,20
4	25	10,40	625	260,00
8	25.	10,00	625	250,00
16	25	10,22	625	255,50
6	27	10,36	729	279,72
10	27	9,90	729	267,30
25	27	10,26	729	277,02
29	28	10,30	784	288,40
9	29	10,30	841	298,70
19	30	10,29	900	308,70
13	32	10,48	1024	335,36
18	32	10,71	1024	342,72

12	33	10,61	1089	350,13
17	33	10,84	1089	357,72
23	33	10,82	1089	357,06
28	34	10,50	1156	357,00
14	35	10,56	1225	369,60
20	35	10,67	1225	373,45
27	38	10,74	1444	408,12
3	39	10,98	1521	428,22
24	39	10,93	1521	426,27
22	45	11,00	2025	495,00
Итого	849	309,31	25601	8840,05

Параметры линейного уравнения регрессии определим по формулам

$$a_1 = \frac{\overline{xy} - \overline{x}\,\overline{y}}{\overline{x^2} - \overline{x}^2} = \frac{\frac{8840{,}05}{30} - \frac{849}{30} \cdot \frac{309{,}31}{30}}{\frac{25601}{30} - \left(\frac{849}{30}\right)^2} = \frac{294{,}67 - 28{,}3 \cdot 10{,}31}{853{,}37 - 800{,}89} = \frac{2{,}9}{52{,}48} = 0{,}055;$$

$$a_0 = \bar{y} - a_1\bar{x} = 10{,}31 - 0{,}055 \cdot 28{,}3 = 8{,}754.$$

Тогда линейное уравнение парной регрессии примет вид:

$$\hat{y} = 8{,}754 + 0{,}055\,x.$$

2. Для построения теоретической линии зависимости объемов продаж от численности менеджеров фирм необходимо знать теоретические значения результативного признака – объемов продаж фирм.

Их определяем, используя полученное линейное уравнение парной регрессии.

Так как в нашем случае – линейная зависимость, то для того, чтобы получить искомые точки для построения теоретической линии зависимости, нам достаточно подставить в это уравнение два значения факторного признака (например, два крайних его значения – 15 и 45).

$$\hat{y}_{\min} = 8{,}754 + 0{,}055 \cdot 15 = 9{,}579 \text{ (млн сум)};$$

$$\hat{y}_{\max} = 8{,}754 + 0{,}055 \cdot 45 = 11{,}229 \text{ (млн сум)}.$$

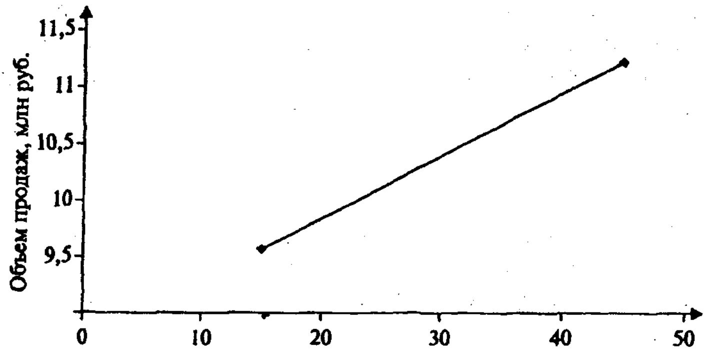


Рис. 8.3. **График теоретической линии зависимости объемов продаж от численности менеджеров фирм в одном из регионов РУз в I кв. исследуемого года**

Задание 2

Используя полученные данные задания 1 сквозной задачи, *требуется* определить линейный коэффициент корреляции и сделать *выводы* о силе связи между численностью менеджеров и объемом продаж.

Решение:

Прежде всего построим вспомогательную табл. 8.7.

Используя полученные данные вспомогательной табл. 8.7, определим линейный коэффициент корреляции:

$$r=\frac{\overline{xy}-\bar{x}\,\bar{y}}{\sqrt{\overline{x^2}-\bar{x}^2}\cdot\sqrt{\overline{y^2}-\bar{y}^2}}=\frac{\frac{8840{,}05}{30}-\frac{849}{30}\cdot\frac{309{,}31}{30}}{\sqrt{\frac{25601}{30}-\left(\frac{849}{30}\right)^2}\cdot\sqrt{\frac{3194{,}59}{30}-\left(\frac{309{,}31}{30}\right)^2}}=$$

$$=\frac{294{,}67-28{,}3\cdot 10{,}31}{\sqrt{853{,}37-800{,}89}\sqrt{106{,}49-106{,}30}}=\frac{294{,}67-291{,}7}{\sqrt{52{,}48}\sqrt{0{,}19}}=$$

$$\frac{2{,}90}{7{,}24\cdot 0{,}44}=\frac{2{,}90}{3{,}19}=0{,}91$$

Таблица 8.7

Вспомогательная таблица для определения линейного коэффициента корреляции

Исходные данные			Расчетные данные		
номер фирмы	численность менеджеров, чел., x_i	объем продаж, млн сум, y_i	x_i^2	y_i^2	x_iy_i
1	15	9,50	225	90,2500	142,50
15	17	9,54	289	91,0116	162,18
21	18	9,58	324	91,7764	172,44
5	20	10,07	400	101,4049	201,40
7	20	10,01	400	100,2001	200,20
26	20	9,77	400	95,4529	195,40
11	22	9,88	484	97,6144	217,36
30	22	9,79	484	95,8441	215,38
2	24	10,30	576	106,0900	247,20
4	25	10,40	625	108,1600	260,00
8	25	10,00	625	100,0000	250,00
16	25	10,22	625	104,4484	255,50
6	27	10,36	729	107,3296	279,72
10	27	9,90	729	98,0100	267,30
25	27	10,26	729	105,2676	277,02
29	28	10,30	784	106,0900	288,40
9	29	10,30	841	106,0900	298,70
19	30	10,29	900	105,8841	308,70
13	32	10,48	1024	109,8304	335,36
18	32	10,71	1024	114,7041	342,72
12	33	10,61	1089	112,5721	350,13
17	33	10,84	1089	117,5056	357,72
23	33	10,82	1089	117,0724	357,06
28	34	10,50	1156	110,2500	357,00
14	35	10,56	1221	111,5136	369,60
20	35	10,67	1225	113,8489	373,45
27	38	10,74	1444	115,3476	408,12
3	39	10,98	1521	120,5604	428,22
24	39	10,93	1521	119,4649	426,27
22	45	11,00	2025	121,0000	495,00
Итого	849	309,31	25601	3194,5941	8840,05

Таким образом, связь между объемом продаж и численностью менеджеров фирм прямая и близка к функциональной (линейной).

Сквозная задача

Задание 3

Используя расчетные данные заданий 1 и 2 сквозной задачи, требуется определить теоретическое корреляционное отношение и сделать выводы.

Решение:

Теоретическое корреляционное отношений определяется по формуле:

$$\eta_T = \sqrt{\frac{\delta_T^2}{\sigma^2}},$$

где σ^2 – общая дисперсия результативного признака (объема продаж), полученная на базе эмпирического материала. Ее значение мы возьмем из решения задания 6.3. $\sigma^2 = 0,1834$ (см. гл.6).

Для определения δ_T^2 построим вспомогательную табл. 8.8 на базе табл. 8.5, с учетом полученного линейного уравнения регрессии (см. решение задания 1).

Таблица 8.8

Вспомогательная таблица для расчета межгрупповой дисперсии выравненных значений результативного признака

(объема продаж)

Номер группы	Число фирм, ед. $f_{ю}$	Центральное значение численности менеджеров в группе, чел., $x_{цi}$	Теоретическое значение объемов продаж фирмы, млн сум, $\hat{y}_j$	$(y_j - \bar{y})^2 f_j$, где $\bar{y}$ = 10,31
1	3	17,5	9,7165	1,0567
2	6	22,5	9,9915	0,6087
3	8	27,5	10,2665	0,0151
4	7	32,5	10,5415	0,3751
5	5	37,5	10,8165	1,2827
6	1	42,5	11,0915	0,6107
Итого	30	–	10,3100	3,9490

$$\delta_T^2 = \frac{\sum_{j=1}^{k}(y_j - \bar{y})^2 f_i}{\sum_{j=1}^{k} f_i} = \frac{3,949}{30} = 0,1316;$$

$$\eta_T = \sqrt{\frac{0,1316}{0,1834}} = \sqrt{0,7176} = 0,8471.$$

На основании шкалы Чеддока можно сделать заключение о том, что связь между объемом продаж и численностью менеджеров фирм тесная (табл. 8.9).

Таблица 8.9

Шкалы Чеддока

Показания тесноты связи	0,1 -0,3	0,3–0,5	0,5–0,7	10,7–0,9	0,9–0,90
Характеристика силы связи	слабая	умеренная	заметная	высокая	*весьма высокая*

Сквозная задача

Задание 4

Используя полученные данные заданий 2 и 3 сквозной задачи, *требуется* дать оценку значимости корреляционной зависимости объемов продаж от численности менеджеров.

Решение:

$$F_{расч} = \frac{\sigma^2_{факт}}{\sigma^2_{ост}} = \frac{\sum_{j=1}^{k}(y_j - \bar{y})^2 f_j}{k-1} : \frac{\sum_{j=1}^{k}(y_j - \bar{y})^2 - \sum_{j=1}^{k}(y_j - \bar{y})^2 f_j}{n-k};$$

$$F_{расч} = \frac{3,9490}{6-1} : \frac{5,5036 - 3,9490}{30-6} = 0,7898 : 0,0648 = 12,19;$$

$F_{табл}$ = 4,53 и 9,47 (эти табличные значения F-критерия берутся из справочников по числу степеней свобод: 5 и 24 и для уровней значимости: 0,05 и 0,01).

Таким образом, $\mathbf{F_{расч} > F_{табл}}$, т.е. гипотеза о несоответствии заложенных в уравнение, регрессии связей отвергается, или, другими словами, построенное корреляционное уравнение зависимости является значимым, и связь между численностью менеджеров и объемом продаж является существенной.

Если п > 30.

Применительно к совокупностям, у которых n < 30, <u>для провер</u>ки типичности параметров уравнения регрессии <u>ис-</u>

пользуется t-критерий Стъюдента. При этом вычисляются фактические значения t-критерия:

для параметра a_0

$$t_{a_0} = a_0 \frac{\sqrt{n-2}}{\sigma_s}$$

для параметра a_1

$$t_{a_1} = a_1 \frac{\sqrt{n-2} \cdot \sigma_x}{\sigma_s}$$

где $\sigma_s = \sqrt{\frac{\Sigma(y_i - y_{xi})^2}{n}}$ – среднее квадратическое отклонение результативного признака $У_I$ от выравненных значений $У_{XI}$;

$\sigma_x = \sqrt{\frac{\Sigma(x_i - \bar{x})^2}{n}}$ – среднее квадратическое отклонение факторного признака X_I от общей средней $\bar{x}$.

Полученные по вышеуказанным формулам фактические значения t_{a_0} и t_{a_1} сравниваются с критическим t_k, который получают по таблице Стьюдента с учётом принятого уровня значимости а и числа степеней свободы К *(см. приложение).*

Полученные в анализе корреляционной связи параметры уравнения регрессии признаются типичными, если $t_Ф$ фактическое больше t_K критического:

$$t_{a_0} > t_k < t_{a_1}$$

По проверенным на типичность параметрам уравнения регрессии производится синтезирование (построение) математической модели связи. При этом параметры примененной в анализе математической функции получают соответствующие количественные значения.

Смысловое содержание синтезированных таким образом моделей состоит в том, что они характеризуют среднюю величину результативного признака $\tilde{y}_x$ в зависимости от вариации признака-фактора X.

Важным этапом корреляционного анализа связи является оценка практической значимости синтезированных моделей. Смысл такой оценки состоит в том, чтобы обосновать применение метода функционального анализа при изучении корреляционной зависимости. Правомерность такого приема анализа будет оправданной лишь в тех случаях, если изучаемая корреляционная (соотносительная) связь не столь значительно отстоит от функциональной (жесткой) связи. При этом необходимо доказать, что применение метода функционального анализа при изучении корреляционной зависимости не дает существенных погрешностей.

Проверка практической значимости синтезированных в корреляционно-регрессионном анализе математических моделей осуществляется посредством показателей тесноты связи между признаками x и y. (см. решение задач 1, 2, 3, 4).

8.6. Множественная корреляция

До сих пор мы рассматривали корреляционные связи между двумя признаками: результативным (у) и факторным (х). Например, выпуск продукции зависит не только от размера основного капитала, но и от уровня квалификации рабочих, состояния оборудования, обеспеченности и качества сырья и материалов, организации труда и т.д. В связи с этим возникает необходимость в изучении, измерении связи между результативным признаком, двумя и более факторными. Этим занимается множественная корреляция.

Построение уравнений множественной корреляции

Множественной называется такая корреляция, при которой результативный признак связан с несколькими факторными признаками. Например, зависимость между урожайностью и такими факторами, как качество почвы, сроки и способы посева, размеры внесенного удобрения, сроки уборки, зависимость дневной выработки бетонщиков от уровня механизации труда на формовке и одновременно квалификации рабочих и др.

В случае множественной корреляции, так же как и при парной, в соответствии с характером связи возможно построение как прямолинейных, так и криволинейных корреляционных уравнений. Например, выше было показано, что между дневной выработкой бетонщиков и уровнем механизации труда существует прямолинейная форма связи. Точно так же прямолинейная форма корреляционного уравнения может быть выведена и для связи между величиной дневной выработки и квалификацией рабочих. Исходя из этого составим линейное уравнение множественной корреляции, выражающее зависимость величины дневной выработки бетонщиков (y) от уровня механизации труда на формовке (x) и квалификации рабочих, (v):

$$\bar{y}_{xv} = a_0 + a_{1x} + a_{2v},$$

где a_0 – средняя дневная выработка неквалифицированного бетонщика на ручной формовке, a_x – среднее значение разницы между уровнем дневной выработки при выполнении работы механизированным и ручным способом и a_2 – среднее значение роста дневной выработки при повышении квалификации рабочих на один разряд.

Необходимо отметить, что это уравнение будет правильно выражать форму связи между рассматриваемыми признаками при условии независимости между любым из параметров и любым из факторных признаков либо такой слабой связи между параметрами и факторными признаками, которой можно пренебречь. Но для выражения формы связи не имеет значения, связаны ли отдельные факторные признаки между собой либо в такой связи находятся параметры.

Выше указанное уравнение содержит три параметра: $\mathbf{a_0}$, $\mathbf{a_1}$ и $\mathbf{a_2}$. Для их нахождения способом наименьших квадратов составляется и решается следующая система нормальных уравнений:

$$\left.\begin{aligned} a_0 n + a_1\Sigma x + a_2\Sigma v &= \Sigma y \\ a_0\Sigma x + a_1\Sigma x^2 + a_2\Sigma xv &= \Sigma xy \\ a_0\Sigma v + a_1\Sigma xv + a_2\Sigma v^2 &= \Sigma vy \end{aligned}\right\}$$

В расчетной табл. 8.10 имеются все данные для составления нормальных уравнений множественной корреляции.

Таблица 8.10

Расчет сумм для определения коэффициентов системы нормальных уравнений по не сгруппированным данным

№ п/п	x	v	y	xy	x2	vy	v2	xv	y2
1	35	2	3,0	105,0	1 225	6,0	4	70	9,00
2	59	3	6,5	383,6	3 481	19,5	9	177	42,25
3	44	3	4,8	211,2	1 936	14,4	9	132	23,04
4	55	3	5,7	313,5	3 025	17,1	9	165	32,49
5	39	2	2,8	109,2	1 521	5,6	4	78	7,84
6	56	3	4,7	263,2	3 136	14,1	9	168	22,09
7	78	2	4,2	327,6	6 084	8,4	4	156	17,64
8	44	4	5,3	233,2	1 936	21 ,2	16	176	28,09
9	43	3	2,0	86,0	1 849	6,0	9	129	4,00
10	76	3	6,5	494,0	5 776	19,5	9	228	30,25
11	58	3	5,1	295,8	3 364	15,3	9	174	26,01 ‘
12	41	4	5,5	225,5	1 681	22,0	16	164	30,25
13	49	о	3,0	147,0	2 401	6,0	4	98	9,00
14	58	о	3,6	208,8	3 364	7,2	4	116	12,96
15	58	4	4,5	261,0	3 364	18,0	16	232	20,25
16	61	4	6,7	408,7	3 721	26,8	16	244	44,89
17	42	3	5,6	235,2	1 764	16,8	9	126	31,36
18	46	3	5,2	239,2	2 116	15,6	9	138	27,04
19	35	о	3,2	112,0	1 225	6,4	4	70	10,24
20	55	4	5,4	297,0	3 025	21,6	16	220	29,16
21	38	3	4,5	171,0	1 443	13,5	9	114	20,25
22	35	3	5,5	192,5	1 225	16,5	9	105	30,25
23	25	3	2,5	62,5	625	7,5	9	75	6,25
24	90	4	6,2	558,0	8 100	24,8	16	360	38,44
25	47	2	4,1	192,7	2 209	8,2	4	94	16,81
26	69	3	5,0	345,0	4 761	15,0	9	207	25,00
27	48	2	2,5	120,0	2 304	5,0	4	96	6,25
28	82	4	6,8	557,6	6 724	27,2	16	328	46,24
29	98	4	6,6	646,8	9 604	26,4	16	392	43,56
30	63	3	6,3	396,9	3 969	18,9	9	189	39,69
31	79	4	7,9	624,1	6 241	31,6	16	316	62,41
32	41	3	4,6	188,6	1 681	13,8	9	123	21,16
33	45	3	4,2	189,0	2 025	12,6	9	135	17,64
34	75	2	4,8	360,0	5 625	9,6	4	150	23,04
35	45	3	5,8	261,0	2 025	17,4	9	135	33,64
36	51	4	4,9	249,9	2 601	19,6	16	204	24,01

37	55	3	4,3	236,5	3 0.25	12,9	9	165	18,49
38	95	4	6,4	608,0	9 025	25,6	16	380	40,96
39	90	4	7,0	630,0	8 100	28,0	16	360	49,00
40	70	4	7,1	497,0	4 900	28,4	16	280	50,41
41	56	3	4,4	246,4	3136	13,2	9	168	19,36
42	57	3	5,1	290,7	3 249	15,3	9	171	26,01
43	48	2	5,0	240,0	2 304	10,0	4	96	25,00
44	72	3	6,0	439,2	5 184	18,3	9	216	37,21
45	52	3	5,9	306,8	2 704	17,7	9	156	34,81
46	33	2	3,8	125,4	1 089	7,6	4	66	14,44
47	55	3	4,6	253,0	3 025	13,8	9	165	21,16
48	30	2	3,4	102,0	900	6,8	4	60	11,56
49	67	2	5,5	368,5	4 489	11,0	4	134	30,25
50	57	3	5,9	336,3	3 249	17,7	9	171	34,81
Итого	2800	150	250	14 752	171 536	781,4	476	8 672	1 325,96

$50a_0 + 2\,800a_1 + 150a_2 = 250;$

$2800a_0 + 171\,536a_1 - 8\,672a_2 = 14\,752;$

$150a_0 + 8\,671a_1 + 476a_2 = 781,4.$

Совместным решением этих уравнений определяются параметры искомого уравнения: $a_0 = 0,5$; $a_1 = 0,036$ и $a_2 = 0,835$. Подставив значение найденных параметров в уравнение, получим

$$\bar{y}_{xv} = 0,5 + 0,036x + 0,835v.$$

Оно означает, что не имеющий квалификации рабочий при ручной формовке бетона вырабатывает за смену в среднем 0,5 $м^3$, возрастание уровня механизации на один процент вызывает рост дневной выработки в среднем на 0,036 $м^3$ и повышение квалификации рабочих на один разряд вызывает рост дневной выработки в среднем на 0,835 $м^3$.

Уравнение множественной корреляции дает более точные значения параметров, чем уравнения парной корреляции.

В частности, если в уравнении парной корреляции параметр a_1 показывает возрастание дневной выработки при повышении уровня механизации труда при прочих равных или средних условиях, то в уравнении множественной корреляции этот параметр показывает возрастание дневной выработки в связи с повышением уровня механизации труда при прочих равных условиях и постоянном значении уровня квалификации рабочих. Точно так же если в уравнении парной корреляции параметр a_2 показывает возрастание дневной выработки при повышении квалификации рабочих при прочих равных средних условиях, то в уравнении множественной корреляции этот параметр показывает возрастание дневной выработки в связи с повышением квалификации рабочих при прочих равных условиях и постоянном значении уровня механизации труда. Иными словами, уравнение множественной корреляции позволяет элиминировать не только влияние прочих факторов, которые не являются предметом исследования, но и влияние факторов, корреляционно связанных с y.

Оценка силы (тесноты) корреляционной связи

При статистическом изучении взаимосвязей между признаками общественных явлений важное значение имеет оценка тесноты связи, или, иными словами, оценка степени близости корреляционной связи к функциональной. Эта задача решается путем построения и исчисления коэффициентов корреляции.

Коэффициенты множественной корреляции

При линейной форме множественной корреляции теснота связи между результативным (y) и двумя факторными $(x$ и $v)$ признаками измеряется с помощью показателя совокупного коэффициента корреляции:

$$R_{yxv}=\sqrt{\frac{r_{yx}^2+r_{yv}^2-2r_{yx}\times r_{yv}\times r_{xv}}{1-r_{xv}^2}},$$

где r_{yx}, r_{yv} и r_{xv} – линейные коэффициенты парной корреляции между соответствующими признаками.

Совокупный коэффициент корреляции показывает, какую часть ***общей*** колеблемости, у составляют колебания, вызванные исследуемыми факторами x и у.

Этот показатель характеризует степень связи результативного признака с двумя факторами в их совокупности и колеблется между 0 и 1. Когда он равен единице, то y связан с x и v точной линейной связью.

Когда же совокупный коэффициент корреляции равен 0, то **у** не может быть линейно связан с x и v, хотя возможна нелинейная корреляционная и даже функциональная связь. Следовательно, совокупный коэффициент корреляции R_{yxv} характеризует тесноту линейной связи y с x и v.

Определим, например, совокупный коэффициент множественной корреляции, характеризующий тесноту линейной связи между величиной дневной выработки бетонщиков *(у)* и уровнем механизации труда *(х)* и квалификацией рабочих *(у)* по данным табл. 8.10.

Для этого рассчитаем сначала средние значения для каждого признака и их сочетаний, а также средние квадратические отклонения:

$$\bar{x} = \frac{2800}{50} = 56; \ \bar{v} = \frac{150}{50} = 3; \ \bar{y} = \frac{250}{50} = 5;$$

$$\overline{xv} = \frac{8672}{50} = 173,44; \ \overline{xv} = \frac{14752}{50} = 295,04;$$

$$\overline{\upsilon y} = \frac{781,4}{50} = 15,63; \ \overline{x^2} = \frac{171536}{50} = 3430,72;$$

$$\overline{\upsilon^2} = \frac{476}{50} = 9,52; \ \overline{y^2} = \frac{1325,96}{50} = 26,52;$$

$$\overline{\sigma_x} = \sqrt[2]{\overline{x^2} - \left(\bar{x}\right)^2} = \sqrt{3430,72 - 56^2} = \sqrt{3430,72 - 3136} = \sqrt{294,72} = 17,17;$$

$$\sigma_\upsilon = \sqrt{\overline{\upsilon^2} - \left(\bar{\upsilon}\right)^2} = \sqrt{9,52 - 9} = \sqrt{0,52} = 0,72;$$

$$\sigma_y = \sqrt{\overline{y^2} - \left(\bar{y}\right)^2} = \sqrt{26,52 - 25} = \sqrt{1,52} = 1,23;$$

Далее найдем линейные коэффициенты корреляции между двумя признаками. Линейный коэффициент корреляции между величиной дневной выработки на формовке бетона (**у**) и уровнем механизации труда бетонщиков *(х)*:

$$r_{yx} = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\sigma_x \cdot \sigma_y} = \frac{295,04 - 56 \times 5}{17,17 \times 1,23} = \frac{15,04}{21,12} = 0,712.$$

Линейный коэффициент корреляции между величиной дневной выработки и квалификацией бетонщиков (v) составляет:

$$r_{y\upsilon}=\frac{\overline{\upsilon y}-\overline{\upsilon}\cdot\overline{y}}{\sigma_{\upsilon}\cdot\sigma_{y}}=\frac{15{,}63-3\times5}{0{,}72\times1{,}23}=0{,}708.$$

Наконец, линейный коэффициент корреляции между факторными признаками **x** и **v** равен:

$$r_{x\upsilon}=\frac{\overline{x\upsilon}-\overline{x}\cdot\overline{y}}{\sigma_{x}\cdot\sigma_{\upsilon}}=\frac{173{,}44-56\times3}{17{,}17\times0{,}72}=0{,}44.$$

На основе произведенных расчетов определим совокупный линейный коэффициент множественной корреляции, измеряющий силу связи, между величиной дневной выработки бетонщиков, уровнем механизации труда на формовке и одновременно квалификацией рабочих:

$$R_{yx\upsilon}=\sqrt{\frac{0{,}712^2+0{,}708^2-2\times0{,}712\times0{,}708\times0{,}44}{1-0{,}44^2}}=0{,}837.$$

Это свидетельствует о сильной степени линейной связи между рассматриваемыми, признаками.

Наряду с совокупным коэффициентом множественной корреляции и в дополнение к нему при линейной множественной корреляции определяются частные коэффициенты корреляции между результативным признаком и каждым из факторных признаков при исключении влияния второго факторного признака. Эти показатели дают оценку тесноты связи между результативным признаком и одним из факторных признаков при элиминировании влияния другого факторного признака. Частный коэффициент корреляции между **y** и фактором **x** при элиминировании фактора **v** показывает, какую часть колеблемость **y**, вызванная фактором **x**, составляет в колеблемости **y**, вызванной всеми факторами, кроме ***v***. Так же как и парные коэффициенты корреляции, они могут принимать значения от -1 до $+1$.

Значение частного коэффициента корреляции, равное нулю, соответствует отсутствию линейной связи между результативным признаком и одним из факторных, хотя возможна нелинейная корреляционная и даже функциональная связь между ними. Значение частного коэффициента корреляции, равное единице, соответствует линейной функциональной связи между результативным и факторным признаками.

