ВЫСШЕЕ ОБРАЗОВАНИЕ

Общая теория статистики

УЧЕБНИК

ВЫСШЕЕ ОБРАЗОВАНИЕ

серия основана в 1996 г.

М.Р. ЕФИМОВА
Е.В. ПЕТРОВА
В.Н. РУМЯНЦЕВ

ОБЩАЯ ТЕОРИЯ СТАТИСТИКИ

1887

УЧЕБНИК

2-е издание, исправленное и дополненное

Рекомендовано в качестве учебника
для студентов высших учебных заведений,
обучающихся по специальности: финансовый,
банковский, производственный менеджмент,
бухгалтерский учет и аудит, международные
экономические отношения

Москва
ИНФРА-М
2002

УДК (075.8)311
ББК 60.6
Е91

М.Р. Ефимова – глава 3 (3.1, 3.2, 3.3), глава 4 (4.1, 4.2, 4.3), главы 5, 6, 7, 8, 9; Е.В. Петрова – глава 1; В.Н. Румянцев – глава 2; Н.К. Агеева, Н.Е. Варакина – 4.4, 4.5; Н.Е. Варакина, Е.А. Саблина – 3.4.

Общая редакция проф. М.Р. Ефимовой.

Ефимова М.Р., Петрова Е.В., Румянцев В.Н.

Е91 Общая теория статистики: Учебник. Изд. 2-е, испр. и доп. – М.: ИНФРА-М, 2002. – 416 с. – (Серия «Высшее образование»).

ISBN 5-16-000012-7

Рассматриваются вопросы организации статистики, сбора статистической информации в условиях рыночной экономики. Излагается методология расчета относительных и средних величин, их применение в экономико-статистическом анализе. Особое внимание уделено статистическим методам изучения взаимосвязей, группировкам, анализу рядов динамики, выборочному и индексному методам анализа. Использованы данные статистических ежегодников.

Для студентов высших учебных заведений и факультетов, обучающихся по направлению «Экономика».

УДК (075.8)311
ББК 60.6

ISBN 5-16-000012-7

Глава 1
ЗАДАЧИ СТАТИСТИКИ И ЕЕ ОРГАНИЗАЦИЯ

1.1. Общее представление о статистике и краткие сведения из ее истории

Термин «статистика» произошел от латинского слова «статус» (status), что означает «определенное положение вещей». Употреблялся он первоначально в значении слова «государствоведение»; впервые был введен в обиход в 1749 г. немецким ученым Г.Ахенвалем, выпустившим книгу о государствоведении.

В настоящее время термин «статистика» употребляется в трех значениях. Во-первых, под статистикой понимают особую отрасль практической деятельности людей, направленную на сбор, обработку и анализ данных, характеризующих социально-экономическое развитие страны, ее регионов, отраслей экономики, отдельных предприятий. Во-вторых, статистикой называют науку, занимающуюся разработкой теоретических положений и методов, используемых статистической практикой. Между статистической наукой и статистической практикой существует тесная связь. Статистическая практика применяет правила, выработанные наукой; в свою очередь статистическая наука опирается на материалы практики и, обобщая опыт практики, разрабатывает новые положения. В-третьих, статистикой часто называют статистические данные, представленные в отчетности предприятий, организаций, отраслей экономики, а также публикуемые в сборниках, справочниках, периодической прессе, которые представляют собой результат статистической работы.

Особенность статистики заключается в том, что статистические данные сообщаются в количественной форме, т.е. статистика говорит языком цифр, отображающих общественную жизнь во всем многообразии ее проявлений. При этом статистику прежде всего интересуют те выводы, которые можно сделать на основе анализа надлежащим образом собранных и обработанных цифровых данных.

Приведем пример статистических цифр (данных) в форме статистической таблицы (табл. 1.1):

Таблица 1.1

Численность населения Российской Федерации

	1980 г.	1985 г.	1990 г.	1996 г.
Численность населения (на конец года), млн. человек	138,8	143,6	148,2	147,95
Изменение численности по сравнению с 1980 г., %	100,0	103,5	106,8	106,3

Источник. Российский статистический ежегодник 1997. Статистический сборник/Госкомстат России. М., 1997. С.67.

Как видно из приведенного примера, данные статистики отражают состояние (уровень) изучаемого явления на определенной ступени его развития в конкретных условиях места и времени, а также качественные сдвиги в процессе развития. Абсолютные цифры численности населения имеют ознакомительное значение; их сопоставления друг с другом позволяют выявить определенную тенденцию в изменении численности населения.

Следовательно, называя то или иное статистическое число, всегда указывают, какое явление изучается, время и место, в границах которых оно наблюдается, в каких единицах измерения выражается его величина.

Статистика имеет многовековую историю, уходя своими корнями в глубокую древность. 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.

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.

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). 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), В. Госсета (1876–1936), Р. Фишера (1890-1962) и др.

В России становление статистической науки 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 в 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 г. – 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). 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 западноевропей-

ских 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 (С.Г. Струмилин, В.С. Немчинов, Б.С. Ястремский, А.Я. Боярский и др.).

Во второй половине века развитие производства становится все более сложным, социально-экономические процессы более многообразными, это поставило перед практической статистикой новые более трудные задачи, в решении которых принимает участие статистическая наука. Усложнение межрегиональных, межотраслевых и отраслевых связей потребовало разработки новых и улучшения существующих показателей, которые достаточно полно характеризовали бы эти связи и пропорции. В свет вышло много работ, посвященных обсуждению важнейших вопросов совершенствования статистической методологии, применения статистических методов исследования производства. При этом развитие статистической науки шло не только по пути дальнейшего углубления и усовершенствования методов исследования, но и по пути их специализации применительно к особенностям отдельных отраслей хозяйства и социальных отношений. Особое внимание уделялось формированию самостоятельных отраслей статистики: статистики промышленности, строительства, транспорта и др. Дифференциация областей статистического исследования и вместе с тем дифференциация приемов изучения явились необходимым условием наиболее полного выявления резервов производства и сознательного их использования.

Серьезного внимания заслуживают работы по изучению статистической связи, теории средних и закону больших чисел, теории индексного метода. Развитие электронно-вычислительной техники создало благоприятные условия для прогресса статистической методологии и углубленного исследования социально-экономических процессов.

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 и др.

Нельзя не сказать о значении для развития теории статистики издания специальных журналов, статистических сборников. С 1914 по 1917 г. регулярно издавался журнал «Статистический вестник». Ранее вопросы статистической методологии и практики освещались в других периодических изданиях, например, в журналах «Статистический журнал», «Юридический вестник», «Вестник Императорского Русского Географического Общества» и др. С 1919 по 1929 г. издавался журнал «Вестник статистики», в 1949 г. его издание возобновилось, а в 1995 г. он получил название «Вопросы статистики». В 50-х годах начали выходить «Ученые записки по статистике», издаваемые Академией наук. Отдельные тома посвящались актуальным проблемам: статистическому контролю хода производственных процессов и качества продукции, измерению и анализу производительности труда, международным сравнениям статистических показателей и др.

В настоящее время издаются учебники, монографии и научные труды по статистике, в том числе наиболее интересные работы иностранных статистиков, а также статистические сборники («Российский статистический ежегодник», информационные статистические бюллетени и др.). Большое значение придается укреплению контактов с органами ООН, занимающимися статистикой, Международным статистическим институтом; изучается все полезное, разработанное статистической наукой в других странах.

1.2. Предмет статистической науки и ее методология

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 позна-

ния, 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. Социально-экономическая жизнь общества проявляется в различного рода массовых явлениях, как, например, производство разнообразных видов продукции, потребление этой продукции, экспорт и импорт продукции, перевозка грузов и пассажиров и другие явления экономической, культурной и политической жизни. Статистика также изучает природные ресурсы и природные условия, поскольку они влияют на жизнь общества.

Явления и процессы в жизни общества характеризуются статистикой с помощью статистических показателей. **Статистический показатель** - это количественная оценка свойств изучаемого явления. При этом следует имеет в виду, что в статистическом показателе проявляется единство качественной и количественной сторон. Если не определена качественная сторона явления, то нельзя определить и его количественную сторону. Так, правильно определить размер валового национального продукта страны можно лишь при условии, если известна его качественная характеристика.

В каждый исторический момент социально-экономические явления имеют конкретные размеры, структуру, интенсивность развития, ту или иную распространенность и определенные соотношения друг с другом. Статистика при помощи статистических показателей характеризует размеры изучаемых явлений, их особенности, закономерности развития и их взаимосвязи. При этом статистические показатели подразделяются на учетно-оценочные и аналитические. *Учетно-оценочные показатели* отражают объем или уровень изучаемого явления; *аналитические показатели* используются для характеристики особенностей развития явления, распространенности в пространстве, соотношения его частей, взаимосвязи с другими явлениями. В качестве аналитических показателей используются средние величины, показатели структуры, вариации, динамики, степени тесноты связи и др.

Например, статистика показывает численность отдельных групп населения страны и ее регионов на начало отчетного года, используя учетно-оценочные показатели, а с помощью аналитических показателей характеризует изменение численности населения по сравнению с началом предшествующего года, отражает половозрастной состав населения, плотность населения по регионам и т.д.

Статистическая методология представляет собой совокупность общих правил (принципов) и специальных приемов и методов статистического исследования. Общие правила статистического исследования исходят из положений социально-экономической теории и принципа диалектического метода познания. Они составляют теоретическую базу статистики.

Теоретический (качественный) анализ явления, основанный на социально-экономических науках, всегда предшествует его статистическому изучению и является необходимым условием правильной организации статистического исследования и безошибочного толкования его результатов. Необходимым условием статистического изучения является понимание сущности изучаемого объекта или процесса, знание законов развития и особенностей конкретной обстановки. Так, прежде чем провести статистическое исследование влияния отдельных факторов на изменение производительности труда работников промышленного предприятия, необходимо предварительно уяснить понятие производительности труда, обосновать метод расчета показателя для той отрасли экономики, к которой относится предприятие, определить состав факторов и характер их воздействия. Решение этих вопросов требует соответствующих знаний экономики отрасли.