В рассматриваемом примере наряду с совокупным коэффициентом корреляции возможно исчисление двух частных коэффициентов корреляции: 1) частного коэффициента корреляции между величиной дневной выработки и уровнем механизации труда при исключении влияния квалификации рабочих и 2) частного коэффициента корреляции между величиной дневной выработки и квалификацией рабочих при исключении влияния уровня механизации труда.

Частные коэффициенты множественной корреляции при двух факторных признаках, влияние одного из которых исключается, определяются по следующим формулам:

$$\left.\begin{aligned} r_{yx}(\upsilon) &= \frac{r_{xy} - r_{x\upsilon} \times r_{\upsilon y}}{\sqrt{\left(1 - r_{x\upsilon}^2\right)\left(1 - r_{\upsilon y}^2\right)}}; \\ r_{y\upsilon}(x) &= \frac{r_{\upsilon y} - r_{x\upsilon} \times r_{xy}}{\sqrt{\left(1 - r_{x\upsilon}^2\right)\left(1 - r_{xy}^2\right)}}; \end{aligned}\right\}$$

где $r_{yv(x)}$ измеряет тесноту связи между **у** *и* **х** при исключении влияния **v**, а $r_{yv\,(x)}$ – между **у** и **v** при исключении влияния **х** (корень в этих формулах берется со знаком плюс).

Так, для рассматриваемого примера частный коэффициент линейной множественной корреляции между величиной дневной выработки и уровнем механизации труда при исключении влияния квалификации рабочих составляет:

$$r_{yx(\upsilon)} = \frac{0{,}712 - 0{,}44 \times 0{,}708}{\sqrt{\left(1 - 0{,}44^2\right)\left(1 - 0{,}708^2\right)}} = 0{,}635.$$

Частный коэффициент линейной множественной корреляции между величиной дневной выработки и квалификацией рабочих при исключении влияния уровня механизации труда на формовке бетона составляет:

$$r_{y\upsilon(x)} = \frac{0{,}708 - 0{,}44 \times 0{,}712}{\sqrt{\left(1 - 0{,}44^2\right)\left(1 - 0{,}712^2\right)}} = 0{,}625.$$

Если линейный коэффициент корреляции между величиной дневной выработки (у) и уровнем механизации труда на формовке бетона (x) равен 0,712, то частный коэффициент линейной множественной корреляции между этими показателями только 0,635; если линейный коэффициент корреляции между величиной дневной выработки и квалификацией рабочих (v) составляет 0,708, то соответствующий частный коэффициент линейной множественной корреляции только 0,625. Меньшая величина частного коэффициента линейной множественной корреляции по сравнению с линейным коэффициентом корреляции – результат исключения влияния второго факторного признака.

8.7. Измерение связей неколичественных переменных

Коэффициент ассоциации

Показателем силы связи двух явлений, выраженных качественными признаками, каждый из которых в свою очередь является альтернативным, служит **коэффициент ассоциации.**

Он выводится из четырехклеточной аналитической таблицы. Например, обследуют группу населения одного из регионов РУз в отчетном периоде.

I вопрос – о месте проживания (следует выбрать правильный ответ):

1. Проживаю в городе (*a*).
2. Проживаю в сельской местности (b).

II вопрос – о принадлежности к полу (следует выбрать правильный ответ):

1. Мужчина (c).
2. Женщина (d).

Представив суммарную численность ответов на каждый вопрос буквенными символами (*a*, *b*, **c**, d), покажем, как можно построить комбинационную четырехклеточную ***таблицу*** которая может помочь нам в дальнейших расчетах (табл. 8.11)

Таблица 8.11

Взаимосвязь между ответами на два вопроса

Ответы на I вопрос	Ответы на II вопрос		Итого
	мужчина B	женщина $\overline{B}$	
Проживают в городе A	a	b	a + b
Проживают сельской местности $\overline{A}$	c	d	c + d
Итого	a + c	b + d	a+b+c+d

Здесь A и B – признаки, связь между которыми анализируется; $\overline{A}$ и $\overline{B}$ обозначают противоположные признаки; a, b, c и d – частоты соответствующих комбинаций признаков.

На основании данных табл. 8.6 можно построить коэффициенты контингенции и ассоциации, предложенные английскими статистиками К. Пирсоном и Д. Юлом, соответственно. Коэффициент контингенции вычисляется по следующей формуле:

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

Коэффициент ассоциации

Коэффициент ассоциации рассчитывается по формуле

$$K_{асс} = \frac{ad - bc}{ad + bc}.$$

Свойства этих коэффициентов такие же, как линейного коэффициента корреляции.

Покажем расчет этих коэффициентов на примере аналитической табл. 8.12, характеризующей связь между уровнем образования и процентом выполнения нормы выработки.

Таблица 8.12

Аналитическая таблица

Группа рабочих по уровню образования	Выполнение нормы	Не выполнение нормы	Всего
Среднее специальное	78 (a)	22(b)	100 (a+b)
Среднее	32(c)	68(d)	100 (c+d)
Итого	110(a+c)	90(b+d)	200

$$K_{асс} = \frac{76 \cdot 68 - 32 \cdot 22}{78 \cdot 68 - 32 \cdot 22} = \frac{4600}{6008} = 0{,}766;$$

$$K_{конт} = \frac{76 \cdot 68 - 32 \cdot 22}{(78+22)(22+68)(78+32)(32+68)} = \frac{5304 - 704}{\sqrt{99000000}} = 0{,}46$$

Если $К_{асс} \geq 0{,}5$ и $К_{конт}$, наличие связи подтверждается.

Если ad > bc ($К_{асс} = 0$), то связь прямолинейная.

Если ad < bc, то это значит связь криволинейная.

Если ad = bc фв ($К_{асс} = 0$), то связь отсуствует.

Коэффициент взаимной сопряженности

Показателем силы связи двух атрибутивных признаков, каждый из которых имеет более двух вариантов, служит коэффициент взаимной сопряженности. Это одна из мер тесноты связи для качественных признаков. Предложены разные формулы расчета КВС англичанином К.Пирсоном и русским статистиком А.И.Чупровым

КВС Пирсона находится по формуле:

$$KBC = \sqrt{\frac{\varphi^2}{1 - \varphi^2}},$$

где φ^2 – показатель средней квадратической сопряженности. Эта величина получается путем вычитания единицы из суммы отношений квадратов частот каждой клетки корреляционной таблицы к произведению итоговых частот соответствующего столбца и строки.

Таблица 8.13

Распределение фермерских хозяйств по степени удобрения посевов и урожайности

Степень удобрения посевов	Всего хозяйств	В том числе получивших урожайность		
		низкую	среднюю	высокую
Слабая	20	8	1ы1	1
Средняя	60	11	43	6
Хорошая	20	1	6	13
Итого	100	20	60	20

КВС по формуле А.Н.Чупрова

$$K = \frac{\varphi^2}{\sqrt{(K_1 - 1)(K_2 - 1)}},$$

где К1 – число групп по столбцам статистической таблицы;

К2 – число групп по строкам.

Покажем расчет КВС Чупрова на примере аналитической таблицы, характеризующей связь между размерами внесенного удобрения и величиной урожайности хлопка сырца (данные условные).

Для получения величины φ^2 составим вспомогательную табл. (8.14).

Таблица заполняется следующим образом. Сначала в верхнем углу каждой клетки записываются частоты. Затем в середине каждой клетки, кроме графы «Итого», указываются квадраты частот.

Таблица 8.14

Расчет показателей средней квадратической сопряженности

Степень удобрения \ Урожайность	Низкая	Средняя	Высокая	Итого	
Слабая	8 64 3,2	11 121 2,02	1 1 0,05	20 – 5,57	0,264
Средняя	11 121 6,05	43 1 849 30,82	6 36 1,68	60 – 38,67	0,644
Хорошая	1 1 0,05	6 36 0,6	13 169 8,45	20 – 9,10	0,455
Итого	20	60	20	100	1,363

В третьей строке каждой клетки проставляются частные от деления квадратов частот на сумму частот по соответствующему столбцу: 3,2 как результат деления 64 на 20; 6,05 как результат деления 121 на 20 и т.д. Сумма этих чисел заносится в нижнюю часть клеток графы «Итого»: 3,2 +2,02+ 0,05 = 5,27 – в первой клетке и т.д. Наконец, частное от деления этих сумм на итог частот по каждой строке проставляется в последнем столбце таблицы: 5,27 : 20 = 0,264 по первой клетке; такой же расчет применяется и по второй и третьей клеткам. Сумма чисел последнего столбца без единицы и равна φ^2. Таким образом, $\varphi^2 = 1{,}363 - 1{,}0 = 0{,}363$.

Подставляя найденное значение в формулу находим искомый коэффициент:

$$KBC = \sqrt{\frac{\varphi^2}{1-\varphi^2}} = \sqrt{\frac{0{,}363}{1{,}363}} = 0{,}51.$$

Коэффициент взаимной сопряженности характеризует наличие сопряженности двух данных признаков. Величина его варьирует от нуля до единицы. Чем больше сопряженность изучаемых признаков, тем его значение ближе к единице.

Из произведенного расчета видно, что величина коэффициента взаимной сопряженности зависит исключительно от величин частот рассматриваемых признаков и не связана со значениями самих признаков. Такой характер данного коэффициента делает его удобным при изучении связи между атрибутивными признаками.

Интеллектуальный тренинг

1. В чем сущность корреляционной связи между признаками?
2. Назовите основные методы выявления корреляционной связи.
3. В чем состоит основное содержание корреляционно-регрессионного анализа?
4. Как определяются параметры уравнения парной регрессии при линейной зависимости?
5. Что показывают линейный коэффициент корреляции и линейный коэффициент детерминации и как они рассчитываются?
6. Какие показатели используются для оценки силы корреляционной связи при криволинейной зависимости и как они рассчитываются?
7. В чем сущность метода оценки значимости корреляционной связи с использованием критерия Фишера?

8. С какой целью и каким образом рассчитывают ошибку аппроксимации?

9. Какие вы знаете формы статистической связи?

10. Что такое уравнение регрессии?

11. Каковы предельные значения корреляционного отношения?

12. Что такое множественная корреляция?

13. Даст ли количественную оценку качества модели средняя ошибка аппроксимации?

14. Коэффициент корреляции r_{xy} применяется в тех случаях, когда между явлениями существует прямолинейная связь. Так ли это? Если нет, то докажите.

15. Что Вы понимаете под детерминированной зависимостью?

16. Помимо факторного признака (К) на результативный признак влияют и другие факторы, в том числе и неучтенные. Как они называются и обозначаются?

17. Каковы особенности измерения связей неколичественных переменных?

18. Как определяется коэффициент взаимной сопряженности А.И.Чупрова?

19. Как определяется коэффициент контингенции К.Пирсона? Что он характеризует?

20. Покажите порядок расчета коэффициента ассоциации?

21. Можно ли осуществлять прогноз на основе парной регрессионной модели? Если да, то как?

22. Когда признаются типичными полученные при анализе корреляционные связи параметры уравнения регрессии?

23. Для чего используется t-критерий Стьюдента?

24. Что означает $F_{расч.}$ и $F_{табл.}$?

25. В чем состоит отличие между корреляционной и функциональной связью?

26. Какие основные проблемы решает исследователь при изучении корреляционных зависимостей?

27. Какова роль групповых и корреляционных таблиц при анализе взаимосвязей?

28. Какие методы целесообразно использовать для выявления возможного наличия связи между факторным и результативным признаком при небольшом объеме фактических данных?

29. Какие показатели являются мерой тесноты связи между двумя признаками?

30. Как сцепить существенность линейного коэффициента корреляции?

31. Какие показатели используют для измерения степени тесноты связи между качественными признаками?

32. В чем состоит значение уравнения регрессии?

33. Что характеризует коэффициент регрессии?

34. Какая существует связь между линейным коэффициентом корреляции и коэффициентом регрессии?

35. Какое значение имеет расчет индекса корреляции?

36. Способы расчета средней квадратической ошибки уравнения, ее роль в оценке надежности уравнения регрессии.

37. Как осуществить прогноз значений результативного признака, опираясь на использование для этой цели уравнения регрессии?

38. Как измерить долю общей вариации результативного признака, которая объясняется влиянием вариации признака фактора x?

39. Если в случае линейной зависимости между признаками 60% вариации результативного признака объясняется влиянием факторного признака, чему будет равна величина коэффициента корреляции?

40. Как подходить к отбору факторов для включения их в уравнение множественной регрессии?

41. Каким образом можно выделить факторы, в изменении которых заложены наибольшие возможности в управлении изменением результативного признака?

42. Что означает величина коэффициента эластичности 0,58%.

43. Для чего рассчитываются частные коэффициенты корреляции?

44. В каких пределах заключена величина совокупного коэффициента корреляции и как она соотносится с величиной парных коэффициентов корреляции?

Использованная и рекомендуемая специальная литература

1. Бородкин Ф.М. Статистическая оценка связей экономических показателей. М.: Статистика, 1968.

2. Вайну Я.Я. Корреляция рядов динамики. М.: Статистика, 1977.

3. Дюран Б., Одел П. Кластерный анализ. М.: Статистика, 1977.

4. Езикиел М., Факс К. Методы анализа корреляции и регрессии. М.: Статистика, 1966.

5. Елисеева И.И. Статистические методы измерения связей. Л.: Ленинградский университет, 1982.

6. Кендэл М. Ранговые корреляции. М.: Статистика, 1975.

7. Ковалева Л.Н. Многофакторное прогнозирование на основе рядов динамики. М.: Статистика, 1980.

8. Кайкин В.П., Найденов В.С., Галуза С.Г. Корреляция и статистическое моделирование в экономических расчетах. М.: Экономика, 1964.

9. Крастинов О.П. Разработка и интерпретация моделей корреляционных связей в экономике. Рига: Занятие, 1983.

10. Льюис К.Д. Методы прогнозирования экономических показателей. М.: Финансы и статистика. 1986.

11. Лукомский Я.И. Теория корреляции и ее приминение к анализу производства. М.: Госстатиздат, 1958.

12. Окунь Д. Факторный анализ. М.: Статистика, 1974.

13. Статистические методы исследования корреляции в экономике. М.: Статистика, 1972.

14. Сиськов В.И. Корреляционный анализ в экономических исследованиях. М.: Статистика, 1975.

15. Урланис Б.Ц. Статистические методы изучения зависимости явлений. М.: Госстатиздат, 1959.

16. Чупров А.А. Основные проблемы теории корреляции. М.: Госстатиздат, 1960.

17. Четвериков Н.С. О ложной корреляции. В кн. Применение методов корреляции в экономических исследованиях. М.: Наука, 1969.

18. Фестер Э., Ренц Б. Методы корреляционного и регрессионного анализа. М.: Финансы и статистика, 1983.

Глава IX
СТАТИСТИЧЕСКОЕ ИЗУЧЕНИЕ ДИНАМИКИ

Дорожная карта

9.1. Понятие ряда динамики

Одно из основных положений научной методологии – необходимость изучать все явления в развитии и во времени. Это относится и к статистике: она должна дать характеристику изменений статистических показателей во времени. Как изменяются год за годом валовой национальный продукт и национальный доход страны? Как возрастает или снижается уровень оплаты труда? Велики ли колебания урожайности зерновых культур и существует ли тенденция ее роста? На все аналогичные вопросы ответ может дать только специальная система статистических методов, предназначенная для изучения развития, изменений во времени или, как принято в статистике говорить, изучения динамики.

Основная цель статистического изучения динамики общественных явлений состоит в выявлении и измерении закономерностей их развития во времени. Это достигается посредством построения и анализа статистических рядов динамики.

Рядами динамики называются статистические данные, отображающие изменение изучаемого явления во времени.

В каждом ряду динамики имеются два основных элемента:

1) показатель времени ***t***;

2) соответствующие им уровни изменения изучаемого явления *y*.

В качестве показаний времени в рядах динамики выступают либо определенные даты (моменты) времени, либо отдельные периоды (годы, кварталы, месяцы, сутки).

Таблица 9.1

Социально-экономические показатели по Республике Узбекистан за 2014-2018 гг.

Показатели	2014	2015	2016	2017	2018
Численность постоянного населения на конец года (млн чел)	31,0	31,5	32,1	32,7	33,3
Среднегодовая численность занятых в экономике (млн чел)	12,8	13,1	13,3	13,5	13,3
То же в процентах к предыдущему периоду (%)	102,4	102,3	101,5	101,5	101,6
Удельный вес городского населения в общей численности населения страны (%)	50,8	50,6	50,6	50,6	50,7
Выпуск промышленной продукции (трлн. сум)	84,0	97,6	111,9	148,8	228,9

Примечание. *По данным статистического ежегодника Госкомстата РУз. Т., 2018. С. 33-35.*

Уровни рядов динамики отображают количественную оценку (меру) развития во времени изучаемого явления. Они могут выражаться абсолютными, относительными или средними величинами, а также в графическом виде (рис. 9.1 и 9.2). Примером такого представления уровней ряда динамики являются данные об изменении отдельных социально-экономических показателей по Республике Узбекистан (табл. 9.1).

В приведенных рядах численность населения выражена в абсолютных величинах (в млн чел), численность занятых в экономике выражена средними величинами, в удельный вес городского населения населения – относительными величинами структуры (в %).

Рис. 9.1. **Сочетание различных составляющих ряда динамики, определяющих случайные колебания и основную тенденцию**

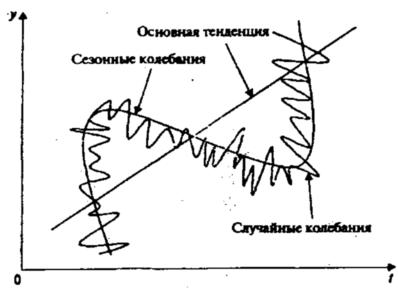


Рис. 9.2. **Сочетание различных составляющих ряда динамики, определяющих случайные, сезонные колебания и основную тенденцию**

На изменение уровней ряда влияют различные факторы:

• постоянно действующие факторы, которые определяют основную тенденцию развития явления;

• случайные факторы, влияние которых отразится на колебании отдельных уровней ряда, но на общей тенденции, скорее всего, нет;

• периодически действующие факторы, которые определяют сезонные колебания.

Влияние различных факторов на изменение уровней ряда динамики показано на рис. 9.1 и 9.2.

В статистической теории приняты следующие *основные общепринятые обозначения уровней рядов динамики:*

y_1 – начальный уровень;

y_n – конечный уровень;

y_0 – базисный уровень;

y_i – данный (текущий) уровень;

y_{i-1} – предыдущий уровень;

y_j – уровень одного из промежутков времени, общее число которых *к;*

$\bar{y}$ – средний уровень.

Всего в ряду динамики *п* уровней.

9.2. Виды рядов динамики

Данные, приведенные в табл. 9.1, могут быть использованы для иллюстрации рядов динамики разного типа. Различают следующие виды рядов динамики (рис. 9.3).

Виды рядов динамики

• моментные и интервальные
• ряды с равно- и неравноотстоящими во времени уровнями
• стационарные и нестационарные
• с нарастающим итогом
• с абсолютными величинами
• с относительными величинами
• со средними величинами

Рис. 9.3. **Классификация видов рядов динамики**

Моментные РД *Моментным* является ряд динамики, уровни которого характеризуют изучаемое явление в конкретный момент времени (табл. 9.1, строка 1). Такие ряды используют для описания величин типа запаса (остатки средств на счетах клиентов, объем собственных средств (капитала), объем основных фондов и т.д.).

Интервальные РД Следует помнить, что *моментные ряды нельзя суммировать.* Бессмысленно, например, складывать численность населения РУз по стоянию на конец 2014, 2015, 2016 годов (табл. 9.1). Полученная сумма ничего не выражает, так как в ней многократно повторяются одни ***и те же численности населения. Каждый последующий уровень ряда включает основную часть предыдущего уровня.***

Интервальные ряды динамики отображают итоги развития (функционирования) изучаемых явлений за отдельные периоды (интервалы) времени (табл. 9.1, строка 5). Их можно *суммировать* для получения новых числовых значений за более длительный период времени. Например, в нашей стране за 2014-2018 гг. выпущена промышленная продукция (в ценах соответствующих лет) более чем 670 трлн. сум.

Таким образом, интервальные ряды динамики используют для описания величин типа экономического потока, операции (проценты полученные, проценты уплаченные, комиссионные доходы и расходы, выпуск продукции, текущие затраты и т.п.). Уровни интервальных рядов динамики обладают свойством суммарности, показатели моментных рядов такого свойства не имеют. Можно сложить показатели объема промышленной продукции за кварталы и получить итог производства за год. Но если за год сложить данные о числе рабочих на начало каждого квартала, полученная сумма не будет иметь реального смысла.

В рядах с *равностоящими* уровнями даты регистрации или окончания периодов представлены через равные, следующие друг за другом отрезки времени (табл. 9.1). В рядах с *неравностоящими* уровнями принцип равенства отрезков времени не соблюдается.

Стационарные и нестационарны РД

Ряд динамики, в изменении уровней которого не наблюдается общей направленности (тенденции), является *стационарным.* Напротив, *нестационарный* ряд отличается наличием общей направленности в изменении уровней изучаемого показателя. Так, уровни рядов динамики, которые представлены в табл. 9.1, имеют разнонаправленные тенденции изменения: объем промышленной продукции и номинальные совокупные доходы на душу населения РУз имеют тенденцию к росту[1]. А доля общей численности занятых в экономике имеет тенденцию к снижению (табл. 9.1, строка 3).

Ряды динамики с нарастыющим итогом

Статистическое отображение развития изучаемого явления во времени может быть представлено рядами динамики с нарастающими итогами. Их применение обусловлено потребностями отображения результатов развития изучаемых показателей не только за данный отчетный период, но и с учетом предшествующих периодов. При составлении таких рядов производится последовательное суммирование смежных уровней. Этим достигается суммарное обобщение результата развития изучаемого показателя с начала отчетного периода (месяца, квартала, года и т.д.).

Ряды динамики с нарастающими итогами строятся при определении общего объема товарооборота в розничной торговле. Так, объем продажи товаров в «Корзинке» определяется каждый месяц обобщением товарно-денежных отчетов за отдельные операционные периоды (пятидневки, недели, декады и т.д.).

В качестве примера воспользуемся следующими данными о ходе реализации товаров в «Корзинке» за октябрь 2018 г. (табл. 9.2).

[1] Статистический ежегодник Госкомстата РУз. Т., 2018. С. 17.

Данные гр. 3 табл. 9.2 отображают обобщенные с начала месяца, результаты продажи товаров по отдельным периодам месячного цикла работы «Корзинки».

РД абсолютных величин

Рядом динамики абсолютных величин называют такой ряд, члены которого выражают абсолютные значения изучаемого показателя за ряд последующих моментов или отрезков времени.

Таблица 9.2

Реализация товаров в магазине «Корзинка» за октябрь 2018 г.

(млн сум)

Пятидневки	Розничная реализация товаров за октябрь	
	за пятидневку	с начала месяца
1	2	3
Первая	5,2	5,2
Вторая	4,3	5,2+4,3=9,5
Третья	12,4	9,5+12,4=21,9
Четвертая	18,0	21,9+18,0=39,9
Пятая	21,9	39,9+21,9=61,8
Шестая	18,4	61,8+18,4=80,2

Оба приведенных выше ряда (табл. 9.1 строки 1 и 5), являются рядами абсолютных величин. Они выражают в абсолютных величинах: первый – динамику численности постоянного населения Республики Узбекистан (млн чел), пятый – динамику выпуска промышленной продукции за последние пять лет в триллионах сум.

РД относительных величин

Динамическим рядом относительных величин называют такой ряд, члены которого выражают относительные размеры изучаемого показателя за ряд последовательных промежутков или моментов времени. Рядом динамики относительных величин являются, например, данные об удельном весе городского населения в общей численности населения нашей страны в процентах (табл. 9.1, строка 4).

РД средних величин Рядом средних величин называется такой ряд динамики, члены которого выражают средний уровень изучаемого показателя за отдельные промежутки или моменты времени. Примером такого вида рядов динамики могут служить приводимые выше данные о динамике среднегодовой численности занятых в экономике нашей страны, в млн чел (табл. 9.1, строка 2).

Данная выше классификация рядов динамики в известной мере, условна. Однако она помогает выяснить некоторые характерные особенности рядов динамики, влияющие на порядок построения рядов, их обработку и анализ.

9.3. Сопоставимость в рядах динамики

Основным условием для получения правильных выводов при анализе рядов динамики является сопоставимость его элементов.

Ряды динамики формируются в результате сводки и обработки материалов периодического наблюдения. Повторяющиеся во времени (по отчетным периодам) значения одноименных показателей в ходе статистической сводки систематизируются в хронологической последовательности.

При этом каждый ряд динамики охватывает отдельные обособленные периоды, в которых могут происходить изменения, приводящие к несопоставимости отчетных данных с данными других периодов. Поэтому для анализа ряда динамики необходимо приведение всех составляющих его элементов к сопоставимому виду.

Для этого в соответствии с задачами исследования устанавливаются причины, обусловившие несопоставимость анализируемой информации, и применяется соответствующая обработка, позволяющая производить сравнение уровней ряда динамики. Несопоставимость в рядах динамики вызывается различными причинами, Уровни рядов динамики должны быть сопоставимы:

- по территории;
- по способу исчисления;
- по единице измерения;
- по периоду времени и т.п.

Несопоставимость по кругу охватываемых объектов возникает, например, в силу особенности учета изучаемых единиц совокупности, в результате изменения административно-территориальных границ, сравниваемых объектов (областей, административных районов, предприятий и т.п.).

Подобные изменения неизбежно изменяют границы исследуемого объекта, изменяют число единиц, подвергаемых наблюдению до и после территориальных изменений, а потому и величину характеризующих их показателей. Так, например, если изучается развитие хозяйства и культуры Самаркандской и Бухарской областей за ряд лет, то, естественно, данные этих областей должны относиться к одинаковой территории. Известно, что в 1982 г. произошло изменение границ этих областей, в результате которого образовалась новая область – Навоийская, что данные, характеризующие развитие Бухарской и Самаркандской областей за годы, предшествовавшие 1982 и после 1982, будут несравнимыми вследствие неодинаковости территорий. Поэтому в нынешнем стат ежегоднике территории всех областей Республики Узбекистан приведены к сопоставимому виду за все анализируемые годы (1970–2018). Показатели по Самаркандской и Сырдарьинской областям приведены в соответствующих границах с учетом отошедших районов в Джизакскую область (1972), а в 1982 г. в связи с образованием Навоийской области вновь пересчитаны все показатели Бухарской и Сырдарьинской областей. Таким образом, обеспечена административно-территориальная сопоставимость границ областей республики.

Пример: В 2017 г. произошло укрупнение обслуживаемого торговой организацией региона, результаты которого отображены в следующих изменениях объемов товарооборота (млн сум).

Несопоставимые ряды	2016	2017	2018
В прежних границах	432	450	–
В новых границах	–	630	622,5

Для приведения этой информации к сопоставимому виду производится так называемое смыкание рядов динамики. При этом для 2017 г. определяется коэффициент соотношения двух уровней: 630:450 = 1,4. Умножая на этот коэффициент объем товарооборота 2016 г. (432·1,4 = 604,8 млн сум), можно построить ряд динамики сопоставимых уровней в новых границах региона 9 млн сум.

	2016	2017	2018
Сомкнутый ряд	604,8	630	622,5

Смыкание рядов

Смыкание рядов используют для сопоставления двух рядов показателей, характеризующих динамику одного и того же явления в новых и старых административных границах. Если имеются данные в новых и старых границах по одному и тому же кругу объектов, то такие динамические ряды можно сомкнуть.

Общность способов исчисления

Иногда уровни научаемого явления бывают исчислены различными способами. Например, надо изучить динамику численности населения РУз за длительный период времени. Однако имеющиеся данные о численности населения исчислены разными способами. Так, за один год численность определена на начало каждого года, то есть на определенную дату, а за другие – как среднегодовая. В таком виде эти данные несопоставимы между собой. Для обеспечения их сопоставимости необходимо численность населения начала года пересчитать в среднегодовую численность.

Общность единиц измерения

Нередко одни и те же данные выражаются: в различных единицах измерения. Например, валовой надой молока может быть выражен в литрах и в килограммах. Чтобы обеспечить сравнимость таких данных, необходимо выразить их в одних и тех же единицах измерения.

Проблема сопоставимости в рядах динамики возникает в связи с применением в статистической информации различных по экономическому значению денежных измерителей. Так, для денежной оценки объема поставки (оптовой продажи) товаров применяются оптовые цены промышленности, а для оценки объема продаж товаров населению применяются розничные цены. К разновидностям розничных цен относятся кооперативные и договорные цены, цены базарной торговли, закупочные и сдаточные цены на сельскохозяйственную продукцию и др.

Поскольку уровни цен изменяются во времени, то для стоимостной оценки товарооборота используются цены соответствующих периодов. Но для изучения динамики физического объема продаж товаров денежная оценка товарооборота в ценах соответствующих периодов не подходит. На объем товарооборота влияет не только фактор реализованной товарной массы, но и фактор изменения цен. Для устранения влияния изменения цен товарооборот пересчитывается в неизменные (базисные) цены. В результате получают ряды динамики объема товарооборота в сопоставимых ценах.

Сопоставимость периодов времени

Уровни динамического ряда должны быть также сравнимы во времени. Так, например, производство молока характеризуется значительными сезонными колебаниями. Поэтому было бы неправильно сравнивать между собой данные о производстве молока в разные годы не за одни и те же месяцы или кварталы. Такого рода данные в силу своей несравнимости по условиям сезонности не позволяют выявить закономерности развития явления. Чтобы установить, растет или уменьшается производство молока в хозяйстве или группе хозяйств, необходимо сравнить между собой надой молока за одни и те же месяцы или кварталы, например, надои за январь 2018 г. с надоем за январь 2017 г. или январь 2016 г., данные за июнь 2018 г. с данными за июнь 2017 г. или за июнь 2016 г. и т.п.

Вопрос о сопоставимости встает и при анализе уровней разных динамических рядов. Так, для совместного анализа уровней моментного и интервального рядов динамики моментные динамические ряды должны быть преобразованы таким образом, чтобы пересчитанные уровни охватывали те же промежутки времени, что и уровни интервального динамического ряда. Трудности сравнения взаимосвязанных рядов динамики возникают из-за наличия, так называемого, *временного лага*, который является мерой отставания во времени изменений одних явлений по сравнению с другими. Например, для расчета доли закончивших обучение студентов показатели года выпуска студентов следует сравнивать с показателями года приема этих же студентов, т.е. смещая данные с учетом срока обучения. Вопрос о сопоставимости встает и при анализе уровней, когда ряды характеризуют динамику различных явлений, непосредственно несопоставимых между собой (например, ВВП различных стран в разных денежных единицах, производство электроэнергии и хлопка сырца и т.п.).

В этих случаях приведение рядов динамики к одному основанию и сопоставление темпов роста и прироста является единственно возможным способом сравнения и анализа. В этих случаях за базу сравнения для всех уровней принимается начальный уровень ряда и все последующие уровни сопоставляются этим уровнем приняв за 100.

9.4. Основные показатели анализа динамических рядов

При изучении динамики необходимо решить целый ряд задач и осветить широкий круг вопросов с тем, чтобы охарактеризовать особенности и закономерности развития изучаемого объекта. К числу основных задач, возникающих при изучении динамических рядов, относятся следующие:

1) характеристика интенсивности отдельных изменений в уровнях ряда от периода к периоду или от даты к дате;

2) определение средних показателей временного ряда за тот или иной период;

3) выявление и количественная оценка основной тенденции развития (тренда);

4) изучение периодических колебаний;

5) экстраполяция и прогнозирование.

Динамический ряд представляет собой ряд последовательных уровней, сопоставляя которые между собой можно получить характеристику скорости и интенсивности развития явления. В результате сравнения уровней получается система абсолютных и относительных показателей динамики, к числу которых относятся абсолютный прирост, коэффициент роста, темп прироста, абсолютное значение одного процента прироста и пункты роста. Если цели сравнения подлежат несколько последовательных уровней, то возможны два варианта сопоставления:

1. Каждый уровень динамического ряда сравнивается с одним и тем же предшествующим уровнем, принятым за базу сравнения.

В качестве базисного уровня выбирается либо начальный уровень динамического ряда или же уровень, с которого начинается какой-то новый этап развития явления. Такое сравнение называется сравнением с постоянной базой.

2. Каждый уровень динамического ряда сравнивается с непосредственно ему предшествующим, такое сравнение называют сравнением с переменной базой.

Показатели динамики с постоянной базой (базисные показатели) характеризуют окончательный результат всех изменений в уровнях ряда от периода, к которому относится базисный уровень, до данного (i-того) периода. **Показатели динамики с переменной базой (цепные показатели)** характеризуют интенсивность изменения уровня от периода к периоду (или от даты к дате) в пределах изучаемого промежутка времени (рис. 9.4).

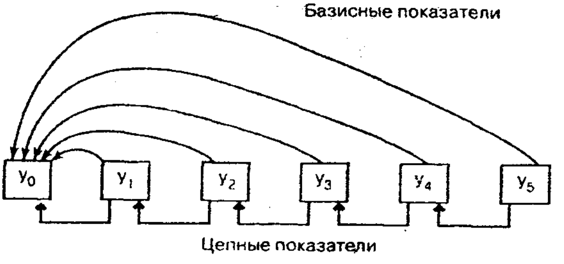


Рис. 9.4. **Принципы построения цепных и базисных показателей динамики**

В анализе рядов динамики обычно используются следующие показатели динамики, характеризующие изменения отдельных уровней ряда динамики. Покажем это на примере изучения динамики численности занятых в экономике (табл. 9.3).

Абсолютный прирост

Абсолютный прирост (сокращение) – абсолютное отклонение ($\Delta_{абс}$), показывающее, на сколько в абсолютном выражении один уровень больше (или меньше) другого, более раннего:

$$\Delta_{абс.ц} = y_i - y_{i-1};$$

$$\Delta_{абс.б.} = y_i - y_0.$$

Необходимо отметить, что $\Delta_{абс}$ со знаком (+) – абсолютный прирост, а со знаком (–) – абсолютное сокращение (см. 2018 г. табл. 9.3).

Если уровни ряда относительные величины, выраженные в процентах, то на этом этапе анализа динамики слово «проценты», во избежание дальнейшей путаницы, принято заменять словом «пункты», т.е. тогда $\Delta_{абс}$ будет выражен в «пунктах».

Таблица 9.3

Среднегодовая численность занятых в экономике Республики Узбекистан за 2013–2018 гг.[1]

Годы	Среднегодовая численность занятых в экономи-ке (тыс. чел)	Абсолютный прирост (тыс. чел)		Коэффициент роста (в разах)		Темп роста, % (Тр = Кр · 100)		Тем прироста (ΔТр = Кр – 100)	
		цеп-ной $\Delta_{абс.ц} = y_i - y_{i-1}$	базис-ный $\Delta_{абс.б.} = y_i - y_0$	цепной $K_p^ц = \frac{y_i}{y_{i-1}}$	базисный $K_p^б = \frac{y_i}{y_0}$	к предыдущему периоду	к базисному периоду	к предыдущему периоду	к базисному периоду
1	2	3	4	5	6	7	8	9	10
2013	12522	–	–	–	–	–	–	–	–
2014	12818	296	296	1,0236	1,0236	102,36	102,36	2,36	2,36
2015	13058	240	536	1,0187	1,0428	101,87	104,28	1,87	4,28
2016	13298	240	776	1,0184	1,0620	101,84	106,20	1,84	6,20
2017	13520	222	998	1,0167	1,0797	101,67	107,97	1,67	7,97
2018	13273	–247	751	0,9817	1,0600	98,17	106,00	–1,83	6,00

[1] *По данным статистического ежегодника Госкомстата Республики Узбекистан. Т., 2018. С. 33-35.*

Абсолютный прирост (сокращение) с переменной базой $\Delta_{абс.ц}$ иначе называют ***скоростью роста (снижения).***

Примечание. Для цепного варианта можно определить ***абсолютное ускорение*** $\Delta_{абс.ц}$ – разность между данным абсолютным приростом (сокращением) и предшествующим

$$\Delta_{абс.ц} = \Delta_{абс.i} - \Delta_{абс.i-1}.$$

Отрицательная величина ускорения говорит о замедлении роста или об ускорении снижения уровней ряда.

Цепные и базисные абсолютные приросты связаны между собой: сумма последовательных приростов равна соответствующему базисному приросту за весь период.

Так, в табл. 9.2 сумма показателей цепных абсолютных приростов (гр. 3: $\Sigma\Delta y^{ц} = 296 + 240 + ... + (-240) = 751$ равна базисному абсолютному приросту за 2013–2018 гг. (гр. 4 : $\Delta y^{б}_{2018} = 751$).

Коэффициент роста

Для оценки эффективности изменения уровня динамического ряда используют относительные показатели динамики:

– *коэффициент роста,* выраженный в долях единицы;

– *темп роста*, выраженный в %.

Коэффициент роста K_p определяют по формулам:

• цепной $K_p^{ч} = \dfrac{y_i}{y_{y-1}}$;

• базисный $K_p^{б} = \dfrac{y_i}{y_0}$.

Взаимосвязь цепных и базисных коэффициентов роста заключается в следующем:

а) произведение цепных коэффициентов роста равно базисному коэффициенту роста за весь период. Так, в табл. 9.2 произведение цепных коэффициентов роста (графа 5:$1{,}0236 \cdot 1{,}0187 \cdot ... \cdot 0{,}9817 = 1{,}0600$) равно базисному коэффициенту роста за 2013–2018 гг. (графа 6: $K_p^{б} = 1{,}0600$).

б) частное от деления последующего базисного коэффициента роста на предыдущий равна соответствующему цепному коэффициенту роста. Например, в гр. 6 табл. 9.2, разделив базисный коэффициент роста за 2015 г. на базисный коэффициент роста за 2014 г., получим цепной коэффициент роста за 2015 г. (1,0428 : 1,023 = 1,0187).

Для большей простоты и наглядности доказательства этой взаимосвязи используем данные за три периода:

а) $\frac{y_2}{y_1} \cdot \frac{y_3}{y_2} = \frac{y_3}{y_1}$; б) $\frac{y_3}{y_1} : \frac{y_2}{y_1} = \frac{y_3}{y_2}$.

Коэффициент роста показывает, во сколько раз увеличился уровень динамического ряда по сравнению с базисным, а в случае уменьшения какую часть базисного составляет сравниваемый уровень. Темпы и коэффициенты роста отличаются только единицами измерения.

Темп изменения

Темп изменения (роста или снижения) – относительная величина динамики, выраженная в процентах

$$T_{изм.ц} = \frac{y_i}{y_{i-1}} \cdot 100; \; T_{изм.б} = \frac{y_i}{y_0} \cdot 100.$$

Темп изменения показывает, сколько процентов один уровень составляет от другого, принятого за базу сравнения (постоянную или переменную). При этом следует учитывать, что $T_{роста} > 100\%$; $T_{снижения} < 100\%$ (см. табл. 9.3, за 2018 г. гр. 5).

Темпы прироста

Темпы прироста (сокращения) так же, как и темпы роста, исчисляют по годам (цепным методом) и накопленным итогом за длительный период (базисным методом).

Существует два способа расчета этого показателя:

1) с использованием абсолютного прироста:

$$T_{пр(сокр)_ц} = \frac{\Delta_{iаб_ц}}{y_{i-1}} \cdot 100; \; T_{пр(сокр)_б} = \frac{\Delta_{iаб_ц}}{y_0} \cdot 100;$$

2) с использованием темпа роста:

$$T_{пр(сокр)} = T_{р.(сн)} - 100\%.$$

Это формула подходит как для цепного варианта, так и для базисного. При этом следует учитывать, что $Д_{отн}$ со знаком (+) – относительный прирост, а со знаком (–) – относительное сокращение.

Относительное отклонение, выраженное не в процентах, а в коэффициентах, называют ***коэффициентом прироста***; определяется он через коэффициент роста:

$$K_{пр} = K_р - 1.$$

Темп прироста показывает, на сколько процентов изменилась величина уровня динамического ряда за изучаемый период времени. Если она сокращается, то темпы прироста будут иметь знак «минус» и характеризовать относительное уменьшение уровня ряда.

Для правильной интерпретации относительных показателей динамики явлений рекомендуется рассматривать их совместно с исходными уровнями ряда.

Если уровень ряда принимает положительные и отрицательные значения (например, финансовый результат деятельности организации может быть прибылью или убытком), то темпы изменения и прироста не имеют экономической интерпретации и не рассчитываются.

Абсолютное значение 1% прироста

Для цепных показателей приросту и его темпов рассчитывают показатель абсолютного значения одного процента прироста. Он равен отношению абсолютного прироста (цепного) к темпу прироста (цепному).

$$A = \frac{\Delta_{абс}}{\Delta_{отн\%}} \quad \text{или} \quad A = \frac{\Delta_{абс}}{T_{пр(сокр)}}$$

Эти формулы в цепном варианте можно преобразовать следующим образом:

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

Аналитическое значение данного показателя состоит в том, что при возрастающей скорости (и растущем уровне) темпы роста могут иметь тенденцию к уменьшению или оставаться без изменения.

В результате абсолютное значение одного процента прироста будет расти. Так, численность занятых в экономике Республики Узбекистан за 2016–2017 годы увеличилась почти с одинаковым темпом прироста: соответственно, 1,87 и 1,84%. При этом один процент относительного прироста в 2016 г. по сравнению с *2015* г. составляет 130,6 тыс. чел, а в 2018 г. по сравнению с 2017 г. – 135,2 тыс. чел. (табл. 9.4).

Таким образом, затухающий темп прироста вовсе не означает приостановки роста: при высоких абсолютных уровнях развития изучаемого явления может значительно увеличиться его абсолютный объем даже при небольшой величине темпов. Следовательно, чтобы правильно оценить значение показателя темпа, его нужно рассматривать не изолированно, а совместно с абсолютными показателями уровня и прироста.

Таблица 9.4

Изменение численности занятых в экономике Республики Узбекистан за 2014-2018 гг.

Годы	Численность занятых в экономике (тыс. чел)	Абсолютный прирост (цепной, тыс. чел)	Темп прироста (цепной), %	Абсолютное значение одного процента прироста (тыс. чел)
2014	12818	296	–	–
2015	13058	240	2,36	128,1
2016	13292	240	1,87	130,6
2017	13520	222	1,84	132,9
2018	13273	–247	–1,67	135,2

Например, в целом за пять лет (2017–2013) прирост ВВП Республики Узбекистан составил 2,1 раза против 2,5 раза в предыдущие пять лет (2012–2008). Следовательно, темп прироста снизился с 2,5 раза до 2,1 раза. Однако за этот период значение одного процента прироста ВВП возросло (в сопоставимых ценах 1991) с 3,3 трлн. сум до 8,9 трлн. сум, т.е. 2,7 раза.

Коэффициент опережения

Для анализа интенсивности изменения во времени одного явления по сравнению с другим рассчитывают *коэффициент опережения* ($K_{оп}$). Он представляет собой отношение базисных темпов роста двух динамических рядов одинаковые отрезки времени:

$$K_{оп} = \frac{K_1}{K_2};$$

где K_1, K_2 – базисные темпы роста, соответственно, первого и второго рядов динамики.

Коэффициент опережения показывает, во сколько раз быстрее растет уровень одного ряда динамики по сравнению с уровнем другого.

При таком сопоставлении темпы должны характеризовать тенденции одного направления. Так, по данным Госкомстата Республики Узбекистан темпы роста ВВП в 2017 г. увеличились по сравнению с темпом роста ВВП в 2016 г. на 5,3% (319,3:303,2).

В ряде случаев при сравнительном анализе рядов динамики рассчитывается коэффициент опережения среднегодовых темпов прироста.

Он представляет собой отношение большего среднегодового темпа прироста к меньшему, исчисленному за тот же период

$$K_{\substack{опережения \\ (среднегодовой)}} = \frac{\overline{T}_{пр}(>)}{\overline{T}_{пр}(<)}. \quad (95)$$

Коэффициент опережения показывает, во сколько раз быстрее растет уровень одного ряда по сравнению с другим.

Так, например, за 2005–2014 гг. среднегодовые темпы прироста занятых в экономике РУз составляли: в Кашкадарьинской области – 1,3%, а в г. Ташкенте – только 0,4%. Отсюда коэффициент опережения равен:

$$K_{опережения} = \frac{1,3}{0,4} = 3,2$$

Таким образом, в 2005–2014 гг. в Кашкадарьинской области ежегодный темп прироста занятых в экономике был в 3,2 раза больше, чем в г. Ташкенте.

Коэффициент опережения может быть также использован для характеристики более быстрого роста составных частей по сравнению с целым. Так, например, в Указе Президента РУз «О пяти приоритеных направлениях стратегического развития Республики Узбекистан на 2017–2021 гг.» предусматривается опережающее развитие отраслей промышленности, обеспечивающих технический прогресс и повышение эффективности производства. Если вся промышленная продукция будет увеличиваться в среднем ежегодно на 7,0–7,4%, то, например, производство электроэнергии будет возрастать ежегодно на 9,4–10,1%, производство химических волокон – на 10,9–13,3%, химического оборудования – на 12,4–13,4%. Следовательно, ежегодный темп прироста производства электроэнергии будет больше, чем всей промышленной продукции, в 1,3 раза, производства химических волокон – в 1,4–1,5 раза, химического оборудования – в 1,6–1,8 раза.

Коэффициент ускорения

Наряду с коэффициентом опережения в процессе анализа может быть исчислен коэффициент ускорения (или замедления) среднегодовых темпов прироста. Этот коэффициент представляет собой отношение среднегодового темпа прироста за тот или иной период времени к среднегодовому темпу прироста за предыдущий период. Так, например, в 2012–2016 гг. производительность труда в промышленности РУз увеличивалась в среднем ежегодно на 4,6%, а в 2017–2021 гг. она должна возрастать ежегодно в среднем на 6%. Следовательно, в республике предусматривается ускорение среднегодовых темпов прироста производительности труда по сравнению с предыдущим пятилетием в 1,3 раза (6:4,6 = 1,3).

Отметим, что коэффициенты ускорения (в отличие от коэффициентов опережения) могут также исчисляться на базе средних темпов прироста за разные по длине периоды.

Типы изменения уровней динамики

Показатели динамики с переменной базой сравнения (цепные) используют для выявления типа изменения уровней ряда. В статистической практике в соответствии с показателями динамики различают следующие типы изменений:

➢ равномерный рост или снижение (цепные абсолютные приросты одинаковы);

➢ ускоренный рост или снижение (цепные приросты систематически увеличиваются по абсолютной величине);

➢ замедленный рост или снижение (цепные приросты систематически уменьшаются тоже по абсолютной величине).

Чтобы получить обобщенную характеристику скорости темпов развития изучаемого явления в пределах рассматриваемого периода, рассчитывают средние показатели динамического ряда за единицу времени.

9.5. Средние характеристики ряда динамики

Для обобщающей характеристики динамики используют два типа средних показателей:

✓ средние уровни ряда;

✓ средние показатели изменения уровней ряда.

При расчете средних показателей динамики необходимо соблюдать основные принципы теории средних величин, рассмотренные в гл. 4. Прежде всего это относится к расчету средних величин за период времени, в течение которого условия развития изучаемых явлений существенно менялись. В этом случае общая для всего периода средняя, как правило, является малоинформативной и должна быть дополнена расчетом средних за отдельные этапы развития, т.е. с учетом, предварительной периодизации динамики.

Порядок расчета среднего уровня различается для отдельных видов рядов динамики, которые были рассмотрены в гл. 9.1.

Для рядов динамики с равноотстаюшими по времени уровнями порядок расчета среднего уровня следующий:

а) находим средний уровень интервального ряда абсолютных величин:

$$\bar{y} = \frac{\Sigma y_i}{n};$$

б) определяем средний уровень моментного ряда абсолютных величин:

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

Средний уровень интервального ряда абсолютных величин соответствует рассмотренной выше категории определяющего показателя. Поскольку, как уже отмечалось, уровни такого ряда можно суммировать, то справедливо равенство:

$$y_1 + y_2 + ... + y_n = \bar{y} + \bar{y} + ... + \bar{y} = n\bar{y}$$

Следовательно,

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

где n – число уровней ряда.

Средний уровень моментного **ряда** с равноотстающими уровнями рассчитывается в предположении, что в пределах каждого периода, разделяющего моментные наблюдения, развитие явления происходило по линейному закону. Тогда общий средний уровень вычисляется как среднее значение из средних по каждому интервалу:

$$\bar{y} = \frac{\frac{y_1 + y_2}{2} + \frac{y_2 + y_3}{2} + ... + \frac{y_{n-1} + y_n}{2}}{n-1}.$$

В итоге получаем следующую формулу средней хронологической:

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

Согласно данным, приведенным в табл. 9.1 среднегодовое число постоянного населения нашей страны составило:

$$\bar{y} = \frac{\frac{31,0}{2} + 31,6 + ... + 32,7 + \frac{33,3}{2}}{5-1} = 32,1 \text{ млн чел.}$$

Для моментного ряда с неравными промежутками времени при известных точных датах изменения уровней ряда средний уровень определяется по формуле

$$\bar{y} = \frac{\Sigma yt}{\Sigma t},$$

где t – время, в течение которого сохранялся уровень.

Например, количество обслуживаемых филиалом банка счетов клиентов с 1 по 15 января составляло 300 счетов (t_1 = 15 дней); с 16 января по 20 марта – 370 счетов (t_2 = 65 дней), с 21 марта по 1 апреля – 390 счетов (t_3 = 10 дней). Среднее количество обслуживаемых счетов клиентов за квартал

$$\bar{y} = \frac{300 \cdot 15 + 370 \cdot 65 + 390 \cdot 10}{15 + 65 + 10} = \frac{32450}{90} = 360 \text{ } \textit{счетов.}$$

Средние показатели изменения уровней ряда включают:

- средним абсолютный прирост ($(\Delta\bar{y})$;
- средний коэффициент роста $(\bar{K}_p)$;
- средний темп роста $(\bar{T}_p)$;
- средний темп прироста $(\overline{\Delta T_p})$.

*Средний **абсолютный** прирост* показывает, на сколько единиц в среднем увеличивался или уменьшался каждый уровень ряда по сравнению с предыдущим за ту или иную единицу времени (в среднем ежемесячно, ежегодно и т.п.) и характеризует среднюю абсолютную скорость роста (или снижения) уровня ряда. Его рассчитывают в зависимости от исходных данных следующими способами:

1) **как простую среднюю арифметическую из абсолютных приростов *(цепных)* за последовательные промежутки времени**

$$\Delta\bar{y} = \frac{\Sigma \Delta y^{Ц}}{t},$$

где t – продолжительность периода.

По данным, приведенным в табл. 9.4, среднегодовой абсолютный прирост численности занятых за 2013–2014 гг. составил:

$$\Delta\bar{y} = (296 + 240 + 240 + 222 + (-247)) : 5 = 150{,}2$$ млн чел.;

2) **как частное от деления базисного абсолютного прироста конечного уровня ряда на продолжительность периода** (число усредняемых отрезков времени от базисного до сравниваемого периода):

$$\Delta\bar{y} = \frac{y_n - y_1}{t}.$$

Согласно данным, приведенным в табл. 9.4,

$$\Delta\bar{y} = \frac{(-247) - 296}{5} 9{,}8$$ млн чел.;

3) **через накопленный (базисный) абсолютный прирост** ($\Delta y^б$):

$$\Delta\bar{y} = \frac{\Delta y^б}{t}.$$

По данным, приведенным в табл. 9.4 (графа 4), $\Delta y^б_{2017} = 751$ млн чел., тогда

$$\Delta\bar{y} = \frac{751}{5} = 150{,}2$$ млн чел.