Одновременно, руководствуясь положениями социально-экономической теории, статистика обогащает социально-экономические науки фактическими данными, полученными в статистическом исследовании; статистическая информация используется для проверки, обоснования или иллюстрации их теоретических положений.

В соответствии с диалектическим методом познания статистика изучает все явления в их взаимосвязи, в движении и изменении, выделяя их различные типы и формы, устанавливает то новое, прогрессивное, что зарождается в существующем и определяет направление развития. В процессе развития наряду с количественными изменениями в изучаемом предмете происходят коренные качественные изменения. Поэтому необходимо располагать методами, позволяющими изучать количественные изменения в явлениях, оценивать существенность или несущественность наблюдаемых различий, улавливать переход количественных изменений в качественные. Так, при изучении производительности

труда работников предприятия статистика не только определяет достигнутый уровень производительности труда и характеризует его динамику, но и выявляет зависимость производительности труда от степени использования производственного оборудования, уровня механизации тяжелых и трудоемких работ, квалификации рабочих, применяемой системы оплаты труда и других факторов, а также определяет влияние роста производительности труда на изменение объема произведенной продукции, уровень ее себестоимости. В том числе характеризуются результаты работы отдельных подразделений предприятий - цехов, участков, выявляются различия в условиях производства, показывается степень внедрения новой технологии, новых форм организации труда и т.п.

В процессе исследования своего предмета статистика может использовать и другие общенаучные методы, например аналогию (перенесение свойств одного предмета на другой) или гипотезу (научно обоснованное предположение о возможных причинных связях между явлениями).

Опираясь на теоретическую базу, **статистика применяет специфические методы цифрового освещения явления**, которые находят свое выражение в трех этапах (стадиях) статистического исследования:

1. Массовое научно организованное наблюдение, с помощью которого получают первичную информацию об отдельных единицах (фактах) изучаемого явления. Например, при переписи населения регистрируются заранее обусловленные признаки всех жителей страны по тщательно разработанному плану. Массовое статистическое наблюдение (учет большого числа или всех входящих в состав изучаемого явления единиц) представляет исходный материал для статистических обобщений, для получения объективных выводов об изучаемом явлении. Получение сведений о достаточно большом числе единиц дает возможность освободиться от влияния случайных причин и установить характерные черты изучаемого объекта.

2. Группировка и сводка материала, представляющие собой расчленение всей массы случаев (единиц) на однородные группы и подгруппы, подсчет итогов по каждой группе и подгруппе и оформление полученных результатов в виде статистической таблицы. Группировки дают возможность выделить из состава всех случаев единицы разного качества, показать особенности явле-

ний, развивающихся в различных условиях. После проведения группировки приступают к обобщению данных наблюдения по выделенным частям и целому, т.е. к получению статистических показателей в форме абсолютных величин (учетно-оценочные показатели), при помощи которых измеряют объемы (размеры) явлений. Эта ступень работы носит название сводки. Например, первичная информация, полученная при переписи населения, подразделяется на социальные группы, группы по полу, возрасту и т.д.; по каждой выделенной группе подсчитывается численность населения;

3. Обработка статистических показателей, полученных при сводке, и анализ результатов для получения обоснованных выводов о состоянии изучаемого явления и закономерностях его развития. Выводы, как правило, излагаются в текстовой форме и сопровождаются графиками и таблицами.

При обработке данных исчисляют аналитические показатели, отражающие особенности отдельных однородных групп (подгрупп), соотношения и взаимосвязи меду ними. Они определяются в форме средних, относительных величин, показателей вариации, индексных показателей. Для этого этапа исследования характерно применение всего арсенала статистических методов; применение специальных методов предопределяется поставленными задачами и особенностями первичной информации.

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.

Последовательность статистического исследования, особое содержание и познавательное значение статистических цифр проиллюстрировано примером, представленным в форме статистической таблицы (табл. 1.2).

Таблица 1.2

Отправление грузов в России по видам транспорта общего пользования

Вид транспорта	Отправлено грузов, млн т		Изменение отправления грузов, раз	Удельный вес в общем объеме отправления, %	
	1992 г.	1996 г.		1992 г.	1996 г.
1	2	3	4	5	6
Транспорт - всего	4869, 4	2850, 8	0, 585	100, 00	100, 00
железнодорожный	1640, 0	911, 0	0, 555	33, 68	31, 95
автомобильный	1882, 0	1002, 0	0, 532	38, 65	35, 15
трубопроводный	947, 0	783, 0	0, 827	19, 45	27, 47
морской	91, 0	54, 0	0, 593	1, 87	1, 89
внутренний водный	308, 0	100, 0	0, 325	6, 63	3, 51
воздушный	1, 4	0, 8	0, 571	0, 02	0, 03

Источник. Российский статистический ежегодник 1994 г. Статистический сборник/Госкомстат России. С. 391.

На первом этапе статистического исследования были получены *отчетные данные* об отправлении грузов по всем транспортным предприятиям общего пользования; этот этап представлял собой массовое статистическое наблюдение. В графах 2 и 3 приведены абсолютные величины (цифры), характеризующие отправление грузов каждым видом транспорта и в целом по транспорту; они получены на основе группировки и сводки материалов и носят название *учетно-оценочных показателей*. Цифры граф 4, 5, 6 получены в результате обработки статистических показателей, полученных при сводке, и носят название *аналитических показателей*. Их анализ позволяет сформулировать общие выводы.

Сопоставление цифр о размере отправлений грузов за два года (гр. 4) позволило обнаружить значительные изменения в размере отправлений грузов: так, в целом по транспорту отправление в 1996 г. по сравнению с 1992 г. уменьшилось на 41,5% [(0,585x100)–100], особенно значительное снижение произошло по автомобильному транспорту (уменьшение на 44,5%). Сравнение размера отправления по каждому виду транспорта с общим размером отправления в целом по транспорту выявляет значение каждого вида транспорта в транспортном балансе страны (гр. 5 и гр. 6). Наибольшее значение в транспортном балансе страны занимают три вида транспорта: железнодорожный, автомобильный и трубопроводный. Так, в 1996 г. 31,95% всех отправлений груза приходилось на железнодорожный транспорт – [(911,0 : : 2850,8)x100]. Доля автомобильного транспорта в 1996 г. по сравнению с 1992 г. снизилась с 38,65 до 35,15%; в то же время доля трубопроводного транспорта возросла.

С вопросом о методе статистики связан вопрос о ее связи с математикой. Эта связь объясняется тем, что для измерения и анализа количественных отношений необходимо применение математических приемов и методов. При этом следует иметь в виду, что в статистике применяется математика различных уровней сложности. Длительное время статистика использовала в своей работе простейшие приемы в математике (правила арифметики, алгебраические выражения и т.п.). Но необходимость изучения массовых случайных процессов вызвала потребность использования специальных математических дисциплин - теории вероятностей и математической статистики.

Теория вероятностей занимается исследованием случайных процессов, ее положениями и теоремами широко пользуется статистика. При исследовании тенденций и закономерностей развития общественных явлений статистика опирается на закон больших чисел, выражающий диалектику случайного и необходимого. 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 не зависящему 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. На этом основано применение выборочного метода наблюдения, имеющего в практической статистике важное значение.

Широко применяется в статистике аппарат математической статистики - анализ вариационных рядов, корреляционный и регрессионный анализ. В последние годы используются и другие приемы высшей математики: методы оптимального программирования, теория распознавания образов и др.

Значение математики для развития статистики резко выросло в современных условиях в связи с широким использованием вычислительной техники. Внедрение математики в статистику позволяет упорядочить систему статистической информации, обеспечивает возможность создания стандартных программ для «перевода» фактических данных на формализованный язык для проведения массовых расчетов, осуществляемых вычислительными машинами. Это приводит к значительному ускорению обработки и передачи данных, упорядочению хранения, облегчению и ускорению их поиска в больших массивах.

Однако следует иметь в виду, что использование сложного математического аппарата не может превратить статистику в математику. Математика - наука, исследующая пространственные формы и количественные соотношения реального мира «в чистом виде», отвлекаясь от определенного материального содержания. Для статистики, изучающей явления социально-экономический жизни в их конкретном своеобразии, математика имеет значение лишь как инструмент исследования.

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. Структура статистической науки представлена на рис. 1.1.

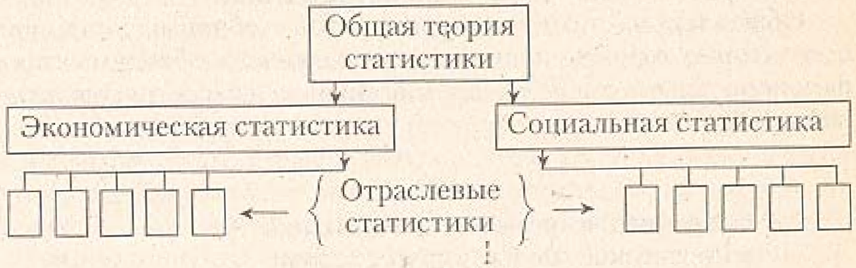

Рис. 1.1. **Структура статистической науки**

Таким образом, в статистической науке выделяются следующие части: общая теория статистики, экономическая статистика и ее отрасли, социальная статистика и ее отрасли.

Общая теория статистики разрабатывает общие принципы и методы статистического исследования общественных явлений, наиболее общие категории (показатели) статистики.

Задачей **экономической статистики** является разработка и анализ синтетических показателей, отражающих состояние национальной экономки, взаимосвязи отраслей, особенности размещения производительных сил, наличие материальных, трудовых и финансовых ресурсов, достигнутый уровень их использования. Отрасли экономической статистики - статистика промышленности, сельского хозяйства, строительства, транспорта, связи, труда, природных ресурсов, охраны окружающей среды и т.д.; их задача - разработка и анализ статистических показателей развития соответствующих отраслей.