Средний коэффициент роста (снижения) показывает, во сколько раз в среднем за единицу времени изменяется уровень ряда динамики. Для его вычисления используют формулу геометрической средней в предположении, что соблюдается равенство фактического отношения конечного уровня к начальному при замене фактических темпов на средние. В зависимости от наличия *исходных данных* расчет проводят следующим образом:

- если исходной информацией служат цепные коэффициенты роста, то формула имеет вид:

$$\overline{K_p} = \sqrt[t]{K_1^ц \cdot K_2^ц \cdot \ldots \cdot K_t^ц} = \sqrt[t]{\Pi K_p^ц}$$

где П – произведение цепных показателей динамики по данным, приведенным в табл. 9.3.

$$\overline{K_p}=\sqrt[5]{1{,}0236\cdot 1{,}0187\cdot 1{,}0184\cdot 1{,}0167\cdot 0{,}9817}=\sqrt{1{,}0600}=1{,}012=1{,}2\%$$

- через базисный коэффициент роста конечного периода $(K^{б}_{р\,кп})$

$$\overline{K_p}=\sqrt[t]{K^{б}_{р.кп}}.$$

По данным, приведенным в табл. 9.3 (графа 6),

$$K^{б}_{P2017}=1{,}0600,$$

тогда $\overline{K_p}=\sqrt[5]{1{,}0600}=1{,}012=1{,}2\%$

- если известны уровни динамического ряда,

$$\overline{K_p}=\sqrt[t]{\frac{y_n}{y_1}}.$$

Согласно данным, приведенным в табл. 9.3 (гр. 2):

$$\overline{K_p}=\sqrt[5]{\frac{13273}{12522}}=\sqrt[5]{1{,}600}=1{,}2\%$$

Подчеркнем выявленную *взаимосвязь базисных и цепных коэффициентов роста*: базисный коэффициент роста равен произведению всех цепных коэффициентов роста, рассчитанных за этот же период.

Соответственно, *взаимосвязь базисных и цепных темпов изменения* может быть определена так: базисный темп изменения равен результату произведения всех цепных коэффициентов роста, рассчитанных за этот же период, помноженному на 100.

Эта правило может пригодиться в анализе рядов динамики;

4) ***средний темп прироста (сокращения)*** рассчитывают с использованием среднего темпа изменения (роста или снижения):

$$\overline{T}_{ПР(сокр)}=\overline{T}_{Р(СН)}-100\%.$$

Аналогично рассчитывают *средний коэффициент прироста*:

$$\overline{K}_{ПР}=\overline{K}-1.$$

Особую осторожность при применении средних показателей следует соблюдать в тех случаях, когда появляется ***перелом в тенденции*** изменения уровней ряда динамики. Если, например, в развитии какого-либо социально-экономического явления сначала наблюдалась тенденция роста, затем по каким-либо причинам, появилась тенденция снижения, то сначала рассчитывают средние показатели для периода роста, затем – для периода снижения.

Если уровни ряда динамики сокращаются, то соответствующие показатели темпа прироста будут со знаком минус, так как они характеризуют относительное уменьшение прироста уровня ряда динамики (табл. 9.4). Численность занятых в экономике РУз в 2018 г. по сравнению с 2017 г. сократилась на 247 тыс. чел, что означает снижение их на 1,83 %.

$$\overline{T}_{ПР(СОК)} = \overline{T}^{Ц}_{Р(СН)} - 100 = 98{,}17 - 100 = -1{,}83\%$$

т.е. произошло сокращение численности занятых в экономике на 1,8%.

Важным статистическим показателем динамики социально- экономических процессов является *темп наращивания,* который в условиях интенсификации экономики измеряет наращивание во времени экономического потенциала.

Вычисляются *темпы* наращивания (Тн) делением цепных абсолютных приростов $\Delta y_{ц_i}$ на уровень, принятый за постоянную базу сравнения, y_{0_i}.

$$T_{H_i} = \Delta y_{Ц_i} : y_{O_i}.$$

Темпы наращивания можно непосредственно определять и по базисным темпам роста:

$$T_{H_i} = \frac{\Delta y_{Ц_i}}{y_{O_i}} = \frac{y_t - y_i - 1}{y_{O_i}} = Tp_{б_i} - Tp_{б_{i-1}}.$$

Это формула удобна для практики, так как статистическая информация о динамике социально-экономических явлений публикуется чаще всего в виде базисных рядов динамики.

9.6. Методы выявления основной тенденции (тренда) в рядах динамики

Важным направлением в исследовании закономерностей динамики социально-экономических процессов является изучение общей тенденции развития *(тренда)*. Это можно осуществить, применяя специальные методы анализа рядов динамики. Конкретное их использование, зависит от характера исходной информации и предопределяется задачами анализа.

Теоретически при анализе рядов динамики различают следующие компоненты:

- *тенденция, или тренд;*
- *периодически повторяющиеся колебания;*
- *случайные колебания.*

Под тенденцией *понимают общее направление изменения уровней ряда: к росту, снижению или стабилизации с течением времени.*

К периодически повторяющимся колебаниям *относят долговременные циклические колебания и кратковременные или сезонные колебания (регулярные изменения внутри года).*

Случайные колебания *складываются под влиянием внешних факторов.*

Выявление основной тенденции развития в статистике называют выравниванием временного ряда. Тенденция выявляется различными методами, к числу которых, как правило, относят следующие:

– метод укрупнения интервалов;

– метод скользящей средней (механическое сглаживание);

– аналитическое выравнивание.

Укрупнение интервала

В ряде случаев общая тенденция развития явления может быть выявлена при помощи укрупнения интервалов. Суть этого приема заключается в том, что первоначальный ряд преобразуется и заменяется другим, показатели которого охватывают большие периоды времени чем те, которые имелись в первоначальном ряду. Выбор величины новых периодов зависит от конкретных особенностей ряда, в частности, от величины их в первоначальном ряду. Так, например, ряд, составленный из дневных (суточных) данных, может быть преобразован в ряд пятидневных, недельных или декадных данных. Ряд, содержащий помесячные данные, может быть преобразован в ряд квартальных или полугодовых данных; ряд годовых данных может быть преобразован в ряд пятилетних, десятилетних данных и т.д.

Вновь образованные периоды могут характеризоваться либо абсолютными (суммарными) величинами, либо средними. В первом случае уровни нового ряда получаются путем простого суммирования уровней первоначального ряда, если они были выражены в абсолютных величинах. Во втором случае находят средние величины показателя на каждый вновь образованный период.

Проиллюстрируем сказанное на конкретных примерах (табл. 9.5).

Таблица 9.5

Валовый сбор хлопка сырца в фермерских хозяйствах области

Годы	Валовый сбор хлопка (тас.т)	Пятилетние итоговые данные	Пятилетние средние	Сумма валового сбора хлопка по пятилеткам	Пятилетняя средняя скользящих
1	2	3	4	5	6=5:5
2004	82,5			–	–
2005	86,6			–	–
2006	103,7	499,4	99,9	499,4	99,9
2007	125,0			552,6	110,5
2008	102,6			585,5	117,1
2009	134,7			607,3	121,5
2010	119,5			613,1	122,6
2011	125,5	650,7	130,1	650,7	130,1
2012	130,8			623,5	124,7
2013	140,2			656,7	131,3
2014	107,5			651,7	130,4
2015	152,1			692,1	138,4
2016	121,1	690,8	140,0	699,8	140,0
2017	171,2			–	–
2018	147,4			–	–

В гр. 2, табл. 9.5 обнаруживается тенденция роста валового сбора хлопка хотя в отдельные годы он и уменьшается, причем в некоторые (2008, 2014, 2016) – довольно значительно. Для того чтобы ясно установить общую тенденцию динамики производства хлопка в рассматриваемый период, найдем суммарный его сбор по пятилетиям (гр. 3).

Пятилетие итоговые данные обнаруживают рост валовых сборов хлопка: от пятилетия к пятилетию они увеличиваются. Определим теперь размеры среднегодовых сборов по пятилетиям, для чего суммарные валовые сборы за каждое пятилетие разделим на 5. В результате получим данные гр. 4.

Из этих данных видно, что среднегодовые валовые сборы хлопка от пятилетия к пятилетию увеличиваются, причем наиболее значительный прирост среднегодового валового сбора хлопка (на 30%) приходится на пятилетие 2009-2013 гг. Следует отметить, что при укрупнении интервалов число уровней динамического ряда существенно сокращается. Кроме того, при анализе не учитывается изменение уровней внутри укрупненных интервалов. В связи с этим для более детальной характеристики тенденции изменения уровней используют выравнивание динамического ряда с помощью скользящей (подвижной) средней.

Скользящие средние

Суть способа сглаживания ряда динамики при помощи скользящей средней заключается в том, что последовательно находятся средние значения из нескольких членов ряда (например, из трех, пяти, десяти), начиная с первого, со второго, с третьего и т.д.

Например, если дан ряд ежегодных уровней: $y_1, y_2, \dots, y_9$, то трехлетнюю скользящую среднюю определяют следующим образом:

- для первого интервала $\breve{y}_1 = \dfrac{y_1 + y_2 + y_3}{3}$;
- второго интервала $\breve{y}_2 = \dfrac{y_2 + y_3 + y_4}{3}$;

• третьего интервала $\breve{y}_3 = \frac{y_3 + y_4 + y_5}{3}$ и т.д.

Таким образом, использование скользящей средней позволяет осуществить замену фактических уровней динамического ряда расчетными, имеющими значительно меньшую колеблемость, чем исходные данные. При низкой колеблемости тенденция развития явления становится более очевидной (см. гр. 5 и гр. 6 табл. 9.5).

Число уровней, по которым рассчитывают скользящую среднюю, называют *периодом (интервалом) сглаживания*. Чем он меньше, тем больше сглаженный ряд приближается к исходному фактическому.

Если требуется получить более плавный вид изменения уровней ряда, то используют более длительный интервал сглаживания, но тогда выравненный ряд будет еще короче. Так, если в рассматриваемом нами примере исходный ряд стал короче на два крайних уровня при периоде сглаживания, равном трем, то при периоде сглаживания, равном пяти, он будет короче на четыре уровня. Вопрос о том, какой период сглаживания следует использовать, решают в зависимости от характера колебаний уровней фактического динамического ряда.

Если колебания имеют определенную периодичность, то период сглаживания следует принять равным (или кратным) периоду колебаний. Так, при наличии динамического ряда с уровнями за каждый месяц, которые ежегодно отличаются сезонными колебаниями, целесообразно использовать 12-месячный (или 24-месячный) период сглаживания, а при наличии уровней за кварталы – четырех или восьмиквартальный. Если колебания уровней беспорядочные, то следует постепенно укрупнять период сглаживания, пока не выявится отчетливая картина тренда. Предпочтительнее **применять период сглаживания с нечетным числом уровней**, поскольку в этом случае расчетное значение уровня окажется в центре числа слагаемых скользящей средней и им легко заменить фактическое значение.

При четном периоде сглаживания используют специальную процедуру центрирования.

Центрирование

Центрирование заключается в нахождении средней из двух смежных скользящих средних. Оно осуществляется для того, чтобы соотнести полученный уровень с определенной датой.

Метод скользящей средней позволяет получать общие представления о направлении развития уровней ряда. Рассмотренные выше отдельные свойства скользящей средней несколько ограничивают возможности этого метода при изучении характера выявленной тенденции:

– выравниванию подлежат не все уровни ряда и сглаженный ряд сокращается;

– не представлена необходимая для целей прогнозирования аналитическая формула тенденции развития.

В связи с этим в ряде случаев метод скользящей средней применяют как вспомогательный, облегчающий использование других методов выявления тенденции и, в частности, метода аналитического выравнивания.

Аналитическое сглаживание

Сглаживание рядов динамики при помощи скользящей средней обычно называют *эмпирическим сглаживанием,* так как оно не основано на каком-либо аналитическом уравнении. Сглаживание же на основе математического уравнения называют *аналитическим сглаживанием или выравниванием рядов динамики.*

Выравнивание рядов динамики, основанное на известном из курса математической статистики способе наименьших квадратов, производится по уравнению прямой или какой-либо кривой линии, при этом уровни ряда динамики рассматриваются как функции времени.

Уравнение, по которому следует произвести выравнивание ряда, выбирают путем анализа характера динамики данного явления. Динамика социально-экономических явлений может быть выражена аналитическим уравнением лишь приближенно. Выбирая уравнение для выравнивания, нужно стремиться к тому, чтобы данные, полученные на его основе, возможно близко подходили к фактическим значениям членов рассматриваемого ряда динамики.

При аналитическом сглаживании, в целях облегчения нахождения параметров уравнения, отсчет времени следует производить так, чтобы сумма показателей времени рассматриваемого ряда динамики была равной нулю. Такой счет времени не сказывается на выводах, так как в любом счете времени есть какое-то условное начало отсчета.

Проиллюстрируем процесс выравнивания ряда динамики на следующих данных (табл. 9.6).

Таблица 9.6

Валовой сбор хлопка-сырца в фермерском хозяйстве «Авангард»

	2014	2015	2016	2017	2018
Валовой сбор хлопка-сырца – тыс. т	4,7	5,2	5,8	6,1	6,5

Данные показывают, что на протяжении периода, охватываемого рядом динамики, сохраняется более или менее стабильный абсолютный прирост (0,3–0,6). Это позволяет произвести выравнивание ряда по прямой. Уравнение прямой применительно к выравниванию рядов динамики запишем в следующем виде;

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

где $\bar{y}_t$ – сглаженные значения ряда динамики;

t – время.

Для этого уравнения способ наименьших квадратов дает следующую систему нормальных уравнений:

$$na_0 + a_1\Sigma t = \Sigma y$$

$$a_0\Sigma t + a_1\Sigma t^2 = \Sigma ty$$

где n – число членов ряда динамики. При условии, что $\Sigma t = 0$, эта система уравнений примет следующий вид:

$$na_0 = \Sigma y$$

$$a_1\Sigma t^2 = \Sigma ty$$

Отсюда находим, что

$$a_0 = \frac{\Sigma y}{n};\quad a_1 = \frac{\Sigma ty}{\Sigma t^2}.$$

При нечетном числе членов ряда динамики будет равна нулю, если члену, находящемуся в середине ряда, придается нулевое значение времени, т.е. он принимается за условное начало отсчета времени. Ниже его будут идти значения времени со знаком плюс, увеличиваясь всякий раз на одну единицу (+ 1, +2, +3 и т.д.), а выше – тот же ряд чисел, но со знаком минус (–1, –2, –3. и т.д.).

Учитывая сказанное, рассчитаем необходимые суммы для определения a_0 и a_1 (табл. 9.7).

Подставив затем соответствующие суммы в приведенные выше уравнения параметров a_0 и a_1 получим:

$$a_0 = \frac{28{,}3}{5} = 5{,}66;\quad a_1 = \frac{4{,}5}{10} = 0{,}45.$$

Таблица 9.7

Расчет сумм, необходимых для выравнивания (сглаживания) ряда динамики по прямой

Годы	**t**	**y**	**t^2**	**ty**	$\bar{y}_t$
А	**1**	**2**	**3**	**4**	**5**
2014	–2	4,7	4	–9,4	4,76
2015	–1	5,2	1,	–5,2	5,21
2016	0	5,8	0	0	5,66
2017	1	6,1	1	6,1	6,11
2018	2	6,5	4	13,0	6,56
–		28,3	10	4,5	28,3

Не трудно заметить, что здесь a_0 представляет собой среднюю арифметическую из уровней ряда, принимаемую за серединный уровень, *а* a_1 – прирост уровня за единицу времени.

Подставив найденные значения параметров в уравнение прямой, будем иметь

$$\bar{y}_t = 5{,}66 + 0{,}45t$$

Для определения сглаженных уровней ряда динамики ($\bar{y}_t$) подставим в это уравнение для каждого года соответствующее ему значение *t*.

Так, подставив вместо *t* минус ***2***, найдем сглаженное значение уровня ряда для 2014 г.:

$$\bar{y}_t = 5{,}66 + 0{,}45(-2) = 4{,}76$$

Проделав соответствующие расчеты для остальных лет, получим ряд выравненных (сглаженных) значений уровня ряда (гр. 5 табл. 9.7), который иногда называют **теоретическим уровнем**. Этот ряд показывает, как бы шло изменение уровней ряда динамики, если бы оно было целиком и полностью подчинено данному математическому закону.

В рассмотренном примере показано, как отсчитывается время при нечетной числе членов ряда динамики. В случае, *если* в ряду четное *число* членов, то в середине его находятся два члена. Тогда за начало отсчета принимается промежуточный момент между ними, а за меру счета времени не единица времени, содержащаяся в данном ряду, а половина ее, например, полугодие. В таком случае каждый в ряду член будет отстоять от соседнего на расстоянии, равном двум единицам времени, а не одной, как в нечетном ряду. Показатели времени в этом случае будут представлять собой ряд чисел с разностью в 2 единицы, т.е. 1, 3, 5, 7, 9 и т.д., причем как и в предыдущем *случае:* вниз (вправо) – со знаком плюс, вверх (влево) – со знаком минус.

Так, если бы ряд динамики содержал данные за 6 лет, например, за 2014-2012 гг., то в этом случае *t* имело бы следующие значения (табл. 9.8).

Таблица 9.8

Годы	**2014**	**2015**	**2016**	**2017**	**2018**	**2019**
t	–5	–3	–1	+1	+3	+5

При подобном отсчете времени всегда будет равно нулю. Для выравнивания рядов динамики часто используются уравнения кривых линий, в частности, парабола второго порядка:

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

Для нахождения параметров этого уравнения используется следующая система нормальных уравнений:

$$na_0 + a_1\Sigma t + a_2\Sigma t^2 = \Sigma y$$
$$na_0\Sigma t + a_1\Sigma t^2 + a_2\Sigma t^3 = \Sigma ty$$
$$a_0\Sigma t^2 + a_1\Sigma t^3 + a_2\Sigma t^4 = \Sigma t^2 y$$

При соблюдении принципа отсчета времени от условного нуля сумма нечетных степеней t всегда будет равна нулю (в рассматриваемом случае Σt и Σt^3). Вследствие этого приведенная выше система нормальных уравнений примет следующий вид:

$$na_0 + a_2\Sigma t^2 = \Sigma y$$
$$a_1\Sigma t^2 = \Sigma ty$$
$$a_0\Sigma t^2 + a_2\Sigma t^4 = \Sigma t^2 y$$

Решая эту сумму уравнений, найдем:

$$a_0 = \frac{\Sigma y \cdot \Sigma t^4 - \Sigma t^2 \cdot \Sigma t^2 y^4}{n\Sigma t^4 - \left(\Sigma t^2\right)^2};$$

$$a_1 = \frac{\Sigma ty}{\Sigma t^2};$$

$$a_2 = \frac{n\Sigma t^2 y - \Sigma y \cdot \Sigma t^2}{n\Sigma t^4 - \left(\Sigma t^2\right)^2};$$

В качестве примера произведем выравнивание по параболе второго порядка данных о производстве электроэнергии в регионе «А» (гр. 1 табл. 9.9).

Необходимые суммы для определения параметров a_0, a_1 и a_2 содержатся в таблице (гр. 3–6). Используя их, найдем:

$$a_0 = \frac{15{,}2 \times 1414 - 70 \times 196}{6 \times 1414 - 70^2} = 2{,}169;$$

$$a_1 = \frac{31}{70} = 0{,}443;$$

$$a_2 = \frac{6 \times 196 - 15{,}2 \times 70}{6 \times 1414 - 70^2} = 0{,}031$$

Подставив найденные значения параметров в уравнение параболы второго порядка, получим:

$$\bar{y}_t = 2{,}169 + 0{,}443t + 0{,}031t^2 .$$

Таблица 9.9

Производство электроэнергии в регионе «А» и расчет необходимых сумм для определения параметров параболы второго порядка

Годы	Произведено электроэнергии млрд. *квт-ч, у*	t	t^2	t^4	ty	t^2y	$\bar{y}_t$
А	1	2	3	4	5	6	7
2014	0,9	–5	25	625	–4,5	22,5	0,729
2015	1,0	–3	9	81	–3,0	9,0	1,119
2016	1,3	–1	1	1	–1,3	1,3	1,757
2017	3,1	1	1	1	3,1	3,1	2,643
2018	3,9	3	9	81	11,7	35,1	3,777
2019	5,0	5	25	625	25,0	125,0	5,159
	15,2		70	1414	31,0	196,0	15,184

Подставляя в это уравнение соответствующие значения, найдем ряд выравненных значений уровня ряда (гр. 7 табл. 9.9). Эти значения почти совпадают с заданными значениями.

При аналитическом выравнивании рядов динамики необходимо учитывать, что получаемый в результате сглаживания плавный уровень устраняет, сглаживает колебания уровней ряда, скачки в их изменении, причем не только вызванные случайными обстоятельствами, но и обусловленные закономерным развитием явления. Вместе с тем сглаживание должно не затушевывать закономерности развития явления, а, напротив, выявлять их. Поэтому период, за который производится сглаживание, не должен содержать скачков, обусловленных закономерным ходом развития. Если же ряд динамики содержит такие скачки, то значит, мы имеем дело с различными по своему характеру периодами и сглаживание должно производиться для каждого из них в отдельности.

Необходимо также иметь в виду, что при аналитическом сглаживании уровень ряда выражается как функция времени (t). Однако уровень ряда изменяется от одного периода времени к другому не просто потому, что прошло какое-то время (скажем, год, как в нашем примере); он изменяется потому, что в течение этого времени действовали и оказывали на него влияние различные факторы, например, повышение производительности труда, ввод в действие новых мощностей и т.п. Это надо учитывать при анализе выравненного ряда динамики.

9.7. Методы изучения сезонных колебаний

Измерение сезонных колебаний

Многие общественные явления и процессы носят сезонный характер, испытывают сезонные колебания. Уровень их из года в год в определенные месяцы повышается, а в другие снижается. Эти внутригодичные *колебания,* имеющие периодический характер, и носят название *сезонных колебаний.*

Сезонный ряд динамики (тренд сезонный) – *динамический ряд с периодически повторяющимися уровнями.*

Сезонные колебания всегда в конечном счете связаны с влиянием природных факторов, различных в разное время года. Особенно, наглядно это проявляется в сельском хозяйстве, где, например, получение большинства продуктов растениеводства происходит только в определенные месяцы. Сезонность производства большинства сельскохозяйственных продуктов приводит в свою очередь к сезонным колебаниям в работе предприятий, перерабатывающих сельскохозяйственное сырье, особенно сырье, не подлежащее длительному хранению. Наблюдаются сезонные колебания и в других отраслях народного хозяйства – в строительстве, на транспорте, в торговле и т.д.

Сезонность – явление отрицательное. Она приводит простоям и неравномерному использованию в течение года трудовых ресурсов и оборудования, к понижению производительности труда и повышению себестоимости продукции. В связи с этим задача заключается в том, чтобы преодолеть, ликвидировать сезонность, а если это пока еще невозможно, то уменьшить, смягчить сезонные колебания. Преодоление сезонности является важным резервом улучшения использования материальных и трудовых ресурсов, и увеличения производства материальных благ.

Отсюда вытекает необходимость изучения сезонности и измерения сезонных колебаний. Измерение сезонных колебаний позволяет судить об эффективности и результатах борьбы с сезонностью. Вместе с тем оно необходимо также для того, чтобы при внутригодичном планировании предвидеть и учесть такие сезонные колебания, которые пока что не удается устранить. Механическая разбивка годового плана по месяцам и кварталам могла бы привести при этом к серьезным неувязкам.

Измерение сезонных колебаний в статистике производится путем исчисления индексов сезонности. *Индекс сезонности представляет собой отношение фактического уровня явления за тот или иной месяц к выравненному уровню за тот же месяц* (или к среднемесячному уровню за год) и выражается обычно в процентах:

$$I_{сезонности} = \frac{y_{факт}}{y_{вырав}} \times 100 . \quad (100)$$

Существуют различные методы исчисления индексов сезонности, отличающиеся друг от друга способами расчета выравненных уровней. При наиболее простом способе выравненный уровень представляет собой простую среднюю арифметическую из месячных уровней, т.е. среднемесячный уровень за год.

Так как в каждом году сезонные колебания имеют обычно свои особенности, то индексы сезонности, как правило, исчисляются не за один год, а за несколько лет. Для получения индексов, свободных от особенностей отдельных лет и отражающих типичный характер сезонных колебаний, из индексов за одноименные месяцы ряда лет вычисляют простую среднюю арифметическую.

Исчислим методом простой средней арифметической индексы сезонности отработанных работниками фермерского хозяйства человеко-дней. Среднемесячный уровень за 2017 г. равен 150:12 = 12,5 (тыс. чел.-дней), а 2018 г. – 15 (тыс. чел.-дней). Разделив уровень каждого месяца на среднемесячный уровень за соответствующий год, получим следующие индексы сезонности (табл. 9.10).

Таблица 9.10

Индексы сезонности затрат труда в фермерском хозяйстве

Месяцы	Отработано работниками (тыс.)		Индексы сезонности (%)					
			метод средней арифметической			метод выравнивания по прямой		
	2015 г.	2016 г.	2017 г.	2018 г.	в среднем за 2017-2018 гг.	2017 г.	2018 г.	в среднем за 2017-2018 гг.
I	8,0	10,6	64	71	68	72	76	74
II	8,2	10,0	66	67	66	73	71	72
III	10,2	11,6	82	77	80	89	81	85-
IV	12,0	12,8	96	85	90	102	88	95
V	14,8	15,8	118	105	112	123	107	115
VI	16,8	19,0	134	127	130	138	126	132
VII	17,2	19,4	138	129	134	138	127	132
VIII	16,4	19,8	131	132	132	129	128	128
IX	15,2	19,0	122	127	124	118	120	119
X	11,6	17,4	93	116	104	88	109	98
XI	10,4	14,6	83	97	90	78	90	84
XII	9,2	10,0	74	67	70	67	61	64
Итого	150,0	180,0	–	–	–	–	–	–

Метод средней арифметической пригоден только в тех случаях, когда в ряду динамики нет отчетливо выраженной тенденции к росту или снижению уровня. Если же уровень явления имеет тенденции к росту или снижению, то отклонения от постоянного среднего уровня могут дать искаженную картину сезонных колебаний. Поэтому в этих случаях нужно измерять колебания не около постоянного уровня, а около таких уровней, в которых выражена эта общая тенденция к росту или снижению. Такими уровнями могут служить звенья 32-месячной скользящей средней, а также уровни, полученные путем аналитического выравнивания.

Рассчитаем индексы сезонности для нашего примера, использовав уровни, полученные методом выравнивания по прямой (табл. 9.11). Для этого фактический уровень каждого месяца выразим в процентах к соответствующему выравненному уровню. Так, для января 2017 г. получим: (8,0 : 11,04) × 100 = 72% и т.д. (табл. 9.10).

Для определения параметров уравнений при Σt = 0 составляется матрица расчетных показателей (табл. 9.11).

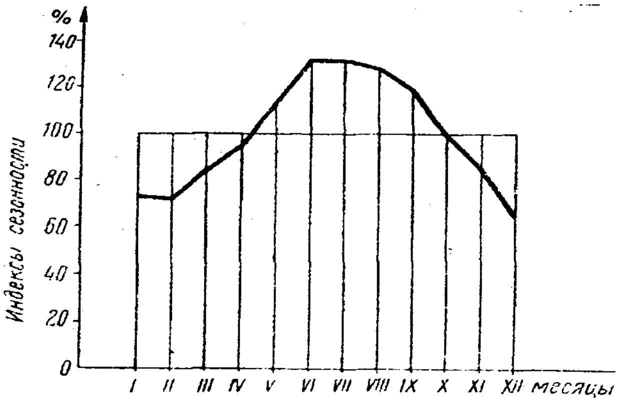


Рис. 9.5. **Сезонность затрат труда в фермерском хозяйстве**
(в среднем за 2017–2018 гг.)

Для прямолинейной функции $\bar{y}_t = a_0 + a_1 t$ расчет параметров (при Σt = 0) производится по формуле:

$$a_0 = \frac{\Sigma y}{n} = \frac{330}{24} = 13{,}75\ ;$$

$$a_1 = \frac{\Sigma yt}{\Sigma t^2} = \frac{542{,}8}{4 \cdot 600} = 0{,}118\ .$$

Отсюда уравнение прямой в нашем примере:
$\bar{y}_t = 13{,}75 + 0{,}118t$.

Подставляя в это уравнение соответствующие значения t найдем выравненные уровни $\bar{y}_t$ (последнюю графу табл. 9.11). Так,

• для января 2017 г. (t= –23) получим: $\bar{y}_t$ = 13,75 + 0,118 × (–23) = 11,04 и т.д.;

• для января 2018 г. (t = +10,6) получим $\bar{y}_t$= 13,75 + 0,118 × (+10,6) = = 13,87 и т.д.

Таблица 9.11

Выравнивание ряда динамики по прямой

(данные условные)

Год и месяц	**Отработано работниками человеко-дней (тыс.) y**	*t*	*yt*	t^2	$\bar{y}_t$
Январь 2017	*8,0*	*–23*	*–184,0*	*529*	*11,04*
Февраль »	*8,2*	–21	–172,2	*441*	*11,27*
Март »	*10,2*	–19	–193,8	*361*	*11,51*
Апрель »	*12,0*	–17	–204,0	*289*	*11,74*
Май »	*14,8*	–15	–222,0	*225*	*11,98*
Июнь »	*16,8*	–13	–218,4	*169*	*12,22*
Июль »	*17,2*	–11	–189,2	*121*	*12,45*
Август »	*16,4*	–9	–147,6	*81*	*12,69*
Сентябрь »	*15,2*	–7	–106,4	*49*	*12,92*
Октябрь »	*11,6*	–5	–58,0	*25*	*13,16*
Ноябрь »	*10,4*	–3	–31,2	*9*	*13,40*
Декабрь »	*9,2*	–1	–9,2	*1*	*13,63*
Январь 2018	*10,6*	*+1*	*+10,6*	*1*	*13,87*
Февраль »	*10,0*	+3	+30,0	*9*	*14,10*
Март »	*11,6*	+5	+58,0	*25*	*14,34*
Апрель »	*12,8*	+7	+89,6	*49*	*14,58*
Май »	*15,8*	+9	+142,2	*81*	*14,81*
Июнь »	*19,0*	+11	+209,0	*121*	*15,05*
Июль »	*19,4*	+13	+252,2	*169*	*15,28*
Август »	*19,8*	+15	+297,0	*225*	*15,52*
Сентябрь »	*19,0*	+17	+323,0	*289*	*15,76*
Октябрь »	*17,4*	+19	+330,0	*361*	*15,90*
Ноябрь »	*14,6*	+21	+306,0	*441*	*16,23*
Декабрь »	*10,0*	+23	*+230,0*	*529*	*16,46*
Итого	*330,0*	*0*	*+542,8*	*4600*	*330,00*

Изобразим исходные данные и найденную прямую на графике (рис. 9.6).

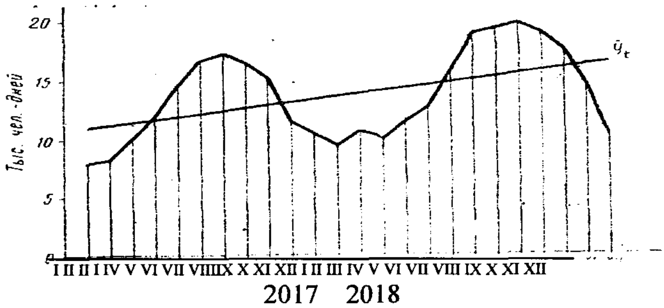


Рис. 9.5. **Количество отработанных чел.-дней**

Отметим, что при перенесении начала отсчета времени в середину периода параметр $\mathbf{a_0}$ представляет собой величину выравненного уровня в середине периода, а параметр a_1 – абсолютный прирост выравненного уровня, за единицу измерения времени (в нашем примере – за полмесяца).

Количество, отработанных работниками человеко-дней испытывает сезонные колебания (летние подъемы и зимние спады), которые затушевывают общую тенденцию. Переход к среднемесячным уровням за каждый год погасил бы не только сезонные колебания, но и нарастание уровня из месяца в месяц на протяжении этих двух лет. Поэтому, чтобы освободить общую тенденцию динамики от переплетающихся с ней сезонных колебаний, сохранив помесячные уровни произведем выравнивание по прямой (табл. 9.11).

9.8. Методы прогнозирования

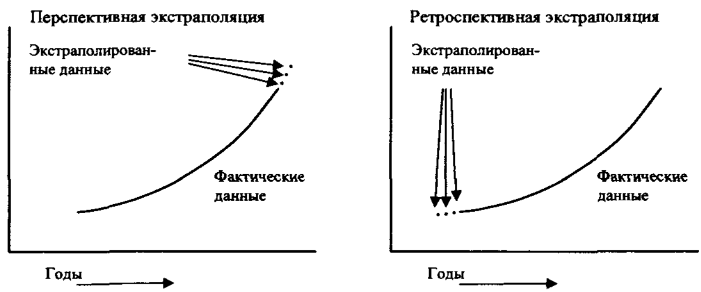
Основы прогнозирования

Прогнозирование в экономике – определение будущих размеров социально-экономических явлений на основе анализа тенденций их развития. Предполагается, что закономерность, действующая в прошлом, сохранится и в прогнозируемом будущем, т.е. прогноз основан на **экстраполяции.**

В зависимости от сроков прогнозирования различают:

1) оперативные прогнозы (до 1 месяца);
2) краткосрочные прогнозы (до 1 года);
3) среднесрочные прогнозы (от 1 года до 5 лет);
4) долгосрочные прогнозы (свыше 5 лет).

Говоря об экстраполяции рядов динамики, чаще подразумевают **перспективную экстраполяцию,** т.е. прогноз на будущее, но есть еще **ретроспективная экстраполяция** – прогноз в прошлое.



Теоретической основой распространения тенденции на будущее является такое известное свойство социально-экономических явлений, как инерционность.

Экстраполяцию следует рассматривать как начальную стадию построения окончательных прогнозов. Механическое использование экстраполяции может стать причиной неправильных выводов. Всегда следует учитывать все необходимые условия предпосылки и гипотезы, связанные с экономико-теоретический анализом.

Чем короче период экстраполяции (период упреждения), тем более надежные и точные результаты дает прогноз. В целях получения научно обоснованных прогнозов в статистике используется правило, при котором период упреждения не должен превышать 1/3 периода исследования.

В зависимости от принципов и исходных данных, положенных в основу прогноза, выделяют следующие элементарные методы экстраполяции:

1) по среднему абсолютному приросту:

$$\hat{y}_{i+t} = y_i + \overline{\Delta}t,$$

где $\hat{y}_{i+t}$ – прогнозируемый уровень; t – период упреждения (число лет, кварталов и т.п.); y_i – базовый для прогноза уровень; $\overline{\Delta}$ – средний за исследуемый период абсолютный прирост (среднегодовой, среднеквартальный и т.п.);

2) по среднему коэффициенту роста:

$$\hat{y}_{1+t} = y_i + \overline{K}^t,$$

где $\overline{K}$– средний за исследуемый период коэффициент роста (среднегодовой, среднеквартальный и т.п.);

3) на основе выравнивания рядов по какой-либо аналитической формуле:

Так, например, модель прямолинейной зависимости имеет вид

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

Параметры a_0 и a_1 искомой прямой, удовлетворяющие принципу наименьших квадратов, находятся путем решения следующей системы уравнений:

$$a_0 n + a_1 \Sigma t = \Sigma y;$$

$$a_0 \Sigma t + a_1 \Sigma t^2 = \Sigma yt,$$

где y – фактические уровни ряда динамики; n – число уровней.

Эта система уравнений значительно упрощается, если начало отсчета времени перенести в середину рассматриваемого периода. При нечетном числе уровней получаем тогда такие значения t:

Годы	2014	2015	2016	2017	2018
t	–2	–1	0	+1	+2

При четном числе уровней значения t устанавливаются так, как это показано в табл. 9.7 гр. 1. В обоих случаях $\Sigma t = 0$, в результате чего система уравнений упрощается и параметры прямой оказываются равны:

$$a_0 = \frac{\Sigma y}{n}; \quad a_1 = \frac{\Sigma yt}{\Sigma t^2}$$

Проиллюстрируем выравнивание по прямой в я следующем примере (см. табл. 9/12).

Известно, что группой фирм одного из региона РУз за год было продано 3244 единицы условного однокачественного товара. Среднегодовой абсолютный прирост количества проданного товара составляет:

$$\overline{\Delta}_{абс} = \frac{y_n - y_1}{n-1} = \frac{3244 - 1994}{5-1} = 312{,}5 \text{ (шт.)}.$$

Среднегодовой темп роста товарооборота в натуральном выражении составляет:

$$\overline{T}_p = \overline{K} \cdot 100;$$

$$\overline{K} = \sqrt[n-1]{\frac{y_n}{y_1}} = \sqrt[5-1]{\frac{3244}{1994}} = \sqrt[4]{1{,}6269} = 1{,}1294;$$

$$\overline{T} = 112{,}94\%$$

Требуется:

1. Используя методы экстраполяции, сделать перспективный прогноз товарооборота в натуральном выражении на два года:

а) по среднему абсолютному приросту;

б) по среднему темпу роста;

в) на основе аналитического выравнивания по прямой.

2. Сделать выводы.

Решение:

1. Рассчитаем прогнозируемый уровень ряда динамики по среднему абсолютному приросту:

$$\hat{y}_{i+1} + \overline{\Delta}t = 3244 + 312{,}5 \cdot 2 = 3869 \text{ (шт.)}.$$

2. Рассчитаем прогнозируемый уровень ряда динамики по среднему коэффициенту роста:

$$\hat{y}_{i+1} + y_i \overline{K}^t = 3244 \cdot 1{,}129^2 = 4135 \text{ (шт.)}.$$

3. Рассчитаем прогнозируемый уровень ряда динамики на основе аналитического выравнивания по прямой. С этой целью построим вспомогательную табл. 9.12, позволяющую рассчитать параметры тренда.

Таблица 9.12

Вспомогательная таблица для расчета параметров тренда

Год	Количество проданного товара, шт., y_i	Условное обозначение периодов, t_i	$y_i t_i$	t_i^2	Выровненные уровни ряда динамики, шт., $\hat{y}_i$
1	2	3	4	5	*6*
1-й	1994	–2	–3988	4	2045,8
2-й	2440	–1	–2440	1	2344,0
3-й	2611	0	0	0	2642,2
4-й	2922	1	2922	1	2940,4
5-й	3244	2	6488	4	3238,6
Итого	13211	0	2982	10	13211,0

Параметры тренда прямой:

$$a_0 = \frac{\Sigma y}{n} = \frac{13211}{5} = 2642{,}2$$

$$a_1 = \frac{\Sigma yt}{\Sigma t^2} = \frac{2982}{10} = 298{,}2$$

Тренд имеет вид: $\hat{y} = 2642{,}2 + 298{,}2t$.

Правильность расчетов уровней выровненного ряда динамики проверяется следующим образом: сумма значений эмпирического ряда совпадает с суммой значений выровненного ряда динамики, т.е. $\sum_{i=1}^{n} y_i = \sum_{i=1}^{n} \hat{y}_i$ (итоги граф 2 и 6 табл. 9.12).

Прогнозируемый уровень товарооборота в натуральном выражении группы фирм одного из регионов РУз на 2 года вперед на основе аналитического выравнивания ряда динамики по прямой составляет:

$$\hat{y} = 2642{,}2 + 298{,}2t = 2642{,}2 + 298{,}24 = 3835 \text{ (шт.)}.$$

Пояснения к расчетам: в уравнении тренда прямой взято $t = 4$, так как для 5-го года исследования t = 2 (табл. 9.*12* графу

3), тогда для 6-го года будет t = 3, а для 7-го (прогнозируемого) t =4.

Как показывают полученные данные, все прогнозы товарооборота условного товара (шт.) группы предприятий одного из регионов РУз на 2 года вперед довольно близки между собой (3869; 4135; 3835). Расхождение полученных данных объясняется тем, что в основу прогнозов положены разные методики расчета.

Интерполяция рядов динамики

Бывают случаи, когда в ряду динамики не достает данных за какой-либо период (или момент) времени, или нужно определить уровень явления на будущее, выходя за пределы данного ряда. Нахождение неизвестного промежуточного члена ряда-динамики называется *интерполяцией,* а нахождение уровня ряда за его пределами, в перспективе на будущее, – *экстраполяцией.*

Наиболее простым приемом интерполяции является следующий расчет: из двух членов ряда динамики, непосредственно примыкающих к неизвестному члену ряда, находится средняя величина, которая и принимается за искомый показатель.

Допустим, например, что нам не известен численность занятых в экономике в РУз в 2017 г., но известны уровни 2015 и 2017 гг. которые равны, соответственно, 13058 и 13520 тыс. чел. Приняв среднюю из этих величин за численность занятых в 2016 г., найдем, что он равен 13289 тыс. $\left(\frac{13038+13520}{2}\right)$, что 9 тыс. чел. меньше фактических данных (см. табл. 9.3). Это операция называется интерполяцией.

Иногда берут не один, а два (или больше) предшествующих уровня и столько же последующих и среднюю из этих уровней принимают за отсутствующий уровень ряда.

Интерполяция производится исходя из предположения о той или иной закономерности изменения уровня явления за рассматриваемый период и носит поэтому условный характер.

Однако в ряде случаев путем интерполяции оказывается возможным определить неизвестные промежуточные уровни с практически приемлемой точностью. При прочих равных условиях интерполяция дает тем более точные результаты, чем меньше колеблется уровень явления и чем короче промежуток времени между периодами или датами, уровни которых известны.

Чаше всего при интерполяции исходят из того, что уровень ряда динамики изменяется равномерно, т.е. при сохранении либо равных абсолютных приростов, либо равных темпов роста за равные промежутки времени. Если есть основания считать, что сохраняются равные абсолютные приросты, то неизвестные уровни определяются на основе среднего абсолютного прироста:

$$y_i = y_1 + \overline{\Pi}(i-1),$$

где y_i – неизвестный уровень ряда с порядковым номером i;

y_1 – начальный уровень;

$(i-1)$ – длина периода, равная разности между порядковыми (или хронологическими) номерами уровней y_i и y_1;

$\overline{\Pi}$– средний абсолютный прирост за весь рассматриваемый период.

Интеллектуальный тренинг

1. Что такое ряд динамики?
2. С какой целью анализируются данные рядов динамики?
3. Какие существуют виды динамических рядов?
4. Какие показатели применяются для характеристики изменений уровней ряда динамики?
5. Какой вид средних величин используется для расчета среднего уровня моментного ряда динамики?
6. Приведите примеры моментных рядов динамики с абсолютными конкретными и абсолютными средними абстрактными уровнями. Приведите примеры интервальных рядов

динамики абсолютных величин, а также интервальных рядов, выраженных относительными величинами.

7. Как рассчитать средний темп роста и темп прироста уровней ряда динамики?

8. Назовите виды колебаний уровней временного ряда.

9. Как может быть выявлена основная тенденция в изменениях уровней ряда динамики?

10.Назовите преимущества и роль аналитического выравнивания уровней временного ряда.

11.Как выполнить прогноз на будущее с помощью уравнения тренда?

12.Какие факторы влияют на величину средней квадратической ошибки уравнения тренда?

13.Как рассчитать скользящую среднюю и для каких целей она может быть использована?

14.Какие методы можно использовать для выявления сезонных колебаний?

15.Как рассчитать индексы сезонности и осуществить экстраполяцию с учетом сезонной составляющей?

17. Какие методы экстраполяции применяются в статистическом прогнозировании и являются ли они достаточными для составления научно обоснованного прогноза?

18. Чем отличается экстраполяция от интерполяции?

19. Какая разница между механическим сглаживанием и аналитическим выравниванием ряда динамики?

20. Составьте таблицу перехода от одного показателя динамики к другому на примере четырех показателей – темпа роста, коэффициента роста, темпа прироста и коэффициента прироста.

21. Обычно рекомендуют, чтобы срок прогноза не превышал одной трети длительности базы расчета тренда. Почему?

22. В отличие от прогноза на основе регрессионного уравнения прогноз по тренду учитывает факторы развития только в неявном виде, и это не позволяет «проигрывать» разные варианты прогнозов при разных возможных значениях факторов, влияющих на изучаемый призрак. Зато прогноз по тренду

охватывает все факторы, в то время как в регрессионную модель невозможно включить в явном виде более 10-20 факторов в самом лучшем случае. Так ли это?

23. Сущность прогноза на основе тренда хорошо иллюстрируется следующим рассказом о греческом философе Диогене, жившем в большой бочке на берегу Саронического залива недалеко от афинского порта Пирея. Как-то вечером Диогена стал окликать снаружи неизвестный. Диоген вышел к нему. – «Скажи, мудрый человек», – спросил путник, – дойду ли я к закату в Афины?» Диоген посмотрел на него и сказал: – «Иди!» Путник повторил свой вопрос... – «Иди!» – закричал Диоген, и путник, пожав плечами, побрел по берегу. – «Вернись!» – снова закричал Диоген, и путник вернулся к нему. – «Вот теперь я тебе скажу, что до заката ты не дойдешь до Афин. Оставайся у меня». – «А почему же ты сразу мне это не сказал, а прогнал меня?» Диоген усмехнулся: – «А как же я скажу, дойдешь ли ты до Афин, если я не видел, как быстро ты ходишь?» Прогноз по тренду – это и есть Диогенов прогноз на основании знания того, как изучаемая система «шла» до настоящего времени.

* * *

Завершая этим признанием главу о статистическом анализе рядов динамики, дадим последние методологические советы изучающим статистику.

Всякая наука – это процесс продолжающегося познания природы и общества. Нет наук законченных, которые следует *лишь* выучить наизусть, чтобы все знать.

Учебники и учебные пособия – *лишь* сжатые и неполные изложения уже достигнутого наукой уровня познания. Изучайте специальную литературу, если хотите больше знать, а также новейшие достижения ученых всего мира.

Не считайте себя только «сосудами для вливания» *знаний.* Познав известное, вы тоже можете (и должны!) внести свой вклад в дальнейшее развитие теории статистики. «Если не я – то кто же?»

Использованная и рекомендуемая специальная литература

1. Андерсон Т. Статистический анализ временных рядов / Пер. с англ. М.: Мир, 1976.

2. Афанасьев В.Н. Статистическое обеспечение проблемы устойчивости сельскохозяйственного производства. М.: Финансы и статистика, 1996.

3. Вайну Я. Я.-Ф. Корреляция рядов динамики. М.: Статистика, 1977.

4. Казинец Л.С. Темпы роста и абсолютные приросты. М.: Статистика, 1975.

5. Чётыркин Е.М. Статистические методы прогнозирования. М.: Статистика, 1977.

6. Юзбашев М.М., Манелля А.И. Статистический анализ тенденций и колеблемости. М.: Финансы и статистика, 1983.

Глава X
ИНДЕКСЫ

Дорожная карта

10.1. Понятие индекса

Слово индекс *(index)* буквально означает показатель. Однако не всякий показатель в статистике называется индексом. **Индекс в статистике** – ***это, как правило,*** **обобщающий показатель сравнения общественно-экономических явлений, состоящих из элементов, непосредственно не поддающихся суммированию.**

Сравнительная характеристика такого рода явлений во времени или пространстве не может быть произведена путем сопоставления их абсолютных или средних уровней, для этого требуется применение особых приемов индексного метода.

С такими явлениями статистика имеет дело, например, тогда, когда возникает потребность дать сводную характеристику изменения объема продукции как массы материальных благ, или дать сводную характеристику изменения цен, себестоимости, затрат труда на единицу продукции или других аналогичных явлений.

Для решения этих задач исчисляются индексы объема произведенной (проданной или потребленной) продукции, индексы цен, себестоимости, производительности труда и другие индексы.

Отдельными элементами в индексах выступают конкретные продукты, которые непосредственно не поддаются суммированию потому, что они берутся в натурально-вещественной форме. Нельзя сложить тепловозы с иголками, так же как и сопоставить цены за килограмм хлеба и литр керосина.

Чтобы просуммировать подобного рода непосредственно несоизмеримые элементы, необходимо найти для них общую единицу измерения.

Общим для различных видов изделий является то, что они все есть продукты труда. В товарном производстве это общее находит свое выражение в стоимости товара. Стоимость разных товаров можно складывать. Поэтому, измеряя динамику стоимости разнородной продукции, можно сравнивать общие ее суммы за разные периоды времени. Но изменение этих общих сумм стоимостей обусловливается совместным изменением количества продукции и цен. Чтобы измерить, как изменился объем (количество) разнородной продукции, нужно принять для сравниваемых периодов одинаковые цены, а чтобы измерить, как изменился уровень цен по группе разнородной продукции, необходимо в индексе исключить изменение ее количества.

Индексный метод наиболее широко применяется для исчисления темпов динамики (сравнение во времени), а также для анализа выполнения плана. Однако в отдельных случаях он может быть использован и для сравнения такого же рода явлений за один и тот же период времени в пространстве, например, для сопоставления уровня цен или производительности труда в разных странах.

С помощью индексного метода производят также анализ изменения средних величин в том случае, если оно обусловлено изменением не только осередняемого признака в отдельных группах единиц совокупности, но и структуры самой совокупности. Индексный метод в этом случае помогает выявить роль каждого из этих факторов в изменении средней.

Таким образом, при помощи индексов:

1) определяются средние изменения сложных, непосредственно несоизмеримых совокупностей во времени;

2) оценивается средняя степень выполнения плана по совокупности в целом или ее части;

3) устанавливаются средние соотношения сложных явлений в пространстве;

4) определяется роль отдельных факторов в общем изменении сложных явлений во времени или в пространстве и, в частности, изучается влияние структурных сдвигов.

При решении первой задачи – изучений изменения явлений во времени – индексы выступают как показатели динамики, при решении второй – как показатели выполнения плана, третьей – как показатели сравнения, четвертой – как аналитическое средство.

Аналитическое направление в развитии индексной теории в значительной степени сформировалось благодаря работам российских статистиков В.Н. Старовского, Н.М. Виноградовой, Л.В. Некраша, И.Ю. Писарева, Г.И. Бакланова, Л.С. Казинец, Б.Г. Плошко, В.Е. Адамова, Г.В. Ковальского а так же зарубежных статистиков Р. Адлена, Н. Фишера и других.

В настоящее время индексный метод продолжает успешно развиваться в трудах как зарубежных, так и отечественных статистиков, и получил широкое применение в практике статистической работы.

Символика в индексном методе

Для удобства восприятия индексов в теории статистики разработана определенная символика. Так, каждый индекс включает два вида данных, оцениваемые данные, которые принято называть отчетными и обозначать значком «1», и данные, которые используются в качестве базы сравнения – базисные, обозначаемые знаком «0».

Индекс, который строится как сравнение обобщенных величин, называется **сводным** или **общим,** и обозначается I. Если же сравниваются необобщенные величины, то индекс называется **индивидуальным** и обозначается i. Как правило, подстрочно дается значок, который указывает, для оценки какой величины построен индекс. Например, $I_{q/0}$ или $i_{q/0}$, т.е. сводный и индивидуальный индексы для величины q.

Основной элемент индексного отношения – **индексируемая величина** – значение признака статистической совокупности, изменение которой является объектом изучения.

Вес индекса – это величина, служащая для целей соизмерения индексируемых величин.

При составлении индексов в статистике пользуются следующими условными знаками (табл.10.1).

Таблица 10.1

Условные обозначения применяемые при исчислении индексов

Показатели	Условные знаки	Периоды	
		базисный	отчетный
Количество	q	q_0	q_1
Цена	p	p_0	p_1
Себестоимость продукции	c	c_0	c_1
Трудоемкость	т	$т_0$	$т_1$
Общее количество затрат времени (чел.час, чел.день)	т	t_0	t_1
Производительность труда	w	w_0	w_1

10.2. Виды индексов

Различают следующие виды индексов:

1. В зависимости от степени охвата подвергнутых обобщению единиц изучаемой совокупности индексы подразделяются на:

- *индивидуальные* (элементарные);
- *групповые;*
- *общие.*

2. В зависимости от содержания индексируемой величины:

- *индексы количественных показателей;*
- *индексы качественных показателей.*

3. По характеру отношения:

- *динамические индексы* – сравнение во времени;
- *территориальные индексы* – сравнение в пространстве;
- *индексы сравнения фактических данных с плановыми* (договорными, нормативными, прогнозируемыми).

4. По способу сравнения:

- *базисные индексы,* когда база сравнения постоянная;
- *цепные индексы,* когда база сравнения переменная.

5. По виду весов:

- *индексы с постоянными весами* – индексы, вычисленные с весами, не меняющимися при переходе от одного индекса к другому;
- *индексы с переменными весами* – индексы, вычисленные с весами, меняющимися при переходе от одного индекса к другому.