Статистики крупных отраслей могут быть подразделены на более мелкие отраслевые статистики: например, статистика промышленности - на статистику машиностроения, металлургии, химии и др.; статистика сельского хозяйства - на статистику земледелия и животноводства и т.п.

Социальная статистика формирует систему показателей для характеристики образа жизни населения и различных аспектов социальных отношений; ее отрасли - статистика народонаселения, политики, культуры, здравоохранения, науки, просвещения, права и т.д.

Отраслевые статистики формируются на базе показателей экономической или социальной статистики, а 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.

1.3. Современная организация статистики в Российской Федерации и ее задачи

Изучением экономического и социального развития страны, отдельных ее регионов, отраслей, объединений, фирм, предприятий занимаются специально созданные для этого органы, совокупность которых называется **статистической службой**. В Российской Федерации функции статистической службы выполня-

ют органы государственной статистики и органы ведомственной статистики.

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.

Правительственная статистика в России провела несколько крупных работ: в 1897 г. была проведена первая в России перепись населения, промышленные переписи 1908 и 1913 г., сельскохозяйственные переписи 1916 и 1917 гг.

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.

В настоящее время главным учетно-статистическим центром в стране является Государственный комитет Российской Федерации по статистике (Госкомстат России). Он осуществляет руководство российской статистикой в соответствии со ст. 71 Конституции Российской Федерации. В его задачи входит представление официальной статистической информации Президенту, Правительству, Федеральному собранию, федеральным органам исполнительной власти, общественным и международным организациям, разработка научно обоснованной статистической методологии, координация статистической деятельности федеральных и региональных органов исполнительной власти, анализ экономико-статистической информации, составление национальных счетов и балансовых расчетов.

Система органов государственной статистики образована в соответствии с административно-территориальным делением страны. В республиках, входящих в Российскую Федерацию, имеются республиканские комитеты, в автономной области, автономных округах, краях, а также в областях, в городах Москве и Санкт-Петербурге действуют государственные комитеты по статистике (комитеты государственной статистики), в районах (городах) - управления (отделы) государственной статистики.

Госкомстат РФ выполняет работу по сбору, обработке и анализу научно обоснованных данных, характеризующих экономическое и социальное развитие страны, процессы становления многоукладной экономики, ход выполнения государственных и региональных программ по решению важнейших народнохозяйственных проблем, эффективность производства. Характеризуя развитие национальной экономики страны, Госкомстат выявляет соотношения отраслей экономики, соотношения между размерами производства продукции и размерами ее потребления, в том

числе потребления различных видов продукции на душу населения, отражает достигнутый уровень валового внутреннего и валового национального продукта, национальный доход и т.д. Этой информацией обеспечиваются законодательная власть, исполнительные, управленческие и хозяйственные органы. Одновременно с этим Госкомстат занимается совершенствованием методологии учета и статистики, разрабатывает формы отчетности. В настоящее время особое значение придается формированию бухгалтерских и статистических показателей в соответствии с требованиями международного бухгалтерского учета и системы национальных счетов (СНС), поскольку эта система наиболее полно отвечает требованиям рыночных отношений. Методы сбора и обработки статистических данных, методология исчисления статистических показателей, установленные Госкомстатом России, являются официальными стандартами Российской Федерации.

Органы государственной статистики выполняют переписи и единовременные учеты, необходимые для глубокого изучения отдельных сторон экономики и образа жизни населения; публикует данные об экономическом и социальном развитии страны и отдельных ее регионов, данные по международной статистике и международным сопоставлениям.

В связи с переходом на принятую в международной практике систему учета и статистики в России создан и функционирует Единый государственный реестр (регистр) предприятий, организаций, учреждений и объединений (ЕГРПО). Цель его создания - 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. В соответствии с Законом Российской Федерации «Об ответственности за порядок представления государственной статистической отчетности» руководители предприятий и другие должностные лица, подписывающие отчет, несут административную ответственность за непредставление отчетов и других данных, необходимых для проведения государственных статистических наблюдений, искажение данных или нарушение сроков представления отчетов.

Держателем ЕГРПО является Госкомстат России, держателями республиканских, краевых, областных, городских уровней Реестра являются соответствующие органы государственной статистики. Госкомстат обеспечивает методическое руководство, координацию и контроль за государственным учетом юридических лиц и ведением ЕГРПО, определяет состав и источники получения экономических показателей, методологию их исчисления и формирования результативной информации. Пользователями ЕГРПО могут быть любые юридические и физические лица, заинтересованные в информации.

Одной из важных задач центрального органа государственной статистики является укрепление контактов с международными статистическими службами ООН, в первую очередь с ее Статистической комиссией. В ее задачи входит разработка методологии статистических работ, системы сопоставимых показателей, разработка и анализ статистической информации, координация статистической работы специализированных органов ООН, подготовка рекомендаций для Статистического бюро Секретариата ООН. Являясь исполнительным органом Статистическое бюро Секретариата ООН собирает статистическую информацию от государств – членов ООН, публикует ее, выполняет доклады по различным вопросам статистики и публикует результаты выполненных исследований в периодических изданиях («Ежегодник по внешней торговле», «Демографический ежегодник» и др.).

Наряду с общегосударственной статистикой существует **ведомственная статистика**, ведущаяся на предприятиях, в объединениях, ведомствах, министерствах. Ведомственная статистика выполняет работы, связанные с получением, обработкой и анализом статистической информации, необходимой для руководства и планирования их деятельности. Для ведения статистики на предприятиях, в объединениях, концернах, 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.

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. 1.2. представлено в общем виде формирование показателей производственной деятельности предприятия (объединения, концерна) любой формы собственности.

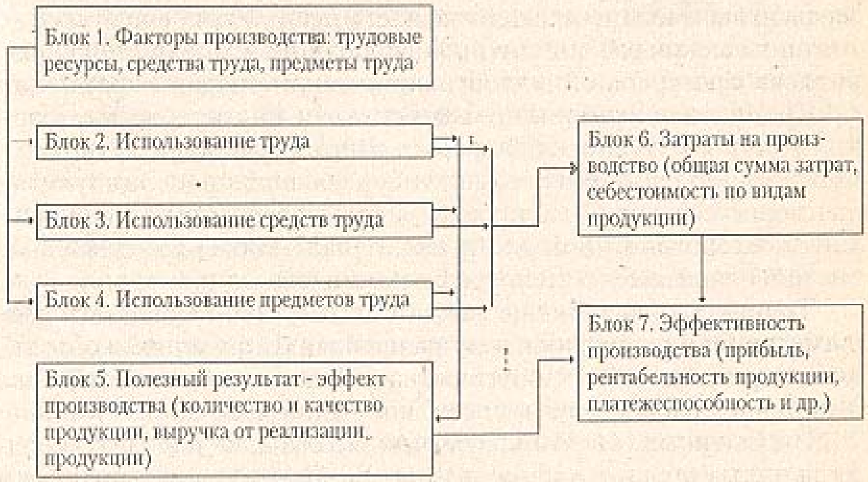

Рис. 1.2. **Блок-схема формирования показателей производственной деятельности предприятия**

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. Но круг показателей статистики шире планового, так как наряду с планируемыми статистика устанавливает и не планируемые показатели. Например, в плане нет и не может быть показателей целодневных и внутрисменных простоев, сверхурочных работ и ряда других показателей. В задачу статистики входит не только проверка выполнения плана, но и выявление качества самого планирования. Так, в частности, в отдельных случаях можно установить, что недовыполнение плана зависело от причин, влияние которых не было должным образом учтено при составлении плана. Выявление неиспользованных резервов производства для более полного удовлетворения потребностей в продукции предприятия и повышения эффективности его работы достигается взаимосвязанным анализом показателей каждого блока, что может быть достигнуто при наличии достаточной информации.

Критерий эффективности предполагает оптимальное соотношение результатов производства и необходимых для его достижения затрат. Так, в процессе анализа статистика должна выявить резервы повышения производительности труда, улучшения использования основных фондов и производственных мощностей, материальных ресурсов, снижения непроизводительных затрат и т.д.

Кроме оценки работы предприятия в целом задачей статистики является изучение результатов работы его подразделений – цехов, участков, бригад, выявление реальных пропорций, складывающихся в процессе производства. Повышение роли экономических методов управления предполагает совершенствование системы показателей оценки их работы.

Точные и объективные данные статистики необходимы для составления планов работы предприятий. Причем в новых условиях хозяйствования требуется укрепление связи прогнозирования, текущего и перспективного планирования.

Необходимость в статистической информации появляется уже на начальном этапе планирования. В общей системе планов на предприятиях определяющим является план выпуска продукции, поэтому обоснованное предвидение темпов роста и пропорций

определенных видов продукции на перспективу имеет исключительно большое значение. С развитием рыночных отношений меняется подход к планированию выпуска продукции, что требует дополнительной статистической информации. В первую очередь статистика обеспечивает информацию о состоянии и сегментации рынка, о конкурентоспособности каждого вида продукции, работ, услуг. Важное значение имеет анализ динамики цен и исследование вопросов ценообразования. Рыночный механизм ценообразования требует системы статистических показателей как информационной основы для моделирования рыночных ситуаций и для обоснованного прогнозирования последствий. Для целей составления плана необходимы статистические данные о наличии трудовых и материальных ресурсов, достигнутом уровне их использования, имеющихся резервах.

Перед органами государственной и ведомственной статистики стоит ответственная задача теоретического обоснования объема и состава статистической информации, которая отвечала бы современным условиям развития экономики, перехода к новым принципам управления. Вопросы улучшения информационной базы требуют решения двух групп задач. Во-первых, система информации должна содействовать полному удовлетворению потребностей в информации различных уровней управления; на предприятиях эта система должна обеспечить переход в управлении преимущественно на режим предотвращения перебоев в работе вместо режима устранения последствий отклонений от нормального хода производства. Во-вторых, необходимо сдерживать рост информации за счет исключения избыточных данных, не имеющих значения для принятия обоснованных управленческих решений; это связано с рационализацией в самой системе учета и статистики и должно способствовать минимизации затрат на выполнение этой функции.

Таким образом, главным содержанием статистики является исчисление статистических показателей и их анализ, благодаря чему управляющие органы получают всестороннюю характеристику управляемого объекта, будь то вся национальная экономика или отдельные ее отрасли, предприятия и их подразделения. Управлять сложными социальными и экономическими системами нельзя, не располагая оперативной, достоверной и полной статистической информацией.

Контрольные вопросы к главе 1

1. Что означает термин «статистика»?
2. Чем обусловлено возникновение и развитие статистической практики и науки?
3. Что является предметом исследования статистической науки? Приведите примеры явлений общественной жизни, изучаемых статистикой.
4. Дайте определение статистического показателя и укажите их виды.
5. В чем заключается сущность статистической методологии?
6. Перечислите стадии статистического исследования, раскройте их основное содержание.
7. Какова роль и значение математики в статистическом исследовании?
8. Перечислите части (разделы) статистической науки и объясните, чем вызвано выделение самостоятельных статистических дисциплин.
9. Каковы принципы организации статистики в России в настоящее время?
10. Каковы задачи государственной статистики в условиях перехода к рыночной экономике?
11. Каковы задачи ведомственной статистики и чем объясняется возрастание ее роли в современных условиях?

Глава 2
СТАТИСТИЧЕСКОЕ НАБЛЮДЕНИЕ

2.1. Формирование информационной базы статистического исследования

С развитием рыночных отношений роль информационной базы возрастает, поскольку усложняются связи субъектов рынка, возникает все более настоятельная потребность в изучении влияния различных факторов на результаты деятельности, социальные последствия, а также в прогнозировании, в обобщениях как на макро-, так и на микроуровне. Важнейшим ресурсом в управлении становится статистическая информация.

Для того чтобы выполнить статистическое исследование, необходима научно обоснованная информационная база. Она формируется в результате статистического наблюдения, которое является начальной стадией экономико-статистического исследования. *Статистическим наблюдением называется планомерный научно обоснованный сбор данных или сведений о социально-экономических явлениях и процессах.*

Статистические данные представляют собой составную часть глобальной информационной системы, которая формируется в соответствии с концепцией информатизации, разработанной в Российской Федерации. Информационная база статистики призвана обеспечить поддержку формирующегося рынка, дать всестороннюю и объективную информацию для разработки вариантов, обоснования и принятия управленческих решений. 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.

Владея информацией, предприятия могут эффективнее решать поставленные задачи. И наоборот, основной причиной неизбежных банкротств в России являются некомпетентные действия руководства предприятий: этому способствует отсутствие в распоряжении руководителей информации о конъюнктуре рынка и многих других жизненно важных для данного предприятия явлениях и процессах. Статистическое наблюдение помогает предприятию реализовать маркетинговую стратегию, гибко реагировать на изменения рынка, сделать обоснованный выбор.

Отметим, что термин «статистические данные» представляет собой, образно говоря, «сырье», полученное в результате статистического наблюдения. В процессе обработки, анализа оно становится информацией.

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.

Применение весьма тонких методов статистического исследования для нужд управления, которые рассматриваются в последующих главах, может оказаться неэффективным или вообще бесполезным, если у исследователя не будет уверенности в качестве первичного материала.

Имеется определенная связь и отличие статистического наблюдения и непосредственного восприятия человеком явлений. С одной стороны, статистическое наблюдение может базироваться на собственном наблюдении человека. Так, установление факта может осуществляться на основе опроса населения или измерения конкретного параметра объекта, подсчета объектов и т.п. Статистический характер такому наблюдению придает регистрация его результатов в определенном инструментарии - опросном листе, анкете и других специальных учетных документах. С другой стороны, статистическое наблюдение может представлять процесс сбора информации и по уже зарегистрированным данным, например, по отчетности. Здесь наиболее выражены отличия статистического наблюдения от непосредственного восприятия человеком конкретных явлений, объектов.

Всякие ли данные, факты, собранные в процессе статистического наблюдения, могут быть использованы для дальнейшего исследования? Нет, не всякие. Они должны отвечать определенным требованиям.

Важнейшим требованием является *достоверность данных*, получаемых в результате статистического наблюдения. Достоверность обеспечивается многими условиями. Это компетентность работника, участвующего в статистическом наблюдении, совершенство инструментария (бланков, инструкций), заинтересованность или готовность объекта и др. Достоверность включает как соответствие данных реальной действительности, так и техническую точность или обоснованность измерения. Например, поставлена задача исследовать загрузку оборудования и возникающие у рабочих внутрисменные простои. Цель требует, чтобы простои были зарегистрированы с точностью до минуты. В такой ситуации можно получить достаточно объективную картину при обобщении данных и выработке рекомендаций.

Достоверность тесно связана с полнотой данных – полученные данные должны быть достаточно полными. 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*. Каждое явление или совокупность, изучаемые во времени или в пространстве, должны быть сопоставимы. Для этого необходимо использовать единые стоимостные оценки, что особенно важно в условиях инфляции, единые территориальные границы, что также актуально для России на современном этапе, т.е. строго соблюдать единство в методологии.

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.

Определенные особенности на статистическое наблюдение налагает сложившаяся в стране в целом практика статистических исследований. Она позволяет выделить *обособленные* и *необособленные системы организации статистического наблюдения*. Обособленная система может быть также *централизованной* и *децентрализованной*.

В РФ действующая в основном централизованная система сочетается с функционированием необособленной системы, когда сбор и обработка определенной статистической информации осуществляется отдельными органами или ведомствами, такими как налоговая инспекция, Госкомимущество РФ, органы здравоохранения и др.

Централизованная система осуществляет сбор данных для наиболее важных общегосударственных исследований, разрабатывает методологическую базу. Она позволяет концентрировать ресурсы на выполнении государственных задач, имеет современное техническое оснащение, квалифицированные кадры. Осуществляет руководство и контроль за организацией необособленной системы сбора данных в министерствах, ведомствах, регионах.

Общее руководство и методологическое обеспечение проводимых в стране статистических наблюдений осуществляет Госкомстат РФ. Можно выделить следующие важнейшие задачи Госкомстата РФ в области статистического наблюдения:

1. Сбор данных о социально-экономическом положении страны.

2. Реализация программ по проведению важнейших общегосударственных наблюдений - переписей населения и др.

3. Методология и организация единой системы статистической отчетности, представляемой государственными учреждениями, профсоюзными и общественными организациями, частными и другими предприятиями.

4. Сбор данных по программам международных органов - СНГ, ООН.

5. Создание и совершенствование новых технологий сбора, хранения, поиска и выдачи данных или информации потребителям.

Наиболее важные общегосударственные статистические наблюдения проводятся Государственным статистическим комитетом РФ. Непосредственный сбор статистических данных ведется региональными статистическими управлениями, отдельными ведомствами, которые представляют информацию в центральные статистические органы. Местные органы осуществляют наблюдение по целому ряду актуальных проблем регионального развития. Предприятия проводят наблюдение в соответствии с разрабатываемой учетной политикой, маркетинговой стратегией фирмы.

Поскольку в РФ становление структуры государственных и ведомственных статистических органов еще незавершено, это ведет к недостаточно эффективному использованию информации в управлении, а нередко и к неиспользованию ее вообще. Все это накладывает отпечаток и на этап статистического наблюдения.

2.2. Программно-методологические и организационные вопросы статистического наблюдения

Статистическое наблюдение проводится строго в соответствии с планом статистических исследований. При подготовке и проведении статистического наблюдения необходимо решить целый ряд вопросов, которые можно подразделить на 1) программно-методологические и 2) организационные. *Программно-методологическими вопросами* определяется цель статистического наблюдения, устанавливается объект и единица наблюдения, разрабатывается инструментарий, обязательный для каждого участника наблюдения, определяется круг признаков, характеризующих единицу наблюдения, по которым производится регистрация данных, или разрабатывается программа. Программой статистического наблюдения называется перечень вопросов или признаков, на которые должны быть получены ответы по единицам наблюдения.

Организационные вопросы охватывают сроки и место проведения наблюдения, положение об организационной стороне наблюдения, подготовку и расстановку кадров и другие, обычно включаемые в организационный план статистического наблюдения. В организационном плане прежде всего указываются органы, выполняющие статистическое наблюдение. Это могут быть органы государственной статистики РФ или конкретные службы. Четко определяются и разграничиваются права и обязанности конкретных учреждений и организаций, участвующих в наблюдении.

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.

Стоящие перед менеджером задачи определяют цель наблюдения. Она может вытекать из постановлений правительственных органов, администрации региона, маркетинговой стратегии фирмы. Общая цель статистического наблюдения состоит в информационном обеспечении управления. Она конкретизируется в зависимости от многих условий.

Цель определяет *объект статистического наблюдения*. Объект статистического наблюдения представляет совокупность явлений, предметов и т.п., охватываемых наблюдением. Объектом может быть население при переписи, предприятия, города, населенные пункты, персонал фирмы или его отдельные категории. Словом, объект наблюдения - это исследуемая статистическая совокупность. Она состоит из отдельных единиц. Единицей совокупности может выступать человек, факт, предмет, процесс и т.п. Единица наблюдения представляет первичный элемент объекта статистического наблюдения. Этот элемент является носителем регистрируемых при наблюдении признаков. *Единица наблюдения* представляет собой элемент совокупности, по которому собираются необходимые данные.

Статистическая совокупность представляет собой широкое понятие. Следовательно, разнообразными могут быть как объекты, так и единицы наблюдения. Так, в торговле статистической совокупностью могут быть торговые предприятия, а единицей наблюдения отдельный акт купли-продажи. Если исследуется качество воды, то объектом исследования является вода, а единицей наблюдения может быть отдельное водохранилище. В других ситуациях объектом исследования являются фермерские хозяйства, а единицей наблюдения могут быть скважины. Выбор объекта и единиц наблюдения зависит от конкретных условий.