6. По форме построения:

- *индивидуальные;*
- *агрегатные;*
- *средние.*

7. По объекту исследования:

- *производительность труда;*

* *себестоимость;*
* *объем продукции;*
* *зарплата и др.*

8. По составу явления:

* *постоянные;*
* *переменные.*

10.3. Индивидуальные индексы

Индивидуальными индексами называют показатели, характеризующие изменение отдельных величин сложного явления, являющиеся обычной относительной величиной. Так, например, индивидуальный индекс динамики – это отношение отчетного уровня индексируемого показателя к базисному уровню, т.е. темп роста, индивидуальный территориальный индекс – это отношение показателя одного предприятия или района к соответствующему показателю другого предприятия или района, т.е. относительная величина сравнения.

На практике в основном используются следующие индивидуальные индексы (табл.10.2).

Таблица 10.2

Индивидуальные индексы

№	Для однотипной продукции	Индивидуальные индексы
1	Индексы физического объема	$i = q_1 : q_0$
2	Индекс цен	$i_q = p_1 : p_0$
3	Индекс себестоимости	$i_c = c_1 : c_0$
4	Индекс затрат времени	$i_t = t_0 : t_1$
5	Индекс производительности труда	$i_w = w_1 : w_0$

Для индивидуальных индексов динамики сохраняет силу взаимосвязь цепных и базисных темпов роста: если базисные индексы исчислены по отношению к начальному уровню, то произведение нескольких последовательных цепных индексов равно базисному индексу за соответствующий период.

Методику анализа индивидуальных индексов покажем на примере агрегатных индексов.

10.4. Общие индексы

Сводные (общие) индексы

Изучаемые статистикой общественные явления обычно состоят из многих отдельных элементов. Так, валовой внутренний продукт представляет сумму валовой продукции секторов экономики как промышленности, сельского хозяйства, строительства и других. Продукция каждой отрасли состоит, в свою очередь, из ряда отдельных продуктов. Так, например, сельское хозяйство производит зерно, хлопок, молоко, шерсть и другие продукты. При этом отдельные продукты могут иметь свои разновидности и сорта (зерно пшеницы, ржи, кукурузы и т.д.). В соответствии с этим по степени охвата элементов изучаемой совокупности различают индивидуальные и сводные (общие) индексы.

Таким образом:

Индивидуальный индекс характеризует соотношение уровней отдельного более или менее однородного элемента сложного явления, а сводный, или общий, индекс соотношение уровней сложного явления, состоящего из нескольких отдельных элементов. Так, например, приведенный выше является индивидуальным, а индекс объемного производства – сводным. Индивидуальными являются также индексы, характеризующие изменение, цен отдельных товаров; индекс же, характеризующий изменение уровня цен нескольких товаров, является сводным индексом.

Сводные индексы делятся, в свою очередь, на *тотальные,* охватывающие всю совокупность и *групповые*, которые охватывают лишь часть ее элементов. Так, индекс промышленной продукции, является групповым по отношению к тотальному индексу совокупного общественного продукта. При изучении только промышленной продукции индексы объема продукции отдельных отраслей промышленности являются групповыми.

Важной особенностью общих индексов является то, что им присущи:

✓**синтетические свойства**: посредством индексного метода производится соединение (агрегирование) в целом разнородных единиц статистической совокупности;

✓**аналитические свойства**: посредством индексного метода определяется влияние факторов на изменение изучаемого показателя.

На основе изучения состава и роли факторов, выявления силы их действия осуществляются возможности квалифицированного управления развитием экономических процессов не только в нужном направлении, но и с заранее заданными параметрами.

Два метода построения сводных индексов

Сводный индекс, должен отражать соотношение уровней такого явления, которое состоит из нескольких отдельных элементов. Так, сводный индекс динамики физического объема продукции должен показать, как изменился общий объем различных видов продукции; сводный индекс цен должен дать общую характеристику изменения уровня цен на отдельные товары и т.п. Следовательно, задача заключается в том, чтобы найти общую меру соотношения отдельных элементов явления.

Возможны два пути решения этой задачи. Первый путь заключается в том, чтобы отыскать и затем сопоставить между собой *общие уровни* явления: общие объемы продукции, если нас интересует изменение ее количества, общие цены всех товаров, если нас интересует изменение цен и т.п. Второй путь заключается в осреднении индивидуальных индексов, характеризующих изменение отдельных элементов явления.

Сводные индексы, построенные первым методом, т.е. на основе сопоставления общих уровней явления, называются *агрегатными индексами*. Сводные индексы, построенные путем осреднения индивидуальных индексов, называются *средними индексами*.

Первый путь в ряде случаев непосредственно приводит к получению индекса. Это имеет место тогда, когда индексируемый показатель является объемным и непосредственно поддается суммированию (товарооборот, фонд заработной платы, посевная площадь, численность работников и т.п.). Так, сводный индекс динамики товарооборота выражается как отношение общей суммы товарооборота отчетного периода к общей сумме товарооборота базисного периода (табл.10.3).

$$I_{w=pq} = \frac{\Sigma w_1}{\Sigma w_0} = \frac{\sum p_1 q_1}{\sum p_0 q_0} = \frac{250450}{214700} = 1{,}167$$ или 116,7%,

где pq – товарооборот по отдельным товарам.

Агрегатный индекс товарооборота

Аналогично индексу товарооборота рассчитываются так же общие индексы затраты времени, продукции, потребления и т.д. Так, приведенная формула индекса товарооборота называется агрегатной (от латинского слова **aggrego – присоединяю**). Агрегатными называются индексы, числители и знаменатели которых представляют собой суммы, произведения или суммы произведений уровней изучаемого явления. Агрегатная форма индекса является основной, наиболее распространенной формой экономических индексов; она показывает относительное изменение изучаемого экономического явления и абсолютного размера этого изменения.

Построенный индекс товарооборота позволяет ответить нам на три вопроса:

➢ что показывает сам индекс;

➢ что показывает относительное отклонение, построенное на базе данного индекса;

➢ что показывает абсолютное отклонение, построенное на базе данного индекса.

Ответим последовательно на эти три вопроса:

• Агрегатный индекс объема продаж (товарооборота, стоимость) показывает, во сколько раз возрос объем продаж отчетного периода по сравнению с базисным (или сколько процентов составляет рост объема продаж):

если $I_{pq} > 1$ (или 100%) рост объема продаж;

если $I_{pq} < 1$ (или 100%) – снижение объема продаж;

если $I_{pq} = 1$ (или 100%) – изменений нет.

• Относительное отклонение, построенное на базе агрегатного индекса объема продаж, показывает, на сколько процентов (долей единицы) увеличился (уменьшился) объем продаж (товарооборота, стоимость) в отчетном периоде по сравнению с базисным:

если индекс рассчитан в процентах, то: $\Delta pq = I_{pq} - 100\%$;

если индекс рассчитан в коэффициентах, то: $\Delta pq = I_{pq} - I$.

• Абсолютное отклонение, построенное на базе агрегатного индекса объема продаж, показывает, на сколько сумов увеличился (уменьшился) объем продаж (товарооборота, стоимость) в отчетном периоде по сравнению с базисным:

$$\Delta pq_{абс} = \Sigma p_1 q_1 - \Sigma p_0 q_0 = 250\ 450 - 214\ 700 = 35750 \text{ тыс. сум.}$$

Абсолютное и относительное отклонения со знаком «+» означают прирост, а со знаком «–» – сокращение.

Таблица 10.3

Количество и цены проданных магазинам продуктов

Наименование продукта	Ед. изм	Базисный период		Отчетный период		Стоимость проданных продуктов, тыс. сум		
		кол-во q_0	цена, сум p_0	кол-во q_1	цена, сум p_1	q_0p_0	q_1p_1	q_1p_0
1	2	3	4	5	6	7	8	9
А	шт	20000	1,5	25000	1,0	30000	25000	37500
Б	кг	16500	2,0	18500	1,2	33000	22000	37000
В	кг	4850	22,0	6250	21,0	106700	131250	137500
Г	л	18000	2,5	24000	3,0	45000	72000	60000
Итого				1500		214700	250450	272000

Агрегатная формула индекса товарооборота показывает, что его величина зависит от двух явлений, от двух переменных величин: физического объема товарооборота, т.е. количества проданных товаров, и цены за каждую единицу реализованных товаров. Чтобы выявить влияние каждой переменной в отдельности, следует влияние одной из них исключить, т.е. принять ее условно в качестве постоянной, неизменной величины на уровне отчетного или базисного периода. Какой же период принять в качестве постоянной величины? В связи с этим возникает вопрос о базисных и отчетных весах агрегатного индекса. Рассмотрим этот вопрос на примере индекса цен и индекса физического объема товарооборота (табл.10.3).

Агрегатный индекс физического объема

Агрегатный индекс физического объема товарооборота – этот индекс должен показывать изменение физического объема в отчетном периоде по сравнению с базисным. Чтобы агрегатный индекс характеризовал только изменение физического объема товарооборота (продукции, потребления) и не отражал изменения цен, в качестве весов берутся неизменные цены как для базисного, так и для отчетного периодов. А неизменные цены, всегда являются ценами базисного периода. Применение в качестве весов неизменных цен дает возможность получить правильное представление о динамике физического объема товарооборота (продукции или потребления), так как устраняет влияние динамики цен на динамику количества выпущенной, проданной или потребленной продукции. Таким образом, в индексе физического объема сомножитель индексируемого показателя берется на уровне базисного периода.

Пользуясь принятыми обозначениями, запишем формулу агрегатного индекса физического объема продукции:

$$I_q = \frac{\sum q_1 p_1}{\sum q_0 p_1}$$

где числитель представляет собой стоимость продукции отчетного периода по ценам базисного, а знаменатель – стоимость продукции базисного периода по ценам того же периода. Подставив в $I_q = \frac{\sum q_1 p_1}{\sum q_0 p_1}$ необходимые данные из таблицы, получим $I_q = \frac{272000}{214700} = 1{,}267$ или 126,7%. Это значит, что в отчетном периоде по сравнению с базисным общий физический объем реализованной продукции увеличился на 26,7%. Абсолютное изменение физическою объема вычисляется как разность между числителем и знаменателем индекса. В нашем примере

$\Sigma q_1 p_0 - \Sigma q_0 p_0 = 272000 - 214700 = 57\,300$ тыс. сум,

т.е. в отчетном периоде по сравнению с базисным физический объем реализованной продукции увеличился в абсолютном выражении на 57 300 сум.

Построение агрегатных индексов с весами базисного периода первым предложил немецкий экономист Э. Ласпейрес (1864), и с тех пор этот метод носит его имя. Отметим, что при построении агрегатных индексов индексируемую величину принято писать сразу после знака суммы, а признак-вес – следующим. Построив агрегатный индекс физического объема, ответим на традиционные три вопроса (см. выше).

✓ **Индекс физического объема** показывает, во сколько раз возрос объем продаж (товарооборота, стоимость) в отчетном периоде по сравнению с базисным из-за изменения физического объема или сколько процентов составляет рост (снижение) объема продаж в результате изменения физического объема.

Допускается упрощенный вариант вывода: во столько-то раз в среднем возрос физический объем в отчетном периоде по сравнению с базисным.

✓ **Относительное отклонение**, построенное на базе-индекса физического объема, показывает, на сколько процентов (долей единицы) увеличился (уменьшился) объем продаж (товарооборота, стоимость) в отчетном периоде по сравнению с базисным из-за изменения физического объема.

Допускается упрощенный вариант вывода: на столько-то процентов (долей единицы) в среднем увеличился (уменьшился) физический объем в отчетном периоде по сравнению с базисным.

Если индекс рассчитан в процентах, то:

$$\Delta^{q}_{pq} = I_q - 100$$

или

$$\Delta^{q}_{pq} = I_q - 1.$$

✓ **Абсолютное отклонение,** построенное на базе индекса физического объема, показывает, на сколько сумов увеличился (уменьшился) объем продаж (товарооборота, стоимость) в отчетном периоде по сравнению с базисным из-за изменения физического объема.

$$\Delta^{q}_{pq} = \sum q_1 p_0 - \sum q_0 p_0.$$

Аналогично индексу физического объема рассчитывается индекс цен товарооборота.

Агрегатный индекс цен

Агрегатный индекс цен представляет собой дробь, числитель и знаменатель которой состоят из двух сомножителей. Один из них является переменной индексируемой величиной (p_1 и p_0), а второй принимается условно в качестве постоянной величины – веса индекса (q_1).

Тогда агрегатный индекс цен будет иметь следующий вид:

$$* I_p = \frac{\Sigma p_1 q_1}{\Sigma p_0 q_1} = \frac{250450}{272000} = 0{,}921 \text{ или } 92{,}1\%;$$

Агрегатный индекс цен с отчетными весами $I_p = 92{,}1\%$ означает, что цены на указанные товары в отчетном периоде снизились по сравнению с базисным на 7.9% (базисный период всегда принимается за 100%), а *абсолютная* фактическая *экономия от снижения цен* составила:

$\sum q_1 p_1 - \sum q_1 p_0 = 25\,0450 - 27\,2000 = -21550$ сум (табл. 10.3).

Построение агрегатных индексов с весами отчетного периода первым предложил немецкий экономист Г. Пааше (1874); и с тех пор этот метод носит его имя.

Ответим на поставленные выше три вопроса.

- Индекс цен показывает, во сколько раз возрос объем продаж (товарооборота, стоимость) в отчетном периоде по сравнению с базисным из-за изменения цен (или сколько процентов составляет рост объема продаж в результате изменения цен).

Допускается упрощенный вариант вывода: во столько-то раз в среднем возросли цены в отчетном периоде по сравнению с базисным.

- Относительное отклонение, построенное на базе индекса цен, показывает, на сколько процентов (долей единицы) увеличился (уменьшился) объем продаж (товарооборота, стоимость) в отчетном периоде по сравнению с базисным из-за изменения цен.

Допускается упрощенный вариант вывода: на столько-то процентов (долей единицы) в среднем выросли (снизились) цены в отчетном периоде по сравнению с базисным.

$$\Delta^{p}_{pq} = I_p - 100 \text{ или } \Delta^{p}_{pq} = I_p - 1.$$

В макроэкономике относительное отклонение цен, рассчитанное на базе динамических цепных индексов цен (в %), характеризует *темп инфляции.*

- Абсолютное отклонение, построенное на базе индекса цен, показывает, на сколько сумов увеличился (уменьшился) объем продаж (товарооборота, стоимость) в отчетном периоде по сравнению с базисным из-за изменения цен

$$\Delta^{p}_{pq\text{абс}} = \sum q_1p_1 - \sum q_1p_0.$$

Между рассмотренными индексами существует взаимосвязь

$$I_{pq} = I_pI_q = 0{,}921 \cdot 1{,}267 = 1{,}167 \text{ или } 116{,}7\%.$$

Абсолютные приросты, полученные на базе построенных индексов, также взаимосвязаны между собой:

$$\Delta_{pq\text{абс}} = \Delta^{q}_{pq\text{ абс}} + \Delta^{p}_{pq\text{ абс}} = 57300 - 21550 = 35750 \text{ тыс. сум.}$$

Аналогично строятся и другие агрегатные индексы с учетом первичности появления признаков и соответствующего ниже правила применения весов в индексе:

• для количественных индексируемых величин веса брать базисные (p_0, c_0 и т.д.);

• для качественных индексируемых величин веса следует брать отчетные (q_1).

Это правило основано на первичности построения показателей. Сначала, например, при планировании производства и реализации продукции появляется показатель «количество продукции» (количественный показатель), а затем уж – средняя цена единицы продукции (качественный показатель).

Помимо индексов товарооборота, цен и физического объема в статистико-экономическом анализе применяются и другие агрегатные индексы. Рассмотрим некоторые из них (табл.10.4).

Таблица 10.4

Индексы качественных показателей

Индекс	**Индексируемые величины**	**Индивидуальный индекс**	**Соизмерители**	**Агрегатная форма общего индекса**
Цен (p)	p_1 и p_0	$i_q = \frac{q_1}{q_0}$	q_1	$I_p = \frac{\Sigma p_1 q_1}{\Sigma p_0 q_1}$
Себестоимости (c)	c_1 и c_0	$i_c = \frac{C_1}{C_0}$	q_1	$J_c = \frac{\sum c_1 q_1}{\sum c_0 q_1}$
Производительность труда	t_1 и t_0	$i_t = \frac{t_0}{t_1}$	q_1	$J_t = \frac{\sum t_0 q_1}{\sum t_1 q_1}$

В табл. 10.4 в дополнение к уже рассмотренным выше категориям принята следующая символика:

c_1 и **c_0** – себестоимость единицы продукции в текущем и базисном периодах;

$\Sigma c_1 q_1$ и **$\Sigma c_0 q_1$** – фактические затраты на производство продукции в текущем периоде;

$\Sigma c_0 q_1$ – расчетные затраты на производство продукции в текущем периоде по себестоимости базисной;

t_0 и t_1 – затраты рабочего времени (труда) на производство единицы продукции данного вида (трудоемкость);

Σt_1q_1 – фактические затраты рабочего времени (труда) на производство продукции в текущем периоде;

Σt_0q_1 – расчетные затраты труда на производство продукции текущего периода по нормативам затрат базисного периода.

Агрегатный индекс себестоимости

Этот индекс строится аналогично индекса цен и показывает во сколько раз себестоимость в отчетном периоде в среднем выше или ниже базисной или плановой себестоимости, а также абсолютный размер экономии или перерасхода в результате изменения себестоимости. Индекс себестоимости продукции является индексом качественных показателей и исчисляется по весам (объему) продукции отчетного периода (табл.10.4, строка 2).

Покажем это на примере табл.10.5.

Таблица 10.5

Количество и себестоимость продукции

Наименование изделия	Себестоимость единицы изделия (тыс. сум)		Количество продукции (тыс. ед)	Затраты на производство (тыс. сум)	
	(c_0)	(c_1)		c_0q_1	c_1q_1
1	2	3	4	5	6
А	15,0	13,8	100	15000	13800
Б	5,0	4,8	200	10000	9600
В	4,0	4,0	50	2000	2000
Итого	–	–	–	27000	25400

Исчислим затраты на производство всей продукции в отчетном периоде: так сумма затрат в отчетном периоде (которую обозначим через Σc_1q_1 составила 25400 тыс. сум. Эта сумма в индексе себестоимости является исходной величиной. На ней базируются при определении изменений себестоимости в отчетном периоде по сравнению с базисным.

Для этого исчисляют другую сумму, характеризующую затраты на продукцию отчетного периода, которые могли бы быть, если бы себестоимость сохранилась на уровне базисного периода (табл. 10.5, гр. 5).

Сравнивая полученную сумму (которую обозначим через $\Sigma c_0 q_1$ с первоначальной, видим, что она больше ее на 160 тыс. сум. (27000-25400) потому что себестоимость изделий *«А»* и «*Б*» была в базисном периоде выше, чем в отчетном.

Эта разница показывает абсолютный размер экономии от снижения себестоимости, которую имело предприятие в отчетном периоде.

Экономия образовалась в результате снижения себестоимости изделия «А» (15000 – 13800 = 1200 тыс. сум) и изделия «Б» (10000 – 9600 = 400 тыс. сум).

Если же взять соотношение этих сумм, то оно и будет характеризовать снижение себестоимости на всю продукцию, т.е. даст искомый индекс себестоимости. Индекс себестоимости равен:

$$J_c = \frac{\sum c_1 q_1}{\sum c_0 q_1}$$ = 25400 : 27000= 0,941 или 94,1%. Это означает, что в целом себестоимость снизилась на 5,9%. Абсолютная фактическая экономия от снижения себестоимости составила:

$\Delta = \Sigma c_1 q_1 - \Sigma c_0 q_1 = 25400 - 27000 = -1600$ тыс. сум.

Агрегатный индекс производительности труда (трудоемкости)

Производительность труда определяется количеством продукции, произведенной в единицу времени, или затратами рабочего времени на производство единицу времени, или затратами рабочего времени на производство единицы продукции. Для определения изменения производительности труда в отчетном периоде по сравнению с базисным, надо затраты рабочего времени на производство единицы продукции в базисном периоде (t_0) поделить на затраты рабочего времени на производство единицы продукции в отчетном периоде (t_1). Таким образом, индивидуальный индекс производительности труда равен

$$i_t = \frac{t_0}{t_1}.$$

Чтобы построить агрегатный индекс производительности труда, необходимо затраты рабочего времени на производство единицы продукции взвесить на количество продукции, произведенной в отчетном периоде:

$$J_t = \frac{\sum t_0 q_1}{\sum t_1 q_1}$$

где $t_1 q_1$ – фактические затраты времени на производство всей продукции в отчетном периоде, а $t_0 q_1$ указывает сколько времени пришлось бы затратить на производство всей продукции отчетного периода в базисном периоде. В этом индексе в отличие от приведенных выше t_0 находится в числителе, а не в знаменателе, потому что время и производительность труда связаны обратной зависимостью: чем меньше затраты времени на производство единицы продукции, тем выше при прочих равных условиях производительность труда. Поэтому, чтобы при исчислении индекса получить прямой показатель производительности труда, t_0 записывается в числитель, а t_1– в знаменатель индекса.

Агрегатный индекс производительности труда, как индекс качественных показателей, также рассчитывается по весам (объему продукции) отчетного периода.

Агрегатный индекс выполнения плана. Особенность этого индекса заключается в том, что при его вычислении фактические данные сопоставляются не с базисными, а с плановыми, причем весами индекса могут быть показатели как плановые, так и фактические.

Постоянные и переменные веса агрегатных индексов

При исчислении индекса за два периода вопрос о весах сводится к выбору между базисными и отчетными периодами. На практике приходится иметь дело не только с двумя, но и большим числом периодов. Если индексы исчисляются за несколько периодов, то для всех них могут быть приняты одни и те же веса – индексы с постоянными весами, или же для каждого периода свои веса – индексы с переменными весами. Покажем это на примере (табл. 10.6).

Таблица 10.6

Количество и цены проданных товаров

Наименование товара	Продано товаров				Цена за единицу,тыс.сум			
	январь	февраль	март	... п	январь	февраль	март	... п
А, кг.	200	210	240	250	4,0	3,8	3,7	3,5
В, шт.	60	75	90	100	20,0	19,0	18,5	18,0

Требуется вычислить помесячные индексы. Их можно вычислить по-разному, в зависимости от решаемой задачи.

Теоретически возможны четыре типа индексов.

1. Общие базисные индексы цен с постоянными (базисными) весами (январскими):

$$I_{\frac{1}{0}}=\frac{\Sigma p_1q_0}{\Sigma p_0q_0};\ I_{\frac{1}{0}}=\frac{3{,}8\cdot200+19\cdot60}{4{,}0\cdot200+20\cdot60}=\frac{1900}{2000}=0{,}950;$$

$$I_{\frac{2}{0}}=\frac{\Sigma p_2q_0}{\Sigma p_0q_0};\ I_{\frac{2}{0}}=\frac{3{,}7\cdot200+18{,}5\cdot60}{4{,}0\cdot200+20\cdot60}=\frac{1850}{2000}=0{,}925;$$

и т.д.

$$I_{\frac{n}{0}}=\frac{\Sigma p_nq_0}{\Sigma p_0q_0};\ I_{\frac{n}{0}}=\frac{3{,}5\cdot200+18\cdot60}{4{,}0\cdot200+20\cdot60}=\frac{1780}{2000}=0{,}890;$$

В данных индексах цены каждого последующего периода (февраля – p_p марта – p_2 и т.д.) сопоставляются с ценами января (p_0) и взвешиваются на одно и то же количество товаров, проданных в январе (q_0). Полученные показатели характеризуют изменение цен по сравнению с начальным периодом, но не отражают изменения в структуре проданных товаров.

2. Общие *базисные* индексы цен с *переменными* (отчетными)

$$I_{\frac{1}{0}}=\frac{\Sigma p_1q_1}{\Sigma p_0q_1};\ I_{\frac{1}{0}}=\frac{3{,}8\cdot210+19\cdot75}{4{,}0\cdot210+20\cdot75}=\frac{2233}{2340}=0{,}950;$$

$$I_{\frac{2}{0}}=\frac{\Sigma p_2q_2}{\Sigma p_0q_2};\ I_{\frac{2}{0}}=\frac{3{,}7\cdot240+18{,}5\cdot90}{4{,}0\cdot240+20\cdot90}=\frac{2553}{2760}=0{,}925;$$

и т.д.

$$I_{\frac{n}{0}}=\frac{\Sigma p_n q_n}{\Sigma p_0 q_n};\ I_{\frac{n}{0}}=\frac{3{,}5\cdot 250+18\cdot 100}{4{,}0\cdot 250+20\cdot 100}=\frac{2675}{3000}=0{,}892;$$

В этих индексах цены каждого последующего периода (февраля – р., марта – p_2 и т.д.) сравниваются с ценами января (p_0), но в качестве весов берется каждый раз количество товаров отчетного периода (q_1, q_2 и т.д.).

В вычисленных индексах находят отражение как изменения цен по сравнению с начальным (базисным) периодом, так и изменения структуры проданных товаров.

3. Общие *цепные* индексы цен с *постоянными* весами (январьскими):

$$I_{\frac{1}{0}}=\frac{\Sigma p_1 q_0}{\Sigma p_0 q_0};\ I_{\frac{1}{0}}=\frac{3{,}8\cdot 200+19\cdot 60}{4{,}0\cdot 200+20\cdot 60}=\frac{1900}{2000}=0{,}950;$$

$$I_{\frac{2}{1}}=\frac{\Sigma p_2 q_0}{\Sigma p_1 q_0};\ I_{\frac{2}{1}}=\frac{3{,}7\cdot 200+18{,}5\cdot 60}{3{,}8\cdot 200+19\cdot 60}=\frac{1850}{1900}=0{,}974;$$

и т.д.

$$I_{\frac{n}{n-1}}=\frac{\Sigma p_n q_0}{\Sigma p_{n-1} q_0};\ I_{\frac{n}{n-1}}=\frac{3{,}5\cdot 200+18\cdot 60}{3{,}7\cdot 200+18{,}5\cdot 60}=\frac{1810}{1850}=0{,}980;$$

Эта группа индексов получена путем сопоставления цен каждого последующего периода с предыдущим, взвешенных на одно и то же количество товаров, проданных в январе (q_0). Эти индексы отражают изменение цен каждого периода по сравнению с предыдущим, но не отражают изменения в структуре проданных товаров.