В реальной действительности в изучаемых явлениях и процессах могут взаимодействовать несколько совокупностей. В производстве участвуют совокупности рабочих, специалистов, станков, сырья, материалов и т.п. В таких ситуациях необходимо определить объект так, чтобы четко выделить исследуемые совокупности.

В некоторых исследованиях выделяются также технические единицы наблюдения или среда. Так, единицей наблюдения при изучении физического развития детей является ребенок, а технической единицей или средой, помогающей извлечь единицу наблюдения - школы, детские дошкольные учреждения и др. При изучении питания единицей наблюдения является человек, а технической единицей - семья, в которой он проживает.

Единицы наблюдения обладают множеством различных признаков. У работника - это возраст, пол, образование, семейное положение и многое др. *Статистический признак - это конкретное свойство, качество, отличительная черта единицы наблюдения.* Исследователя интересуют определенные признаки, поэтому в программе приводится перечень или формулировка признаков, по которым в процессе наблюдения необходимо зарегистрировать данные. Часто это наиболее существенные или взаимосвязанные признаки.

Четкость при определении единицы наблюдения дает возможность обоснованно определить регистрируемые признаки наблюдения при минимальном количестве признаков, имеющих отношение к изучаемой проблеме, явлению. Иначе исследователь рискует регистрировать единицы наблюдения по многим излишним признакам, собираемым «на всякий случай», что затрудняет проведение наблюдения и усложняет обработку результатов.

Уточнение и формулирование признаков единицы наблюдения производится на основании следующих общих правил:

1. Признаки отбирают с учетом целей исследования, возможностей их обработки и анализа полученных данных;
2. Отобранных признаков не должно быть много;
3. Признаки необходимо комбинировать, чтобы они взаимно дополняли друг друга.
4. Отобранные признаки должны учитывать возможности исследователя. Эти общие подходы при определении признаков единицы наблюдения дополняются конкретными особенностями изучаемых процессов.

Единицу наблюдения не следует путать с *отчетной единицей.* Отчетной является такая единица, от которой получают в установленном порядке отчетные данные по утвержденным формам. Если наблюдение проводится путем представления отчетности, то отчетная единица, в основном, может совпадать с единицей наблюдения.

Отчетная единица называется также информирующей. Возможны ситуации, когда она не совпадает с единицей наблюдения. Например, при изучении платежеспособного спроса единицей наблюдения является покупатель. Если исследователь обращается к нему с вопросом, то покупатель является информирующей единицей. Но за данными о продаже товаров можно обратиться к продавцу, менеджеру, маркетологу. В таких случаях информирующие единицы могут быть разными.

Определив объект, исследователь сталкивается с необходимостью выделить границы, которые четко определяют изучаемую совокупность, явление. Например, если ставится цель изучить деятельность предприятий промышленности, то следует определить, какой отрасли, всех ли форм собственности или какой-то конкретно организационно-правовой формы, по какому критерию будут отбираться предприятия, фирмы. Таким критерием может быть, например, численность работников, объем реализации, время с момента регистрации и др. Установление одного или нескольких взаимодополняющих или взаимоисключающих критериев позволяет четко очертить исследуемую совокупность или установить границы объекта наблюдения. Это означает, что в каждом конкретном случае необходимо решить вопрос: подлежит ли регистрации та или иная единица наблюдения и тем самым обеспечить единообразие в подходе к отбору единиц наблюдения, совокупности.

Вот некоторые примеры из действующей практики. По состоянию на 1996 г., не все предприятия - ТОО обязаны представлять по итогам работы за год в налоговую инспекцию заключение аудиторской фирмы, а только такие, которые имеют численность занятых и объем реализации свыше определенных пределов. Или, например, по действующему законодательству декларацию о доходах не представляют граждане, имеющие совокупный годовой доход менее установленного уровня.

Таким образом, на практике для отграничения объекта устанавливаются конкретные значения или пределы признаков. Такие количественные значения признаков называются *цензом.* Это

ряд признаков, количественное значение которых при проведении статистического наблюдения служит основанием для учета (или неотнесения) единицы в изучаемой совокупности. Если объект наблюдения или исследуемая совокупность будет определен недостаточно четко, то наблюдение может охватить явления, не принадлежащие данной совокупности. И наоборот: некоторая часть единиц может выпасть из обследования. Все это отразится на результатах исследования.

Важным вопросом программы является *момент* или *период наблюдения.* Это время, по состоянию на которое регистрируются данные. Момент наблюдения устанавливается в соответствии с целью, особенностями явления. На практике его называют также *критическим моментом.* Он может выбираться исходя из прагматических установок (в наиболее удобное время и т.п.).

Критический момент особенно важно установить при проведении таких общегосударственных наблюдений, как переписи населения. Во всесоюзной переписи населения 1989 г. критическим моментом было 12 часов ночи с 11 на 12 января. Это означает, что все изменения, происшедшие после 12 часов при проведении наблюдения во внимание не принимаются. В эти дни января наблюдается наименьшее движение населения, когда школьные каникулы окончены, а студенческие каникулы еще не начались. По дням недели критический момент приходится на рабочие дни недели.

Период времени, в течение которого проводится наблюдение, устанавливается в четком временном интервале. При этом некоторые явления, процессы имеют сезонные или другие циклические компоненты. Поэтому время проведения наблюдения должно определяться с учетом таких факторов. Например, наблюдение за пассажиропотоком может осуществляться в течение полной недели, включая праздничные и выходные дни, или по часам в течение суток.

Все это излагается в **программе статистического наблюдения.** В программе содержатся конкретные вопросы, на которые необходимо дать ответ в статистическом формуляре, обосновывается вид и метод проведения наблюдения. Программа, характеризующая главную цель статистического исследования, должна носить комплексный характер, дать ответы на все возникающие в процессе наблюдения вопросы. Недооценка программы или недостаточная ее обоснованность в последующем приводит к значительным потерям в процессе обработки данных и снижает обоснованность выводов и рекомендаций.

На практике исследователь сталкивается с определенными ограничениями временных рамок, финансовых и трудовых затрат при проведении наблюдения. Поэтому конкретный объем программы, ее детализация определяются многими условиями. Например, несущественные признаки, хотя и представляющие интерес, могут исключаться, не фиксироваться. Но во всех случаях следует стремиться к тому, чтобы все вопросы, включаемые в программу, соответствовали цели статистического наблюдения и исследования в целом.

Российская земская статистика отличалась применением широких и детальных программ статистического наблюдения крестьянских хозяйств, что высоко ценилось не только в России, но и за рубежом.

Современные технологии сбора и обработки данных позволяют создавать такие программы статистического наблюдения, которые гибко реагируют на особенности явлений или изучаемых совокупностей. К сожалению, в настоящее время ряд проводимых в стране комплексных статистических наблюдений практически остановлен из-за отсутствия финансовых ресурсов. Это не может не отразиться на общем состоянии статистики РФ и ее роли в управлении обществом.

Для реализации программы статистического наблюдения разрабатывается **статистический инструментарий**. Традиционно он включает инструкции по проведению наблюдения и регистрации данных, бланки документов, в которые заносится информация и другие носители первичной информации. Это анкеты, табеля, формы отчетности, переписные листы. В настоящее время получают распространение технические носители информации, т.е. совершенствуется так называемая безбумажная технология сбора информации и введения ее в ЭВМ.

Собранные в результате данные проходят контроль и преобразуются в форму, воспринимаемую современной техникой, переносятся на машинные носители. Чтобы исключить трудоемкие операции при формировании баз статистических данных, используются новые технологии и документы. Так, данные последних переписей населения заносились на машиночитаемые бланки посредством графических отметок, что позволило вводить информацию в ЭВМ непосредственно с первичных документов.

Особое значение имеет **контроль** получаемых в результате наблюдения данных. Как показывает практика, даже при четко организованном статистическом наблюдении встречаются погрешности, ошибки, требующие исправления. На современном этапе контроль данных в условиях сплошной информатизации приобретает особое значение. Предусматриваются различные методы проверки получаемых данных.

Это прежде всего счетный и логический контроль. На основе *счетного контроля* проверяются итоги и расчет показателей, четко устанавливается наличие ошибки. *Логический контроль* проводится путем сопоставления полученных данных с другими известными признаками, показателями. Возможно сопоставление за прошлый период по одной и той же единице, или за один и тот же период с данными по другой единице наблюдения. В результате выявляются неправдоподобные случаи, т.е. логический контроль выявляет возможность ошибки.

Ошибки наблюдения по источнику происхождения можно подразделить на следующие:

1) преднамеренные (злостные),

2) непреднамеренные.

Непреднамеренные ошибки в свою очередь подразделяются на следующие:

а) случайные,

б) систематические,

в) репрезентативности (представительности).

В соответствии с этими особенностями методы контроля данных могут быть разными.

Преднамеренные ошибки завышают или занижают конкретные значения признака, показателя. Они могут грубо искажать действительное положение. Поэтому преднамеренные ошибки требуют сплошного контроля. Известно, что в РФ наблюдается массовое сокрытие фирмами прибыли от налогообложения. Программа статистического наблюдения предусматривает проверку расчетов прибыли налоговой инспекцией на каждом предприятии. Законом предусматриваются экономические и административные меры, применяемые к предприятиям или лицам, включая уголовную ответственность, за злостные ошибки.

Случайные ошибки чаще связаны с невнимательностью регистратора, небрежностью в заполнении документации, неточностью измерительных приборов и т.п. При достаточно большом числе зарегистрированных явлений они могут нейтрализовать друг друга. И все же подчеркнем, что одной из задач наблюдения является исключение таких ошибок. Для этого необходимо тщательнее подбирать и обучать кадры, участвующие в наблюдении, проверять контрольно-измерительные приборы, проводить массово-разъяснительную работу и т.п.