4. Общие *цепные* индексы цен с *переменными* весами:

$$I_{\frac{1}{0}}=\frac{\Sigma p_1 q_1}{\Sigma p_0 q_1};\ I_{\frac{1}{0}}=\frac{3{,}8\cdot 210+19\cdot 75}{4{,}0\cdot 210+20\cdot 75}=\frac{2233}{2340}=0{,}950;$$

$$I_{\frac{2}{1}}=\frac{\Sigma p_2 q_2}{\Sigma p_1 q_2};\ I_{\frac{2}{1}}=\frac{3{,}7\cdot 240+18{,}5\cdot 90}{3{,}8\cdot 240+19\cdot 90}=\frac{2553}{2622}=0{,}974;$$

и т.д.

$$I_{\frac{n}{n-1}}=\frac{\Sigma p_n q_0}{\Sigma p_{n-1} q_n};\ I_{\frac{n}{n-1}}=\frac{3{,}5\cdot 200+18\cdot 100}{3{,}7\cdot 250+18{,}5\cdot 100}=\frac{2675}{2775}=0{,}964;$$

Эти индексы получены путем сопоставления цен каждого последующего периода с предыдущим, но взвешенных в каждом случае на количество товаров отчетного периода (q_1. q_2 и т.д.). В рассчитанных индексах находит отражение как изменение, цен за ряд последовательных периодов, так и изменение структуры проданных товаров. Индексы с переменными весами не дают возможности перехода от цепных индексов к базисным, и наоборот, так как веса их различны:

$$\frac{\Sigma p_1 q_1}{\Sigma p_0 q_1} \times \frac{\Sigma p_2 q_2}{\Sigma p_1 q_2} \neq \frac{\Sigma p_2 q_0}{\Sigma p_0 q_0}.$$

Индексы с постоянными весами допускают возможность перехода от цепных к базисным индексам, и наоборот. Перемножив два (или несколько) цепных индексов с постоянными весами, получим базисный индекс:

$$\frac{\Sigma p_1 q_0}{\Sigma p_0 q_0} \times \frac{\Sigma p_2 q_0}{\Sigma p_1 q_0} = \frac{\Sigma p_2 q_0}{\Sigma p_1 q_0}$$

а поделив два базисных индекса с постоянными весами, получим цепной:

$$\frac{\Sigma p_2 q_0}{\Sigma p_0 q_0} : \frac{\Sigma p_1 q_0}{\Sigma p_0 q_0} = \frac{\Sigma p_2 q_0}{\Sigma p_1 q_0}$$

Аналогично можно построить с постоянными и переменными весами индексы физического объема продукции и т.д.

В связи с разнообразием индексов возникает вопрос о выборе наиболее пригодного из них в каждом конкретном случае. Так, для характеристики изменения цен по сравнению с начальным периодом без учета изменений в структуре проданных (произведенных) товаров применяют общие базисные индексы с постоянными весами, в тех же целях, но с учетом изменения структуры – базисные индексы с переменными весами. Для определения изменения цен каждого периода по сравнению с предыдущим без учета изменений в структуре проданных товаров применяют цепные индексы с постоянными весами, с учетом изменений в структуре – цепные индексы с переменными весами. Выбор периода взвешивания индексов зависит от того, какие индексы вычисляются: индексы количественных (объемных) или качественных показателей. Как было отмечено выше в теории статистики принята следующая система взвешивания: сомножители количественных индексируемых показателей берутся на уровне базисного периода, а качественных – на уровне отчетного.

10.5. Средние индексы

Агрегатные индексы цен, физического объема товарооборота и др. могут быть вычислены при условии, если известны индексируемые величины и веса, т.е. p и q. Но в ряде случаев мы не располагаем необходимыми данными, а имеем произведение pq и индивидуальные индексы i. Возникает проблема построения средних индексов, идентичных агрегатным, путем осреднения индивидуальных индексов. Эта задача решается преобразованием агрегатного индекса в среднеарифметический и среднегармонический индексы (табл.10.7)

Таблица 10.7

Порядок преобразования агрегатных индексов в средние

Индекс	Индивидуальный индекс	Агрегатный индекс	Производные индивидуальных индексов	Средний индекс
Цен	$i_p = \frac{p_1}{p_0}$	$\frac{\sum p_1q_1}{\sum p_0q_1}$ $\frac{\sum p_1q_0}{\sum p_0q_0}$	$p_0 = \frac{1}{i_p}p_1$ $p_1 = i_p \cdot p_0$	$\frac{\sum p_1q_1}{\sum \frac{1}{i_p}p_1q_1}$ * $\frac{\sum i_p p_0 q_0}{\sum p_0q_0}$
Физического объема	$i_q = \frac{q_1}{q_0}$	$\frac{\sum q_1p_0}{\sum q_0p_0}$ $\frac{\sum q_1p_1}{\sum q_0p_1}$	$q_1 = i_q \cdot q_0$ $q_0 = \frac{q_1}{i_q}$	$\frac{\sum i_q q_0 p_0}{\sum q_0p_0}$ $\frac{\sum q_1p_0}{\sum \frac{1}{i_q}q_1p_0}$
Себестоимости	$i_c = \frac{c_1}{c_0}$	$\frac{\sum c_1q_1}{\sum c_0q_1}$ $\frac{\sum c_1q_0}{\sum c_0q_0}$	$c_0 = \frac{c_1}{i_c}$ $c_1 = i_c \cdot c_0$	$\frac{\sum c_1q_1}{\sum \frac{1}{i_c}c_1q_1}$ $\frac{\sum i_c c_0 q_0}{\sum c_0q_0}$
Производительности труда	$i_t = \frac{t_0}{t_1}$	$\frac{\sum t_0q_1}{\sum t_1q_1}$ $\frac{\sum t_0q_0}{\sum t_1q_0}$	$t_0 = i_t \cdot t_1$ $t_1 = \frac{t_0}{i_t}$	$\frac{\sum t_1q_1}{\sum i_t t_1 q_1}$ $\frac{\sum \frac{1}{i_t}t_0q_0}{\sum t_0q_0}$

Среднеарифметический индекс физического объема

Рассмотрим преобразование агрегатного индекса в среднеарифметический на примере агрегатного индекса физического объема товарооборота. В этом случае индивидуальные индексы должны быть взвешены на базисные соизмерители. Из индивидуального индекса физического объема товарооборота $i_q = \frac{q_1}{q_0}$ следует, что $q_1 = i_q \cdot q_0$. Заменив q_1 в числителе агрегатного индекса физического объема товарооборота $J_q = \frac{\sum q_1 p_0}{\sum q_0 p_0}$ на $i_q q_0$ получим

$$I_q = \frac{\sum i_q q_0 p_0}{\sum q_0 p_0}.$$

Это и есть среднеарифметический индекс физического объема товарооборота (табл.10.7, гр.5).

Среднегармонический индекс цен

В тех случаях, когда неизвестны отдельные значения p_1 и q_1, и дано их произведение $q_1 p_1$ – товарооборот отчетного периода, индивидуальные индексы цен $i_p = \frac{p_1}{p_0}$, а сводный индекс должен быть вычислен с отчетными весами, применяется *среднегармонический индекс цен.* Причем индивидуальные индексы должны быть взвешены таким образом, чтобы среднегармонический индекс совпал с агрегатным. Из формулы $i_p = \frac{p_1}{p_0}$ определим неизвестное значение p_0 и, заменив в формуле агрегатного индекса цен I $= \frac{\sum p_1 q_1}{\sum p_0 q_1}$ значение $p_0 = \frac{1}{i_p} p_1$, получим:

$$I = \frac{\sum p_1 q_1}{\sum \frac{p_1}{i_p} q_1} = \frac{\sum p_1 q_1}{\sum \frac{p_1 q_1}{i_p}}.$$

Индекс в такой форме называется среднегармоническим (табл.10.7, гр.5).

Пример. Определим общий индекс цен (по данным табл. 10.8 о продаже товаров в супермаркете «Корзина») по этой формуле.

В гр. 5 по формуле i_p определены индивидуальные (однотоварные) индексы цен:

$$i_p^A = \frac{100-4}{100} = 0{,}96; \quad i_p^B = \frac{100+10}{100} = 1{,}1.$$

В гр. 6 по каждому товару исчислены отношения стоимости продажи товаров в текущем периоде к индивидуальному индексу цен. Например, 185 : 0,96= 192,71 млн сум и т.д.

Итоговые данные гр.3 и гр.4 поставляются в формулу

$$I_p = \frac{\sum p_1 q_1}{\sum \frac{p_1 q_1}{i_p}} = \frac{475{,}0}{459{,}02} = 1{,}035 \quad \textit{или} \quad 103{,}5\%$$

т.е. по данному ассортименту в текущем периоде цены повышены в среднем на 3,5%.

Если в этой формуле из числителя вычесть значение знаменателя, то получают показатель прироста товарооборота в текущем периоде в результате изменения цен:

$$\Sigma\Delta qp(p) = \Sigma q_1 p_1 - \Sigma\frac{q_1 p_1}{i_p}$$

Таблица 10.8

Порядок расчета среднегармонического индекса цен

Товар	Продажи в ценах соответствующего периода (млн сум)		Изменение цен в текущем периоде по сравнению с базисным, %	Расчетные графы	
	базисный период q_0p_0	отчетный период q_1p_1		$i_p = \frac{p_1}{p_0}$	$\frac{q_1p_1}{i_p}$
1	2	3	4	5	6
А	153,5	185,0	– 4	0,96	192,71
Б	245,0	260,6	+ 10	1,1	236,91
В	21,5	29,4	без изменения	1,0	29,4
Итого	420,0	475,0	×	×	459,02

Для данных табл. 10.8 прирост товарооборота в текущем периоде в результате изменения цен составит: 475,0 – 459,02 = 15,98 млн сум, т.е. объем товарооборота возростет на 15,98 млн сум.

Полученное в итоге гр. 6 (табл. 10.8) значение

$$I_q = \frac{\Sigma q_1 p_0}{\Sigma q_0 p_0} = \frac{\Sigma \frac{q_1 p_1}{i_p}}{\Sigma q_0 p_0}.$$

может использоваться для определения общего индекса физического объема товарооборота в сопоставимых (базисных) ценах. Для этого на основе вышеприведенного тождества применяется преобразованная формула агрегатного индекса физического объема:

$$J_q = \frac{\sum i_q q_0 p_0}{\sum q_0 p_0} = \frac{\sum \frac{q_1}{q_0} \cdot q_0 p_0}{q_0 p_0} = \frac{\sum q_1 p_0}{\sum q_0 p_0}$$

При этом вычисляется ip = $p_1 : p_0$, т.е. индивидуальный индекс цен. Подставляя в формулу I_q итоговые данные гр. 2 и гр. 6 (табл. 10.8), вычисляется:

$$I_q = \frac{459{,}02}{420{,}0} = 1{,}093$$

т.е. физический объем продажи товаров, который увеличился в текущем периоде в среднем на 9,3%.

На основе этой же формулы исчисляется прирост суммы товарооборота в текущем периоде в результате изменения физического объема продажи товаров:

$$\Sigma \Delta qp(p) = \Sigma q_1 p_1 - \Sigma \frac{q_1 p_1}{i_p}$$

Подставляя в эту формулу соответствующие данные, получаем:

$$\Sigma \Delta qp(q) = 459{,}02 - 420 = 39{,}02 \text{ млн. сум.}$$

Таким образом, индексный анализ данных табл. 10.8 показывает, что снижение цен по ассортименту в целом в среднем на 3,5% вызвало увеличение товарооборота на 15,98 млн сум. Увеличение физического объема продажи товаров в среднем на 9,3% обусловило рост товарооборота на 39,02 млн сум. В результате совокупного действия этих факторов прирост объема товарооборота в текущих ценах составил 55 млн сум (39,02+15,98). Это соответствует расчету по формуле: ($\sum p_1q_0$: $\sum p_0q_0$);

$\sum\Delta qp\ (qp) = \sum q_0p_1 - \sum q_0p_0 = 4758 - 420 = 55{,}0$ млн сум,
т.е. в текущем периоде прирост товарооборота (в ценах соответствующих периодов) составил 55 млн сум.

Таким образом, анализ средневзвешенных индексов следует проводить подробно анализу агрегатных индексов.

10.6. Территориальные индексы

Построение территориальных индексов в агрегатной форме

При построении территориальных индексов, т.е. при сравнении показателей не во времени, а в пространстве, веса должны быть одинаковы для всех сравниваемых объектов или территорий. В индексах физического объема продукции в качестве таких сопоставимых весов могут быть использованы цены на 1 июля 2018 г. (для промышленной продукции) или сопоставимые закупочные цены (для сельскохозяйственной продукции). Если же, например, соотношение продукции двух районов или объектов интересует нас с точки зрения ее трудоемкости, то в качестве весов можно принять сопоставимую трудоемкость единицы отдельных видов продукции.

При этом возникает вопрос о том, какую именно трудоемкость следует принять в качестве сопоставимой, т.е. зафиксировать в числителе и знаменателе каждого индекса: трудоемкость того объекта или района, который сравнивается, трудоемкость того объекта (района), с которым производится сравнение, или же какую-либо иную трудоемкость. При решении этого вопроса нужно исходить из той экономической задачи, которая ставится перед данным индексом. Если соотношение продукции двух районов необходимо измерить с позиций только самих этих районов, то в качестве сопоставимой можно принять среднюю трудоемкость, исчисленную в целом по обоим сравниваемым районам.

Если же оценка соотношения продукции производится с более широких позиций (например, с позиций области, республики и т.п.), то, соответственно, нужно расширить и границы той территории, по которой исчисляется сопоставимая средняя трудоемкость. Таким образам, индекс физического объема продукции района *А* по отношению к району *Б* будет иметь следующий вид:

$$I_q = \frac{\sum q_A t_{iA}}{\sum q_Б t_{iБ}}$$

где буквами *А* и *Б* отмечены показатели сравниваемых районов или объектов, а *t* – сопоставимая трудоемкость, исчисленная в среднем по более или менее широкой территорий.

При построении индекса физического объема товарооборота колхозного рынка в качестве весов могут быть приняты средние цены отдельных товаров, исчисленные либо по обоим сравниваемым пунктам, либо по более широкой территории:

$$I_q = \frac{\sum q_A \overline{p}}{\sum q_Б \overline{p}_i}$$

где *p* – средние цены.

Пусть, например, по следующим данным нужно определить соотношение количества товаров, проданных в пунктах *А* и В (табл.10.9).

Таблица 10.9

Количество и цены товаров, проданных на дехканских рынках

Товары	Пункт А		Пункт В	
	продано (т)	цена за 1 кг (сум)	продано (т)	цена за 1 кг (сум)
	q_A	p_A	q_B	p_B
Картофель	480	2500	520	2000
Капуста	300	1500	200	2400

Определим среднюю цену каждого товара по формуле:

Средняя цена картофеля: $$\overline{P}_A = \frac{2500 \cdot 480 + 2000 \cdot 1520}{480 + 520} = 2240\, сум$$

Средняя цена капусты составляет:

$$\overline{P}_A = \frac{1500 \cdot 300 + 2400 \cdot 200}{300 + 200} = 1860\, сум$$

Используя эти средние цены в качестве весов, найдем индекс физического объема товарооборота:

$$I_q = \frac{\sum q_A \overline{P}}{\sum q_Б \overline{P}} = \frac{480 \cdot 2240 + 300 \cdot 1860}{520 \cdot 2240 + 200 \cdot 1860} = \frac{1633200}{1536800} = 1{,}063 \quad \textit{ёки} \quad 106{,}3\%.$$

Следовательно, в пункте *A* было продано товаров на 6,3% больше, чем в пункте *B*.

При построении, территориальных индексов качественных показателей также возникает вопрос о том, веса какого района или объекта следует зафиксировать в индексе. В индексах динамики, когда за базу сравнения принимается какой-либо предыдущий период, веса обычно фиксируются на уровне сравниваемого с ним отчетного периода. При построении же территориальных индексов каждый район с одинаковым основанием может быть принят как в качестве сравниваемого, так и в качестве базы сравнения, так что вопрос о фиксировании весов того или иного района остается открытым.

Допустим, что по данным нашего примера нужно определить, в каком пункте и на сколько процентов ниже уровень цен. Если в индексе фиксировать веса того пункта, который сравнивается с другим, то можно построить и исчислить два индекса:

$$I_{\overline{P}_T} = \frac{\sum p_A q_A}{\sum p_Б q_A} = \frac{2500 \cdot 400 + 1500 \cdot 300}{200 \cdot 480 + 2400 \cdot 300} = \frac{1650000}{1680000} = 0{,}982 \quad \textit{или} \quad 98{,}2\%.$$

Это означает, что применительно к составу товаров, проданных в пункте А, уровень цен в этом пункте на 1,8% ниже чем в пункте В. Однако, сопоставляя цены пункта В с ценами пункта А, получим:

$$I_{\overline{P}_T} = \frac{\sum p_Б q_Б}{\sum p_A q_Б} = \frac{2000 \cdot 520 + 2400 \cdot 200}{2500 \cdot 520 + 1500 \cdot 200} = \frac{1520000}{1600000} = 0{,}95 \quad \textit{или} \quad 95{,}0\%.$$

т.е. применительно к составу товаров, проданных в пункте *В*, цены в этом пункте ниже, чем в пункте *А*, на 5%.

Таким образом, в каждом пункте цены оказываются более низкими, чем в другом, если их уровень измерять применительно к составу товаров в сравниваемом пункте.

Чтобы на поставленный вопрос получить однозначный ответ, возможны различные пути. Наиболее простой из них заключается в том, что в качестве весов принимаются количества товаров, проданных в обоих пунктах вместе: $Q = q_A + q_B$. Тогда индекс цен пункта *А* по отношению к ценам пункта *В* будет иметь такой вид:

$$I_P = \frac{\Sigma P_A Q}{\Sigma P_з Q};$$

В нашем примере:

$$I_P = \frac{\Sigma P_A Q}{\Sigma P_з Q} = 2500x000 + 1500x00 / 2000x1000 + 2000x500 = 325;320 = 1{,}016$$

Следовательно, применительно к общему количеству товаров цены в пункте *А* на 1,6% выше, чем в пункте *В*. Этому не противоречит и обратный индекс $J_{\bar{p}_Т} = \frac{\sum p_Б q_Б}{\sum p_А q_Б} = \frac{3200000}{3250000} = 0{,}985$

Таким образом, при построении территориальных индексов качественных показателей в качестве весов могут быть приняты соответствующие объемные показатели в целом по обоим сравниваемым районам (либо по более широкой территории, если сопоставление производить с более широких позиций).

10.7. Индексный метод в факторном анализе средних величин

Индексы переменного и фиксированного состава

В ряде случаев приходится изучать динамику общественных явлений, уровни которых выражены средними величинами (средней себестоимостью, средней заработной платой, средней урожайностью, продуктивностью животных, средней производительностью труда и т.д.).

Динамика *средних* показателей зависит от одновременного изменения вариантов, из которых формируется средние и изменения удельных весов этих вариантов, т.е. от структуры изучаемого явления. Так, например, средняя производительность труда на предприятии может возрасти за счет ее повышения у рабочих отдельных специальностей и повышения удельного веса рабочих с более высокой производительностью труда в общей численности рабочих. Динамика средних надоев молока по региону зависит от динамики средних надоев в каждом хозяйстве, повышения удельного веса хозяйств с более высокими надоями в общем числе хозяйств.

Таким образом, на изменение динамики среднего значения изучаемого явления могут оказывать влияние одновременно два фактора: изменение осредняемого показателя и изменение структуры.

В этой связи возникает задача – показать роль каждого из этих факторов в общей динамике средней. Задача эта разрешается индексным методом, т.е. построением системы взаимосвязанных индексов, в которой динамика среднего показателя (будем называть его *индексом переменного состава*) выступает как произведение двух индексов: индекса показателя в неизменной структуре (будем называть его *индексом постоянного состава)* и индекса влияния изменения структуры явлений на динамику среднего показателя (будем называть его *индексом изменения структуры*).

Характерным для этой сферы применения индексов является непосредственная связь их с методом группировок. Чтобы исчислить индекс постоянного состава и индекс изменения структуры, совокупность предварительно должна быть разделена на группы по интересующему нас признаку и исчислены групповые средние.

В рассматриваемой системе взаимосвязанных индексов, в соответствии с изложенными ранее принципами, в качестве веса индекса постоянного состава должна быть взята структура отчетного периода, а индекса изменения структуры – уровень показателя в базисном периоде.

Если обозначить через x групповые средние и через f численности единиц по группам, то система взаимосвязанных индексов для анализа динамики средней *(x)* будет иметь такой вид:

$$I_{\bar{x}} = \underbrace{\frac{\sum x_1 f_1}{\sum f_1} : \frac{\sum x_0 f_0}{\sum f_0}}_{\substack{\textit{Индекс переменного} \\ \textit{состава}}} = \underbrace{\left(\frac{\sum x_1 f_1}{\sum f_1} : \frac{\sum x_0 f_1}{\sum f_1}\right)}_{\substack{\textit{Индекс постоянного} \\ \textit{состава}}} \cdot \underbrace{\left(\frac{\sum x_0 f_1}{\sum f_1} : \frac{\sum x_0 f_0}{\sum f_0}\right)}_{\substack{\textit{Индекс структурных} \\ \textit{сдвигов}}}$$

Рассмотрим применение такой системы индексов на примере анализа средней себестоимости какой-либо одноименной продукции.

Допустим, что себестоимость единицы одноименной продукции на двух заводах была различной. В отчетном периоде по сравнению с прошлым произошло изменение себестоимости на каждом заводе, а также изменилось соотношение заводов в общем выпуске данной продукции, что видно из следующей таблицы.

Таблица 10.10

Количество выработки изделий и себестоимость единицы одного вида продукции

	Произведено продукции (тыс. единиц)		**Удельные веса заводов в общем выпуске продукции (%)**		**Себестоимость единицы продукции, (тыс. сум)**		**Индексы себестоимости**
	базисный период	**отчетный период**	**базисный период**	**отчетный период**	**базисный период**	**отчетный период**	$i_c = c_1 : c_0$
А	**1**	**2**	**3**	**4**	**5**	**6**	**7**
	q_0	q_1	d_0	d_1	c_0	c_1	i_c
1	400	300	80	50	6	5,7	0,95
2	100	300	20	50	5	4,5	0,90
Итого	500	600	100	100	–	–	–

Как видно из таблицы, себестоимость на первом заводе снизилась на 5%, а на втором заводе – на 10%. Однако первый завод имеет более высокую себестоимость, поэтому производство данной продукции было сокращено на нем в отчетном году на 25%. В то же время на втором заводе производство этой продукции было расширено втрое и в результате общий объем выпуска продукции увеличился на 100 тыс. единиц, а удельный вес второго завода возрос с 20 до 50%. Естественно, что это должно было отразиться на средней себестоимости данной продукции по двум заводам вместе. И, действительно, средняя себестоимость составила:

прошлый год: $\overline{C}_0 = \dfrac{\sum q_0 c_0}{\sum q_0} = \dfrac{6 \cdot 400 + 5 \cdot 100}{500} = \dfrac{2900}{500} = 5{,}8$ *тыс. сум.*

отчетный год: $\overline{C}_1 = \dfrac{\sum c_1 q_1}{\sum q_1} = \dfrac{5{,}7 \cdot 300 + 4{,}5 \cdot 300}{600} = \dfrac{3060}{600} = 5{,}1$ *тыс сум*

Сопоставив среднюю себестоимость отчетного года со средней, себестоимостью базисного года, получим:

$i_c = C_1 : C_0 = 5{,}1 : 5{,}8 = 0{,}879$, или 87,9%.

Следовательно, средняя себестоимость снизилась на 12,1% (или на 0,7 тыс. сум на единицу продукции, что для всей продукции отчетного года дает экономию в 420 тыс. сум). Как видим, снижение средней себестоимости превышает снижение себестоимости на каждом отдельном заводе 5 и 10%.

Этот статистический парадокс, когда изменение средней выходит за пределы изменения частных величин, объясняется тем, что на индекс переменного состава (индекс средний) оказывает влияние также на изменение структуры совокупности – удельных весов заводов в производстве данной продукции.

Исчислим индекс себестоимости постоянного состава, т.е. индекс в неизменной структуре производства отчетного года:

$$J_{пос.\,с.} = \frac{\Sigma c_1 q_1}{\Sigma q_1} : \frac{\Sigma c_0 q_1}{\Sigma q_1} = \frac{5{,}7 \cdot 300 + 4{,}5 \cdot 300}{300 + 300} : \frac{6 \cdot 300 + 5 \cdot 300}{300 + 300} =$$

$$= 5{,}1 \text{ тыс. сум} : 5{,}5 \text{ тыс. сум} = 0{,}927 \text{ или } 92{,}7\%.$$

Таким образом, индекс постоянного состава действительно показывает среднее изменение себестоимости на обоих заводах. За счет снижения себестоимости на отдельных заводах среднее ее снижение составило 7,3%, что дало экономию в отчетном году в 240 тыс.сум.

$$Э = \left(\frac{\Sigma c_1 q_1}{\Sigma q_1} - \frac{\Sigma c_0 q_1}{\Sigma q_1} \right) \cdot \Sigma q_1 = (5{,}5 - 5{,}1) \cdot 600 = 240$$

тыс. сум.

Исчислим влияние изменения структуры производства на динамику средней себестоимости. Для этого возьмем соотношение средних уровней базисной себестоимости, рассчитанных, соответственно, с весами отчетного и базисного годов:

$$I_{с.с.} = \frac{\sum c_0 q_1}{\sum q_1} : \frac{\sum c_0 q_0}{\sum q_0} = \frac{6 \cdot 300 - 5 \cdot 300}{300 + 300} : \frac{6 \cdot 400 + 5 \cdot 100}{400 + 100} = 5{,}5 : 5{,}8 = 0{,}948,$$

или 94,8%.

Расчет показывает, что структурный сдвиг в производстве продукции дополнительно снизил среднюю себестоимость на 5,2%. Благодаря изменению структуры производства мы имеем экономию в сумме 180 тыс. сум.

$$Э = \left(\frac{\Sigma c_0 q_1}{\Sigma q_1} - \frac{\Sigma c_0 q_0}{\Sigma q_0} \right) \cdot \Sigma q_1 = (5{,}5 - 5{,}8) \cdot 600 = 180$$

тыс. сум.

Следовательно, общее снижение средней себестоимости на 12,1 % обусловлено на 7,3% снижением себестоимости на отдельных заводах и на 5,2% структурным сдвигом (0,879 – 0,927 × 0,948).