В целях контроля программа может предусматривать использование опознавательных признаков. Они образуют адресную часть и содержат, например, фамилию, имя и отчество, адрес и т.п. В случае обнаружения ошибки эти признаки позволяют обратиться к источнику информации и исправить ошибку.

Возможно включение в программу контрольных вопросов. Они позволяют, например, уточнить возраст по году рождения или выполнить логический контроль данных.

Непреднамеренные систематические ошибки возникают в самых различных ситуациях, например, при округлении признака в большую или меньшую сторону. Или при изучении бюджетов семей выясняется, что некоторые виды расходов забываются. Такие ошибки требуют корректировки в соответствии с особенностями явлений и процессов. Ошибки представительности, или репрезентативности, определяются специальными методами (см. гл. 6).

Составляя программу, следует оценить степень сложности получения необходимых и объективных данных. Программа не должна быть перегруженной и содержать вопросы, на которые можно получить достоверные ответы.

После того как составлена программа или ее проект, может быть проведено пробное наблюдение. Его цель состоит в уточнении вопросов программы, апробации инструментария, выяснении степени подготовки кадров к проведению наблюдения. Таким наблюдением являются, в частности, пробные переписи населения, пилотаж анкет и др.

В составлении программ крупных общегосударственных статистических наблюдений (переписей населения, обследований и др.) участвуют научные и практические работники. Программа проходит обсуждение, апробацию и утверждается. После этого она готова к реализации. Отдельные предприятия, фирмы, общественные организации проводят статистические наблюдения своими силами или с привлечением других организаций, специализирующихся на конкретных исследованиях.

2.3. Виды статистического наблюдения и их особенности

Отметим определённое 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:

1) по охвату единиц совокупности на сплошное и несплошное;

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

3) по способу организации на специально организованное статистическое наблюдение и отчетность;

4) по источникам сведений на непосредственное наблюдение, документальное наблюдение и опрос.

Такое подразделение на 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.

Сплошное наблюдение охватывает все без исключения единицы совокупности. Таким видом являются переписи населения и др.

Несплошное наблюдение охватывает лишь часть изучаемой совокупности. Эта часть может быть выбрана по-разному. Поэтому несплошное наблюдение можно подразделить на способ основного массива, выборочное и монографическое.

Обследование *основного массива* - это наблюдение за частью наиболее крупных единиц, которые преобладают в исследуемой совокупности. Так, динамика цен может быть исследована по наиболее крупным городам или наиболее крупным оптовым, розничным рынкам.

Выборочное наблюдение предусматривает специальные методы отбора и формирования изучаемой части совокупности (см. гл. 6).

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. В современной практике возможности монографического наблюдения недооцениваются.

Непрерывное статистическое наблюдение, 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.

Единовременное наблюдение проводится по мере возникновения потребности в сборе данных, в исследовании конкретного явления или процесса. Для получения сведений, не собираемых текущей статистикой, проводятся также единовременные учеты как разовое сплошное наблюдение или обследование. Например, единовременный учет народных театров, самодеятельных коллективов. Возможно проведение выборочного единовременного учета. Например, выборочное обследование школ и групп продленного дня и др.

Если наблюдение проводится через определенные промежутки или периоды времени, то такое наблюдение является **периодическим**.

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 кон-

кретных 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 РФ актуальным является изучение статуса занятости населения на основе специальных обследований домашних хозяйств. Это позволило бы скорректировать официальную статистику безработных, которая не дает объективной информации.

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. В нашей стране в последние годы установилась практика проведения переписей населения каждые 10 лет.

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.

Известно, что часть предприятий РФ 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.

Важнейшим общегосударственным статистическим наблюдением является очередная перепись населения. В соответствии с постановлением Правительства Российской Федерации от 15 апреля 1994 г. № 326 Всероссийская перепись населения будет проводиться в 1999 г. Период проведения переписи - с 16 по 23 февраля (включительно). Критическим моментом, или моментом счета населения, устанавливается 12 часов ночи с 15 по 16 февраля, т.е. счет населения проводится на ноль часов 16 февраля (с понедельника на вторник). Перепись проводится специально подготовленными работниками-переписчиками путем опроса населения и записи сведений в переписную документацию.

Программа переписи сочетает сплошное и выборочное наблюдение. Сплошное наблюдение включает такие вопросы, относящиеся ко всему населению, как пол, дата рождения, место рождения, родственное отношение к лицу, записанному первым в домохозяйстве, родной язык, гражданство, брачное состояние, образование, источники средств существования, тип жилого помещения и др.

Выборочное наблюдение, охватывающее 25% населения, содержит такие вопросы, как основное занятие, место работы, число рожденных женщиной детей, продолжительность проживания в данном населенном пункте и др. Выборочное наблюдение, охватывающее 5% населения, включает такие вопросы, как в каком по счету браке состоите или состояли, время вступления во второй брак, для женщин - сведения о каждом рожденном ребенке, число ожидаемых детей. За единицу отбора для выборочного наблюдения принимается жилое помещение - квартира, одноквартирный дом или другое одноквартирное строение, комната в общежитии.

Для отработки организационных и методологических положений Всероссийской переписи населения 1999 г., технологического процесса обработки материалов с применением сканирующих устройств в ноябре 1996 г. в четырех районах (Рыбновском р-не Рязанской обл., Октябрьском р-не г. Ижевск, г. Алейск и Алейском р-не Алтайского края) проводится пробная перепись населения. Для апробации отдельных вопросов переписи могут проходить специальные обследования населения. Проводится периодическая перепись остатков стратегических видов сырья, материалов и т.п. Периодическое повторение наблюдения расширяет возможности в исследовании, позволяет изучать структурные сдвиги, тенденции, более глубоко и всесторонне подойти к исследованию закономерностей развития.

Для совершенствования сбора данных рассматривается вопрос о *модульных обследованиях*. Это специальные наблюдения, позволяющие всесторонне исследовать локальные рынки, отдельные явления и совокупности. В изучении рынка труда модульные обследования позволяют исследовать, например, рынок труда шахтерских регионов, помогают охватить отдельные категории населения - выпускников школ, вузов, работающих пенсионеров и другие с целью их более всестороннего изучения и принятия соответствующих решений.

Реализация программ модульных наблюдений или обследований позволит всесторонне изучить не только безработицу в целом, оценить эффективность политики на рынке труда, но и многие другие важные социально-экономические явления и процессы.

Важнейшим статистическим наблюдением является **отчетность** - составная часть государственной статистики. Отчетность представляется в соответствии с государственной программой статистических работ. Государственная статистическая отчетность на практике включает все виды статистических наблюдений и утверждается Госкомстатом РФ. Отдельные виды отчетности субъектов рынка, характеризующие финансовые результаты деятельности, утверждаются и Министерством финансов РФ.

Специально организованное статистическое наблюдение и отчетность представляют также две основные организационные формы наблюдения.

Отчетность устанавливается для конкретных субъектов рынка - предприятий, организаций. Для нее характерны *обязательность*, т.е. представление по установленной программе на унифицированных формах или бланках в определенные сроки и адреса. Каждый отчет содержит определенные реквизиты - номер или индекс формы, название отчета, отчетный период или на какую дату составляется отчет, название предприятия, организации, административно-территориальной единицы, где расположено предприятие, срок представления отчета, подписи ответственных лиц и др. Отчетность должна вытекать из данных оперативного, бухгалтерского учета или из документации, которая опосредствует деятельность субъекта.

За достоверность данных отчетности предприятия и организации несут ответственность. Так, в соответствии с Законом РФ «О налоге на прибыль предприятий и организаций», если в отчетности налоговой инспекцией выявлено занижение суммы прибыли, эта сумма изымается в бюджет и на предприятие налагается штраф в размере сокрытой от налогообложения прибыли.

В 1992 г. вступил в действие Закон РФ «Об ответственности за нарушение порядка представления государственной статистической отчетности». Органам государственной статистики РФ дано право применять к руководителям и другим должностным лицам предприятий всех форм собственности как административные взыскания, так и налагать штрафы от трехкратного до восьмикратного размера минимальной месячной оплаты труда за нарушение порядка представления государственной статистической отчетности. Эти нарушения включают непредставление данных для проведения статистических наблюдений, несоблюдение сроков представления отчетов, искажение данных и др. За повторные нарушения санкции усиливаются.

По длительности периода можно выделить отчетность периодическую и годовую. Строго говоря, годовая отчетность также является периодической. Выделение ее в особый вид связано в основном с ее расширенным содержанием и некоторыми особенностями. В годовую отчетность могут вноситься изменения. *Периодическая отчетность* представляется ежемесячно, поквартально. Наиболее важные оперативные данные о добыче и производстве энергоресурсов и некоторые другие представляются подекадно, ежесуточно.

Годовая отчетность представляется по итогам работы за год. Она наиболее полно охватывает важнейшие результаты производственно-хозяйственной деятельности.

В зависимости от оперативности представления отчетность может быть *срочной* - передаваемой по телексу, телеграфу, и *почтовой*, высылаемой адресатам по почте. Можно подразделить отчетность на внутреннюю и внешнюю. *Внешняя отчетность* устанавливается государственными органами, министерствами и ведомствами. *Внутренняя отчетность* формируется в соответствии с учетной политикой предприятия, разработка которой в условиях рынка является обязательной.

Как показывает практика, отчетность представляет основной вид статистического наблюдения, используемый для регулирования рынка. Отчетность необходима для оперативного руководства и контроля за ходом производства, состоянием дел как на макро- , так и микроуровне. Например, концерны и другие рыночные структуры широко используют внутреннюю отчетность самостоятельных предприятий для текущего контроля за их деятельностью, оценки эффективности деятельности их администраций, изучения рынка, прогнозирования и других целей.

Отчетность по своему содержанию, форме, срокам представления требует постоянного совершенствования. Так, актуальной проблемой является совершенствование отчетности органов социального обеспечения для выявления наиболее уязвимых групп населения и разработки мер по их социальной поддержке и защите.