В общем виде эта зависимость записывается так:

$$j_{\bar{c}} = J_c \cdot J_{стр}$$

Следовательно, $j_{\bar{c}} = 0{,}879$

$$J_c = 0{,}927,$$

$$I_{стр} = 0{,}948,$$

а общая экономия – 420 тыс. сум от снижения себестоимости в отчетном году образовалась в результате достижений отдельных заводов (240 тыс. сум) и в результате изменения структуры производства (180 тыс. сум)

$$\sum Э = Э_с + Э_{стр.} = 240 + 180 = 420 \text{ тыс. сум.}$$

В указанной системе взаимосвязанных индексов при построении индекса фиксированного состава в качестве весов принята структура отчетного периода, что позволяет нам проследить изменение средней динамики изучаемого явления только за счет изменения осредняемых значений качественного показателя. При построении индекса структурных сдвигов в качестве соизмерителя принята величина осредняемого показателя на уровне базисного периода, что дает возможность изучить изменение средней динамики явления только за счет структурных сдвигов.

10.8. Об индексных соотношениях

Большое значение в экономической практике имеют соотношения в изменениях показателей, т.е. соотношения между величинами их индексов. Например, известно, что в эффективной экономике темпы роста производительности труда должны опережать темпы роста заработной платы:

$$I_{\text{производительности труда}} \geq I_{\text{заработной платы}}.$$

Или же – для развития предприятий оптимально следующее соотношение динамики основных показателей.

$$I_{\text{балансовой прибыли}} \geq I_{\text{реализации}} \geq I_{\text{авансированного капитала}} \geq 100\ \%.$$

Такого рода соотношение принято называть «экономической нормалью» или «динамическим нормативом».

Сравнение с нормалью используется в аудиторской деятельности для заключения о финансовом положении предприятия, его потенциале. Например, для трудоемкого производства в качестве нормали формулируется следующее неравенство:

$$I_{\substack{\text{объем}\\ \text{реализации}}} \geq I_{\substack{\text{материальные}\\ \text{затраты}}} \geq I_{\substack{\text{численность}\\ \text{промышленно-}\\ \text{производственного}\\ \text{персонала}}} \geq I_{\substack{\text{средняя стоимость}\\ \text{основных}\\ \text{производственных}\\ \text{фондов}}}.$$

Чтобы определить соответствие фактической динамики нормали, нужно иметь данные об изменении показателей (индексы) за несколько периодов. Например, оказалось, что поквартальные индексы за два года показывают следующее (табл. 10.11).

Только в трех кварталах соотношение в изменении показателей было близко к нормали. Аудитор обязан указать на это в своем заключении и рекомендовать менеджерам обратить внимание на причины: высокие цены поставщиков, избыточная численность персонала, неэффективная структура и использование основных фондов и т.д.

Таблица 10.11

Поквартальные индексы соотношений

Год, квартал / Индексы	1-й год							
	1	**2**	**3**	**4**	**1**	**2**	**3**	**4**
I объема реализации	x	0,994	0,985	1,041	1,055	0,990	1,002	1,036
I материальных затрат на производство	x	0,995	0,967	1,096	1,007	0,960	0,983	1,056
I численности промышленно производственного персонала	x	0,972	0,995	1,061	1,000	0,989	1,009	0,996
I стоимости основных производственных фондов	x	1,001	0,999	1,001	1,001	1,000	1,002	1,006

Динамика соответствующая экономической нормали, обычно определяет стратегию развития предприятий и для управления компанией (фирмой) важно проводить сравнение фактического соотношения темпов изменения показателей с «нормальным», выявлять, в каком звене нормали возникли *нарушения и* вносить коррективы в деятельность предприятия.

Интеллектуальный тренинг

1. Какова роль индексного метода анализа в экономических исследованиях?

2. На каких принципах базируется расчет агрегатных индексов объемных и качественных показателей?

3. В чем состоит различие агрегатных индексов Ласпейреса и Пааше и какие факторы оказывают влияние на расхождение в величине этих индексов?

4. Какие виды средних индексов используются в статистической практике и для решения каких проблем?

5. Запишите формулу «идеального» индекса Фишера. Какой вид средних величин используется для его расчета?

6. Если производство изделий в натуральном выражении снизилось в 1,2 раза по сравнению с прошлым годом, а цены на это изделие возросли за этот период в 2 раза, на сколько процентов изменилась стоимость продукции в сравнении с прошлым годом?

7. При каких условиях может быть осуществлен переход от базисных агрегатных индексов физического объема продукции к цепным агрегатным индексам физического объема?

8. Какой вариант агрегатных индексов качественных показателей используют при расчете индекса потребительских цен и почему?

9. Чем объяснить различия в величине индекса цен переменного и фиксированного состава?

10. Что характеризует индекс влияния структурных сдвигов? Напишите формулу для его расчета.

11. Назовите виды индексов качественных показателей.

12. Какие допущения (правила) лежат в основе использования индексов в экономическом анализе?

13. Что характеризует разность числителя и знаменателя агрегатных индексов физического объема продукции и цен?

14. Как определить долю влияния различных факторов на изменение результативного показателя?

15. Какие методические приемы используют при расчете индексов фондового рынка?

16. Сформулируйте основные принципы оценки абсолютного и относительного размера влияния факторов на изменение результативного показателя с использованием многофакторных индексных моделей.

17. Какое значение имеет построение факторных индексных моделей?

18. Как измерить уровень инфляции?

19. Какая информация необходима для расчета индекса потребительских цен?

20. Какие информации индексов используются при территориальных сопоставлениях?

21. Какое значение имеет соблюдение соотношений индексов?

Использованная и рекомендуемая специальная литература

1. Адамов В.Е. Факторный индексный анализ. Методология и проблемы. М.: Статистика, 1977.

2. Алан Р. Экономические индексы. М.: Статистика, 1980.

3. Андриенко В.Е. Статистические индексы в экономических исследованиях. Киев: Наукова думка, 1983.

4. Бакланов Г.И. Некоторые вопросы индексного метода. М.: Статистика, 1972.

5. Виноградов Н.М. Теория индексов. М.: Гостехиздат, 1930.

6. Ефимова М.Р., Ипатова И.М. Индексы и их применение в статистико-экономических исследованиях. М.: МИУ, 1974.

7. Казинец Л.С. Теория индексов. М.: Госстатиздат, 1963.

8. Кёвеш П. Теория индексов и практика экономического анализа. М.: Финансы и статистика, 1990.

9. Ковалевский Г.В. Индексный метод в экономике. М.: Финансы и статистика, 1989.

10. Кузуб Н.Л. Использование индексного метода в экономическом анализе. Донецк: ДГУ, 1972.

11. Костюхин В.Н. Индексы. М.: МФИ, 1960.

12. Курышева С.В. Индексы как метод анализа по факторам. Л.: ЛФЭН, 1982.

13. Маслов П.П. Советские индексы. М.: МФИ, 1953.

14. Мересте У. Очерки по индексной теории. Таллин: ТПИ, 1969.

15. Новожилов Е.М. Индексный метод анализа в торговле. М.: Статистика, 1974.

16. Плошко Б.Г. Индексы. Л.: ЛГУ, 1958.

17. Перегудов В.Н. Теоретические вопросы индексного анализа. М.: Госстатиздат, 1960.

18. Торвей Р. Индексы потребительских цен. Методология и руководство. М.: Финансы и статистика, 1983.

19. Фишер И. Построение индексов. М., 1982.

20. Щицман С. Индексный или разностный метод факторного анализа //Вестник статистики. 1990. №10.

ПРИЛОЖЕНИЯ

Приложение 1

Интегральная функция нормального распределения

$$F(t)=\frac{2}{\sqrt{2\pi}}\int_0^t e^{-\frac{t^2}{2}}dt$$

t	F(t)	t	F(t)	t	F(t)	t	F(t)
0,0	0,0000	1,0	0,6827	2,0	0,9545	3,0	0,9973
0,1	0,0797	1,1	0,7287	2,1	0,9643	3,1	0,9981
0,2	0,1585	1,2	0,7699	2,2	0,9722	3,2	0,9986
0,3	0,2358	1,3	0,8064	2,3	0,9786	3,3	0,9990
0,4	0,3108	1,4	0,8385	2,4	0,9836	3,4	0,9993
0,5	0,3829	1,5	0,8664	2,5	0,9876	3,5	0,9995
0,6	0,4515	1,6	0,8904	2,6	0,9907	3,6	0,9997
0,7	0,5161	1,7	0,9109	2,7	0,9931	3,7	0,9998
0,8	0,5763	1,8	0,9281	2,8	0,9949	3,8	0,9999
0,9	0,6319	1,9	0,9426	2,9	0,9963	3,9	0,9999

Приложение 2

Значения критерия t-Стьюдента при уровне значимости p (вероятности) **0,10, 0,05, 0,02 и 0,01 и числе степеней свободы и вариации** (V)

V \ p	**0,10**	**0,05**	**0,02**	**0,01**
1	6,314	12,706	31,821	63,657
2	2,920	4,303	6,965	9,925
3	2,353	3,182	4,541	5,841
4	2,132	2,776	3,747	4,604
5	2,015	2,571	3,365	4,032
6	1,943	2,447	3,143	3,707
7	1,895	2,365	2,998	3,499
8	1,860	2,306	2,896	3,355
9	1,833	2,262	2,821	3,250
10	1,812	2,228	2,764	3,169
11	1,796	2,201	2,718	3,106
12	1,782	2,179	2,681	3,055
13	1,771	2,160	2,650	3,012
14	1,761	2,145	2,624	2,977
15	1,753	2,131	2,602	2,947
16	1,746	2,120.	2,583	2,921
17	1,740	2,110	2,567	2,898
18	1,734	2,101	2,552	2,878
19	1,729	2,093	2,539	2,861
20	1,725	2,086	2,528	2,845
21	1,721	2,080	2,518	2,831
22	1,717	2,074	2,508	2,819
23	1,714	2,069	2,500	2,807
24	1,711	2,064	2,492	2,797
25	1,708	2,060	2,485	2,787

Укрупненная таблица расчета средних темпов роста

$$\overline{T} = \sqrt[n]{Np_{1/0}Tp_{2/1}...Tp_{n/n-1}} = \sqrt[m-1]{\frac{y_n}{y_0}} - \sqrt[m-1]{Tp_{n/0}}$$

Средний темп	**Коэффициенты**								
	$\sqrt{\ }$	$\sqrt[3]{\ }$	$\sqrt[4]{\ }$	$\sqrt[5]{\ }$	$\sqrt[6]{\ }$	$\sqrt[7]{\ }$	$\sqrt[8]{\ }$	$\sqrt[9]{\ }$	$\sqrt[10]{\ }$
0,75	0,562	0,422							
0,900	0,810	0,729	0,656	0,590	0,531	0,478			
0,910	0,828	0,754	0,686	0,624	0,568	0,517	0,470	0,428	0,389
0,920	0,846	0,779	0,716	0,659	0,606	0,558	0,513	0,472	0,434
0,930	0,865	0,804	0,748	0,696	0,647	0,602	0,560	0,520	0,484
0,940	0,884	0,831	0,781	0,734	0,690	0,648	0,610	0,573	0,539
0,950	0,902	0,857	0,814	0,774	0,735	0,698	,0,663	0,630	0,599
0,960	0,922	0,885	0,849	0,815	0,783	0,751	0,721	0,692	0,665
0,970	0,941	0,913	0,885	0,859	0,833	0,808	0,784	0,760	0,737
0,980	0,960	0,941	0,922	0,904	0,886	0,868	0,861	0,834	0,817
0,990	0,980	0,970	0,961	0,951	0,941	0,932	0,923	0,913	0,904
1,001	1,002	1,003	1,004	1,005	1,006	1,007	1,008	1,009	1,010
1,010	1,0200	1,0300	1,0400	1,0500	1,0605	1,0710	1,0820	1,0930	1,1040
1,020	1,0404	1,0612	1,0824	1,1040	1,1261	1,1486	1,1716	1,1950	1,2190
1,030	1,0609	1,0927	1,1255	1,1593	1,1941	1,2290	1,2668	1,3048	1,3439
1,040	1,0820	1,1253	1,1703	1,2171	1,2658	1,3164	1,3691	1,4239	1,4809
1,050	1,1025	1,1576	1,2155	1,2763	1,3401	1,4071	1,4775	1,5514	1,6290
1,060	1,1236	1,1910	1,2625	1,3383	1,4186	1,5037	1,5939	1,6895	1,7909
1,070	1,1449	1,2250	1,3108	1,4026	1,5008	1,6059	1,7183	1,8386	1,9673
1 ,080	1,1664	1,2597	1,3605	1,4693	1,5868	1,7137	1,8508	1,9989	2,1588
1 090	1,1882	1,2950	1,4116	1,5386	1,6771	1,8280	1,9925	2,1718	2,3673
1,100	1,2100	1,3310	1,4641	1,6105	1,7716	1,9488	2,1437	2,3581	2,5939
1,110	1,2321	1,3676	1,5180	1,6850	1,8704	2,0761	2,3045	2,5580	2,5394
1,120	1,2544	1,4049	1,5735	1,7623	1,9738	2,2107	2,4760	2,7831	3,1059
1,130	1,2769	1,4429	1,6305	1,8425	2,0820	2,3527-	2,6586	3,0042	3,3947
1,140	1,3996	1,4815	1,6889	1,9253	2,1938	2,5021,	2,8524	3,2517	3,7069
1,150	1,3225	1,5209	1,7490	1,0114	2,3131	2,6601	3,0591	3,5180	4,0457
1,160	1,3456	1,5609	1,8106	2,1003	2,4363	2,8261	3,2783	3,8028	4,4112
1,170	1,3689	1,6016	1,8739	2,1925	2,5652	3,0013	3,5115	4,1085	4,8069
1,180	1,3924	1,6430	1,9387	2,2877	2,6995	3,1854	3,7588	4,4354	5,2338
1,190	1,4161	1,6852	2,0054	2,3864	2,8398	3,3794	4,0215	4,7856	5,6949
1,200	1,4400	1,7280	1,7280	2,4883	2,9860	3,5832	4,2998	5,1598	6,1918
1,210	1,4641	1,7716	1,7716	2,5938	3,1385	3,7976	4,5951	5,5601	6,7277
1,220	1,4884	1,8158	1,8158	2,7027	3,2973	4,0227	4,9077	5,9874	7,3046
1,230	1,5129	1,8609	1,8609	2,8153	3,4628	4,2592	5,2388	6,4437	7,8258
1,240	1,5376	1,9066	1,9066	2,9316	3,6352	4,5076	5,5894	6,9309	8,5943
1,250	1,5625	1,9531	2,4414	3,0518	3,8148	4,7685	5,9606	7,4508	9,3135
1,260	1,5876	2,0004	2,5205	3,1758	4,0015	5,0419	6,3528	8,0045	10,0857
1,270	1,6129	2,0484	2,6015	3,3039	4,1960	5,3289	6,7677	8,5950	10,9157
1,280	1,6384	2,0972	2,6844	3,4360	4,3981	5,6296	7,2059	9,2236	11,8062
1, 290	1,6641	2,1467	2,7692	3 5723	4,6083	5,9447	7,6687	9,8926	12,7615
1,300	1,6900	2,1970	2,8561	3,7129	4,8268	6,2748	8,1572	10,6044	13,7857
1,350	1,822	2,460	3,321	4,484	6,053	8,172			
1,400	1,960	2,744	3,842						
1,470	2,161	3,176							

Значение F-критерия Фишера

d.f.₂	d.f.₁							
	1	2	3	4	5	6	7	8
1	161	200	216	225	230	234	237	239
2	18,51	19,00	19,16	19,25	19,30	19,33	19,36	19,37
3	10,13	9,55	9,28	9,19	9,01	8,94	8,88	8,84
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95
12	4,75	3,88	3,49	3,26	3,11	3,00	2,92	2,85
13	4,67	3,80	3,41	3,18	3,02	2,92	2,84	2,77
14	4,60	3,74	3,34	3,11	2,96	2,85	2,77	2,70
15	4,54	3,68	3,29	3,06	2,90	2,79	2,70	2,64
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59
17	4,45	3,59	3,20	2,96	2,81	2,7р	2,62	2,55
1-8	4,41	3,55	3,16	2,93	2,77	2,66	2,58	2,51
19	4,38	3,52	3,13	2,90	2,74	2,63	2,55	2,48
20	4,35	3,49	3,10	2,87	2,71	2,60	2,52	2,45
21	4,32	3,47	3,07	2,84	2,68	2,57	2,49	2,42
22	4,30	3,44	3,05	2,82	2,66	2,55	2,47	2,40
23	4,28	3,42	3,03	2,80	2,64	2,53	2,45	2,38
24 '	4,26	3,40	3,01	2,78	2,62	2,51	2,43	2,36
25	4,24	3,88	2,99	2,76	2,60	2,49	2,41	2,34
26	4,22	3,37	2,98	2,74	2,59	2,47	2,39	2,32
27	4,21	3,35	2,96	2,73	2,57	2,46	2,37	2,30
28	4,20	3,34	2,95	2,71	2,56	2,44	2,36	2,29
29-	4,18	3,33	2,93	2,70	2,54	2,43	2,35	2,28
30	4,17	3,32	2,92	2,69	2,53	2,42	2,34	2,27
40	4,08	3,23	2,84	2,61	2,45	2,34	2,25	2,18
50	4,03	3,18	2,79	2,56	2,40	2,29	2,20	2,13
60	4,00	3,15	2,76	2,52	2,37	2,25	2,17	2,10
100	3,94	3,09	2,70	2,46	2,30	2,19	2,10	2,03
0°	3,84	2,99	2,60	2,37	2,21	2,09	2,01	1,94

$d.f._1$ число степеней свободы для большей дисперсии;
$d.f._2$ – число степеней свободы для меньшей дисперсии.

При уровне значимости 0,05

d.f.$_1$								
9	**10**	**11**	**12**	**14**	**16**	**20**	**30**	**∞**
241	242	243	244	245	246	248	250	254
19,38	19,39	19,40	19,41	19,42	19,43	19,44	19,46	19,50
8,81	8,78	8,76	8,74	8,71	8,69	8,66	8,62	8,53
6,00	5,96	5,93	5,91	5,87	5,84	5,80	5,74	5,63
4,78	4,74	4,70	4,68	4,64	4,60	4,56	4,50	4,36
4,10	4,06	4,03	4,00	3,96	3,92	3,87	3,81	3,67
3,68	3,63	3,60	3,57	3,52	3,49	3,44	3,38	3,23
3,39	3,34	3,31	3,28	3,23	3,20	3,15	3,08	2,93
3,18	3,13	3,10	3,07	3,02	2,98	2,93	2,86	2,71
3,02	2,97	2,94	2,91	2,86	2,82	2,77	2,70	2,54
2,90	2,86	2,82	2,79	2,74	2,70	2,65	2,57	2,40
2,80	2,76	2,72	2,69	2,64	2,60	2,54	2,46	2,30
2,72	2,67	2,63	2,60	2,55	2,51	2,46	2,38	2,21
2,65	2,60	2,56	2,53	2,48	2,44	2,39	2,31	2,13
2,59	2,55	2,51	2,48	2,43	2,39	2,33	2,25	2,07
2,54	2,49	2,45	2,42	2,37	2,33	2,28	2,20	2,01
2,50	2,45	2,41	2,38	2,33	2,29	2,23	2,15	1,96
2,46	2,41	2,37	2,34	2,29	2,25	2,19	2,11	1,92
2,43	2,38	2,34	2,31	2,26	2,21	2,15	2,07	1,88
2,40	2,35	2,31	2,28	2,23	2,18	2,12	2,04	1,84
2,37	2,32	2,28	2,25	2,20	2,15	2,09	2,00	1,81
2,35	2,30	2,26	2,23	2,18	2,13	2,07	1,98	1,78
2,32-	2,28	2,24	2,20	2,14	2,10	2,04	1,96	1,76
2,30	2,26	2,22	2,18	2,13	2,09	2,02	1,94	1,73
2,26	2,24	2,20	2,16	2,12	2,06	2,00	1,92	1,71
2,27	2,22	2,18	2,15	2,10	2,05	1,99	1,90	1,69
2,25	2,20	2,16	2,13	2,08	2,03	1,97	1,88	1,67
2,24	2,19	2,15	2,12	2,06	2,02	1,96	1,87	1,65
2,22	2,18	2,14	2,10	2,05	2,00	1,94	1,85	1,64
2,21	2,16	2,12	2,09	2,04	1,99	1,93	1,84	1,62
2,12	2,07	2,04	2,00	1,95	1,90	1,84	1,74	1,51
2,07	2,02	1,98	1,95	1,90	1,85	1,78	1,69	1,44
2,04	1,99	1,95	1,92	1,86	1,81	1,75	1,65	1,39
1,97	1,92	1,88	1,85	1,79	1,75	1,68	1,57	1,28
1,88	1,83	1,79	1,75	1,69	1,64	1,57	1,46	1,00

Значение χ^2-критерия Пирсона при уровне значимости 0,10, 0,05, 0,01

d.f.	0,10	0,05	0,01	d.f.	0,10	0,05	0,01
1	2,71	3,84	6,63	21	29,62	32,67	38,93
2	4,61	5,99	9,21	22	30,81	33,92	40,29
3	6,25	7,81	11,34	23	32,01	35,17	41,64
4	7,78	9,49	13,28	24	33,20	36,42	42,98
5	9,24	11,07	15,09	25	34,38	37,65	44,31
6	10,64	12,59	16,81	26	35,56	38,89	45,64
7	12,02	14,07	18,48	27	36,74	40,11	46,96
8	13,36	15,51	20,09	28	37,92	41,34	48,28
9	14,68	16,92	21,67	29	39,09	42,56	49,59
10	50,99	18,31	23,21	30	40,26	43,77	50,89
11	17,28	19,68	24,72	40	51,80	55,76	63,69
12	18,55	21,03	26,22	50	63,17	67,50	76,15
13	19,81	22,36	27,69	60	74,40	79,08	88,38
14	21,06	23,68	29,14	70	85,53	90,53	100,42
15	22,31	25,00	30,58	80	96,58	101,88	112,33
16	23,54	26,30	32,00	90	107,56	113,14	124,12
17	24,77	27,59	33,41	100	118,50	124,34	135,81
18	25,99	28,87	34,81				
19	27,20	30,14	36,19				
20	28,41	31,14	37,57				

Приложение 6

Критические значения коэффициентов корреляции для уровней значимости 0,05, 0,01

d.f.	α= 0,05	α=0,01	d.f.	α= 0,05	α= 0,01
1	0,996917	0,9998766	17	0,4555	0,5751
2	0,995000	0,990000	18	0,4438	0,5614
3	0,8783	0,96873	19	0,4329	0,5487
4	0,8114	0,9172	20	0,4227	0,5368
5	0,7545	0,8745	25	0,3809	0,4869
6	0,07067	0,8343	30	0,3494	0,4487
7	0,6664	0,7977	35	0,3246	0,4182
8	0,6319	0,7646	40	0,3044	0,3932
9	0,6021	0,7348	45	0,2875	0,3721
10	0,5760	0,7079	50	0,2732	0,3541
11	0,5524	0,6835	60	0,2500	0,3248
12	0,5324	0,6614	70	0,2919	0,3017
13	0,5139	0,6411	80	0,2172	0,2830
14	0,4973	0,6226	90	0,2050	0,2673
15	0,4821	0,6055	100	0,1946	0,2540
16	0,4683	0,5897			

Для простой корреляции d.f. на 2 меньше, чем число пар вариантов; в случае частной корреляции необходимо также вычесть число исключаемых переменных.

ЛИТЕРАТУРА

Учебники и учебные пособия

Для 1 раздела

1. Общая теория статистики: Учебник / Под ред. А. Спирина, О.Баншиной. М.: Финансы и статистика, 1994. 296 с.
2. Громыко Г.Л. Общая теория статистики: Практикум. М.: ИНФРА-М, 1999. 139 с.
3. Ефимова М.Р., Петрова Е.В., Румянцев В.Н. Общая теория статистики: Учебник. М.: ИНФРА-М, 2002. 416 с.
4. Елисеева И.И., Юзбашев М.М. Общая теория статистики: Учебник. М.: Финансы и статистика, 2002. 480 с.
5. Громыко Г.Л. Теория статистики: Учебник. М.: ИНФРА-М, 2005.
6. Ефимов М.Р., Петрова Е.В., Румянцев В.Н. Общая теория статистики: Учебник. М.: ИНФРА-М, 2007.
7. Лысенко С.Н., Дмитриева И.А. Общая теория статистики: Учебное пособие. М.: Вузовский учебник, 2009. 219 с.
8. Абдуллаев Ё. Статистика назарияси: Дарслик. Т.: Ўқитувчи, 2002. 589 б.

Для 2 раздела

1. Экономическая статистика: Учебник / Под ред. Ю.Н.Иванова. М.: ИНФРА-М, 1998. 480 с.
2. В.Н. Салин, В.Г. Медведев, С.Н. Кудряшева, Е.П. Шпаковская. Макроэкономическая статистика: Учебное пособие. М.: Дело, 2001. 336 с.
3. Систма национальный счетов – инструмент макро экономического анализа: Учебное пособие / Под ред. Ю.Н.Иванова. М.: Финстатинформ, 1996. 285 с.
4. Абдуллаев Ё. Макроиқтисодий статистика. Т.: Меҳнат, 1998. 382 б.
5. Умаров Н., Абдуллаев А., Зулинова З. Статистика: Учебник. Т.: Iqtisod-Moliya, 2009. 306 с.

Ё.А. Абдуллаев, У.У. Азизов,
З.Х. Тошматов, М.М. Икрамов

ТЕОРИЯ
СТАТИСТИКИ

Учебник

Редактор Э. Хуснутдинова
Художник К. Бойхужаев
Компьютерная верстка О. Фозилова

Лиц. изд. AI № 305. Подписано в печать 21.02.2021.
Формат 60x84 1/16. Усл.печ.л.27,75. Уч.-изд.л. 28,5.
Тираж 100 экз. Заказ № 20.

Издательство «IQTISOD-MOLIYA».
100000, Ташкент, ул. Амира Темура, 60^{A}.

Отпечатано в типографии
ООО «DAVR MATBUOT SAVDO».
100198, Ташкент, Куйлюк, массив 4, 46.