В настоящее время часть отчетности представляется на основе выборочного метода. Это позволяет повысить оперативность, достоверность данных и снизить затраты на сбор информации.

Несмотря на то, что в настоящее время в организации статистической отчетности в РФ имеется много нерешенных проблем, в этой области происходят существенные изменения. Важнейшее направление совершенствования отчетности как вида статистического наблюдения связано в переходом экономики РФ на *систему национальных счетов* (СНС). В соответствии с постановлением Верховного Совета РФ от 14 февраля 1992 г. «О переходе Российской Федерации на принятую в международной практике систему учета и статистики» разработана Государственная программа перехода РФ на СНС. Выполнение данной программы призвано обеспечить переход к системе показателей учета и статистики в соответствии с международными стандартами, создать условия для повышения уровня государственного регулирования развития экономики за счет повышения эффективности и достоверности информации, расширения сферы учета и статистики, внедрения более эффективных методов организации статистической работы, включая и наблюдение. СНС представляет собой адекватный рыночной экономике национальный учет, завершаемый на макроуровне системой взаимоувязанных статистических показателей. При переходе на СНС происходят существенные преобразования в порядке сбора информации, ее обработке и представлении органам управления, предпринимательским структурам.

СНС требует реформирования информационной базы и статистического наблюдения. В основе организации международной системы учета и отчетности лежит использование положений, создающих возможности получения такой информации, которая позволяет принимать тактические и стратегические управленческие решения.

Формирующаяся новая информационная база статистических показателей требует обеспечения системности и скоординированности всех этапов статистического исследования на разных уровнях управления. В связи с этим необходима и реконструкция статистического наблюдения, которая включает:

а) введение единого регистра для всех хозяйствующих субъектов;

б) адаптацию международных классификаций, стандартов;

в) сближение учета с международными стандартами.

Единый государственный регистр предприятий и организаций всех форм собственности (ЕГРПО) является новым организационным и технологическим видом сплошного наблюдения. ЕГРПО призван обеспечить органы управления и других пользователей своевременной и достоверной информацией обо всех без исключения субъектах рынка с правом и без права юридического лица. Внедрение ЕГРПО позволит обеспечить сплошное наблюдение по определенному набору статистических показателей по всей совокупности объектов, проходящих государственную регистрацию на территории страны. В сочетании с информацией регистров других информационных систем это обеспечит разработку необходимой информации для исследования социально-экономического положения страны, отдельных отраслей, регионов. Включение субъектов рынка в ЕГРПО требует определения классификационных признаков и присвоения идентификационных кодов в соответствии с действующими общероссийскими классификаторами технико-экономической информации.

Статистическое наблюдение может различаться и по источникам сведений. При проведении наблюдения используется три основных источника сведений или способа регистрации данных - это непосредственное наблюдение, документальное наблюдение и опрос.

Непосредственное наблюдение является надежным источником данных, так как регистрация признаков, фактов производится лично исследователем путем подсчета, обмера, взвешивания и т.п. Однако это требует значительных затрат труда. Кроме того в процессе наблюдения нередко регистрируются уже прошедшие случаи, факты. И такой подход оказывается нереальным. Возможны ситуации, когда непосредственно «наблюдать» некоторые признаки, факты просто невозможно (например, требуется установить родной язык или семейное положение, национальность и др.). Поэтому такое наблюдение используется лишь в тех случаях, когда для его проведения имеются все необходимые условия.

Документальное наблюдение основывается на различных документах, таких как счета клиентов, рекламации на качество продукции, свидетельства о рождении и др. Это надежный метод получения данных. На практике важнейшие виды статистического наблюдения основываются на документальном способе сбора данных. На документах базируется и отчетность.

Получает распространение наблюдение на основе **опросов**. Это так называемый анкетный способ. Анкеты могут заполняться самими опрашиваемыми или специальным лицом. Возможно получение информации по конкретным вопросам от специальных корреспондентов. Таким способом фирмы получают, например, информацию от покупателей (потребителей) о своих товарах. При проведении анкетного наблюдения особое внимание уделяется анкете. Она должна быть четкой, не допускать двойного толкования вопросов и ответов. Ответы на анкету должны быть даны добровольно. Анкета распространяется среди определенного круга единиц совокупности - лиц, семей, фирм и др.

Опросы позволяют при достаточно высоком уровне их проведения получить достоверную и оперативную информацию по таким конкретным социально-экономическим явлениям и процессам, по которым другие методы практически неприменимы. Опросы, в частности, используются при проведении переписей населения, бюджетных обследований семей, покупателей, молодежи, студентов и др. Получили распространение социологические опросы, например, о качестве жизни, здоровье и т.п.

С целью проведения наиболее важных исследований развивается сотрудничество статистических органов РФ с международными организациями, со странами СНГ. Это позволяет использовать накопленный зарубежный опыт в проведении отдельных видов статистических наблюдений, скоординировать исследования и повысить эффективность использования получаемой в результате наблюдения информации. Так, Межгосударственный статистический комитет СНГ в сотрудничестве с Международной организацией труда (МОТ) содействовал организации специальных обследований рабочей силы. С 1992 г. в РФ было проведено три специальных обследования, которые позволили получить данные об основных категориях населения - занятых, безработных и экономически неактивном населении. Предполагается решить вопрос о периодичности таких обследований, больше внимания уделить сбору данных о неформальном рынке труда, неполной занятости и т.п.

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.

Контрольные вопросы к главе 2

1. В чем состоит роль статистического наблюдения в комплексном экономико-статистическом исследовании?
2. Каковы основные направления реконструкции статистического наблюдения с целью совершенствования информационной базы статистики?
3. Назовите основные программно-методологические вопросы статистического наблюдения.
4. Какие конкретные виды статистического наблюдения используются для сбора данных?
5. Какие изменения происходят в статистической отчетности предприятий на современном этапе?
6. Какие задачи решаются на основе специальных статистических наблюдений?
7. Поставьте цель, определите объект, единицу наблюдения, конкретные признаки. Составьте программу специального статистического наблюдения.

Глава 3
ГРУППИРОВКА СТАТИСТИЧЕСКИХ ДАННЫХ И ЕЕ РОЛЬ В АНАЛИЗЕ ИНФОРМАЦИИ

3.1. Виды и задачи группировок

Одним из основных и наиболее распространенных методов обработки и анализа первичной статистической информации является группировка. Понятие статистической группировки в широком смысле слова охватывает целый комплекс статистических операций, направленных на объединение зарегистрированных при наблюдении единичных случаев в группы, сходные в том или ином отношении, поскольку целостную характеристику совокупности необходимо сочетать с характеристикой основных ее частей, классов и т.д.; подсчет итогов по выделенным группам и по всей совокупности в целом и, наконец, оформление результатов группировки в виде *статистических таблиц*. При составлении плана статистического наблюдения определяется очередность обработки материалов, разрабатываются макеты сводных таблиц, на основе которых дается характеристика размеров, структуры и взаимосвязи изучаемых явлений. Указывается также, кто и в какие сроки осуществляет сводку, каким способом, куда поступают сводные данные и кто проводит их дальнейшую обработку.

Под **группировкой** в статистике понимают расчленение единиц статистической совокупности на группы, однородные в каком-либо существенном отношении, и характеристику таких групп системой показателей в целях выделения типов явлений, изучения их структуры и взаимосвязей.

Метод группировок является основой применения других методов статистического анализа основных сторон и характерных особенностей изучаемых явлений. По своей роли в процессе исследования метод группировок выполняет некоторые функции, аналогичные функциям эксперимента в естественных науках: посредством группировки по отдельным признакам и комбинации самих признаков статистика имеет возможность выявить закономерности и взаимосвязи явлений в условиях, в известной мере ею определяемых. При использовании метода группировок появляется возможность проследить взаимоотношение различных факторов и определить силу их влияния на результативные показатели.

В развитие метода группировок огромный вклад внесли российские статистики. Им принадлежит первенство в применении комбинационных таблиц, в разработке классификации таблиц и в проведении многочисленных группировок материалов аграрных переписей и обследований, которые оказали благотворное влияние на развитие других отраслевых статистик и общей методологии. Исключительное значение метода группировок в статистике было сформулировано выдающимся русским ученым Д.П. Журавским (1810-1856). Как уже упоминалось в гл. 1, он определил статистику как науку категорического вычисления, т.е. как науку о счете по категориям, по группам. В этом определении подчеркивается одна из специфических черт статистической методологии.

Очень важную роль в развитии теории и практики группировок сыграли работы земских статистиков. Метод группировок начинает рассматриваться ими как специальный научный метод обработки данных, а не просто как подсчет численности групп, естественно выделенных в самом процессе статистического наблюдения. Земская статистика использовала систему показателей для изучения социально-экономической дифференциации и анализа взаимосвязей отдельных факторов крестьянских хозяйств. Большие достижения в области применения метода группировок в статистике отраслей народного хозяйства России характерны для XX столетия.

Огромное значение и роль группировок в статистическом исследовании вытекает из характера объекта статистики, его специфики. Явления общественной жизни, изучаемые статистикой, отличаются многообразием форм и стадий развития, они состоят из существенно различающихся частей, обладающих многими специфическими свойствами.

Изучая количественную сторону массовых общественных явлений в неразрывной связи с их качественными особенностями, статистика стремится показать совокупность явлений в дифференциации, в многообразии их типов, рассмотреть взаимосвязи и соотношения между последними. С помощью метода группиро-

вок решаются сложные задачи статистического анализа. Учитывая, что необходимость группировки обусловливается прежде всего наличием качественных различий между изучаемыми явлениями, первую задачу группировок можно сформулировать как задачу выделения в составе массового явления тех его частей, которые однородны по качеству и условиям развития, в которых действуют одни и те же закономерности влияния факторов. В результате такой группировки выделяются социально-экономические типы (а отсюда и название группировки – типологическая) как выражение конкретного общественного процесса, его форм и разветвлений, как выражение существенных черт, общих для множества единичных явлений.

С задачами **типологической группировки** тесно связаны и две другие задачи группировок: характеристика структуры и структурных сдвигов в исследуемой совокупности и выявление взаимосвязи между отдельными признаками изучаемого явления.

Примерами типологических группировок могут служить группировки хозяйственных объектов по формам собственности; населения по общественным группам; работников на занятых преимущественно физическим и преимущественно умственным трудом и т.д.

Методология типологических группировок определяется тем, насколько ясно выступают качественные отличия в изучаемых явлениях. Например, при группировке отраслей промышленности по экономическому назначению продукции выделяются отрасли, производящие средства производства и отрасли, производящие предметы потребления, в макроструктуре розничного товарооборота выделяются продовольственные и непродовольственные товары. В большинстве случаев качественные отличия между явлениями не выступают столь отчетливо. Например, выделение в отраслях промышленности крупных, средних и мелких предприятий является достаточно сложной в методологическом отношении проблемой. В подобных случаях после предварительной наметки возможных типов на основе четкой формулировки познавательной задачи необходимо определить те признаки, которые будут положены в основу выделения типов, так называемые группировочные признаки.

Выбор **группировочных признаков** всегда должен быть основан на анализе качественной природы исследуемого явления. Всесторонний теоретико-экономический анализ сущности и закономерностей развития явления должен быть направлен на то, чтобы в соответствии с целью и задачами исследования положить в основание группировки существенные признаки. В зависимости от конкретных условий группировочные признаки должны видоизменяться, т.е. должны быть специализированы при выделении одного и того же типа в различных условиях. Так, в настоящее время в промышленности при группировке предприятий по размерам могут использоваться показатели производственной мощности предприятий, стоимости промышленно-производственных основных фондов, численности промышленно-производственного персонала, объем потребленной энергии. Если для трудоемких отраслей в качестве группировочного признака целесообразно использовать численность работников, то для фондоемких отраслей - стоимость промышленно-производственных фондов; для энергоемких отраслей - количество потребленной электроэнергии.

Вместе с тем следует учитывать, что полную характеристику каждому типу явлений можно дать при использовании *системы признаков (системы показателей)*. Только совокупность признаков позволяет отобразить процессы развития, всесторонне выявить реальные связи, взаимоотношения отдельных сторон процесса.

Использование одного признака, характеризующего лишь одну сторону, одну черту в развитии явления, может привести к искажению действительности, поскольку в последней, как правило, переплетаются различные, подчас противоположные тенденции и направления. Множественность признаков, характеризующих объекты, является следствием их многосторонности и многообразия реальных связей между объектами.

Если в основу группировки положено несколько признаков, такую группировку называют *сложной*. Сложная группировка может выполняться как комбинационная и как многомерная. При использовании *комбинационной группировки* группы, выделенные по одному из признаков, затем подразделяются на подгруппы по другому признаку, в свою очередь подгруппы могут быть разделены по следующему признаку и т.д. Общее число выделенных групп будет равно произведению числа группировочных признаков на число выделенных групп в каждом из них. В случае, если группировка осуществляется не последовательно по отдельным признакам, а одновременно по комплексу признаков, ее называют *многомерной* (подробнее см. 3.3).

В зависимости от вида группировочных признаков различают группировки по количественным и качественным признакам.

Качественный признак отражает определенные свойства, качества данного явления и записывается в виде текста. Если качественный признак имеет мало разновидностей, то количество групп определяется числом этих разновидностей. Таковы, например, группировки населения по полу, семейному положению, образованию, деление населения на городское и сельское и т.д.

Но нередки случаи, когда качественный признак имеет большое число разновидностей и перечислить их все не представляется целесообразным. Например, профессии рабочих, номенклатура выпускаемой продукции, виды основных фондов и т.п. В таких случаях разрабатывают *классификацию разновидностей*, т.е. сходные по своим особенностям разновидности объединяются в группы (классы). Под классификацией обычно понимается более устойчивое разграничение единиц наблюдения, чем при группировке. Используются классификации в течение длительного времени, хотя со временем, отразив происходящие изменения в объекте наблюдения, классификации могут подвергаться более или менее существенным изменениям. Утверждаются классификации, как правило, в качестве национального или международного стандарта. Широко используются классификации отраслей народного хозяйства, отраслей промышленности, основных фондов, промышленно-производственного персонала и т.д. На рис. 3.1 приведена классификация форм собственности, действующая в Российской Федерации с 1993 г.

Рис. 3.1. **Классификация форм собственности хозяйственных субъектов РФ**

Любая классификация может состоять из нескольких уровней, т.е. выделенная на первом этапе классификационная группа может в дальнейшем детализироваться и т.д. Необходимость и степень детализации классификационных групп определяется целью и задачами статистического исследования.

Рассмотрим в качестве примера классификацию элементов затрат рабочего времени. В целях изучения структуры затрат внутрисменного рабочего времени можно ограничиться такими видами затрат: 1) время работы; 2) время, неиспользованное по уважительным причинам; 3) потери рабочего времени. Однако, если ставится задача разработки мероприятий, направленных на сокращение потерь рабочего времени, необходимо использовать более развернутую классификацию фонда рабочего времени и, в частности, третьего элемента вышеприведенной классификации. Так, могут быть выделены потери рабочего времени по организационно-техническим причинам и потери по вине рабочего. В свою очередь, в составе потерь по организационно-техническим причинам могут быть выделены простои из-за отсутствия материалов, заготовок, инструмента и технической документации; простои из-за текущего ремонта оборудования и т.д.

Всестороннее статистическое изучение социально-экономических процессов и явлений наиболее плодотворно в том случае, если в основе его лежит система группировок. **Система группировок** – это ряд взаимосвязанных статистических группировок по наиболее существенным признакам, всесторонне отражающим важнейшие стороны изучаемых явлений.

Так, статистика товарооборота внешней торговли опирается на выделение в составе общего объема внешнеторгового оборота объемов экспорта и импорта. Их более подробная характеристика и анализ опираются на ряд важных группировок по следующим направлениям: вид сделки (форма продажи), география связей, товарная специализация, отраслевое происхождение товаров и т.д. С 1991 г. учет экспорта и импорта основывается на базе Гармонизированной системы описания и кодирования товаров (ГС) и имеет 5 ступеней классификации: разделы, группы, подгруппы, товарные позиции и субпозиции. В составе 21 раздела выделяется 96 групп по принципу выделения последовательных стадий обработки товаров: сырье→ полуфабрикаты→готовые изделия. При построении 33 товар-

ных подгрупп, 1241 позиции и 5019 субпозиций в каждой группе применяется своя последовательность группировочных признаков, среди которых основными являются: назначение товара, степень его обработки, вид материала, из которого он изготовлен, удельный вес товара в мировой торговле. Товарные разделы Гармонизированной системы описания и кодирования товаров приведены в Приложении IX.

Товарная классификация является средством упорядочения информации о внешнеэкономических связях, служит основой их системного анализа и обеспечивает сопоставимость национальных статистических данных о внешней торговле на международном уровне.

Приведем пример применения *комбинационной группировки* в анализе результатов деятельности коммерческих банков (см. табл. 3.1).

Коммерческие банки обязаны создавать резерв на возможные потери по ссудам в соответствии со степенью риска непогашения ссуды. В зависимости от степени риска все ссуды делят на пять групп: стандартные, нестандартные, сомнительные, опасные и безнадежные. Для отнесения ссуды к той или иной группе риска используют группировку выданных ссуд банка по двум направлениям, во-первых, в зависимости от реального обеспечения (залог, гарантия банка, гарантия Правительства Российской Федерации); во-вторых, в зависимости от количества дней просрочки выплаты ссуды (см. графу 1 табл. 3.1).

При характеристике обеспеченности ссуды выделяются три группы: 1) обеспеченные ссуды; 2) недостаточно обеспеченные ссуды; 3) необеспеченные ссуды.

В первую группу включаются ссуды, имеющие ликвидный залог, стоимость которого равна 100% от размера ссуды, либо имеется банковская гарантия или гарантия правительства.

Недостаточно обеспеченная ссуда – имеется частичное обеспечение по стоимости не менее 60% от размера ссуды.

Необеспеченная ссуда не имеет обеспечения, либо обеспечение менее 60% от размера ссуды.

В таблице приведены результаты группировки всех кредитов на общую сумму 140 млн. руб. по состоянию на конец июля 1995 г. по одному из московских банков.

Таблица 3.1

Распределение кредитного портфеля банка по группам риска на конец июля 1995 г. (млн руб.)

Количество дней просроченной задолженности	Обеспеченность ссуды	Группы риска по ссудам					Итого по группе
		стандартные	нестандартные	сомнительные	опасные	безнадежные	
1	2	3	4	5	6	7	8
Возврат ссуды в срок	обеспеченная	62					80
	недостаточно обеспеченная	11					
	необеспеченная	7					
до 30	обеспеченная	18					24
	недостаточно обеспеченная		4				
	необеспеченная			2			
30–60	обеспеченная		10				15
	недостаточно обеспеченная			3			
	необеспеченная				2		
60–90	обеспеченная			9			17
	недостаточно обеспеченная				7		
	необеспеченная					1	
более 90	обеспеченная					1	4
	недостаточно обеспеченная					2	
	необеспеченная					1	
Итого в группе:		98	14	14	9	5	140

* Пример предоставлен доц. Н.Е. Варакиной.

К первой группе риска «Стандартные ссуды» относятся ссуды, по которым своевременно и в полном объеме погашается основной долг, а также просроченные до 30 дней обеспеченные ссуды. По нашему примеру «Стандартные ссуды» составляют 98 млн руб., или 70% от общего объема выданных кредитов ($\frac{98}{140} \cdot 100\%$).

Во вторую группу риска «Нестандартные ссуды» относятся просроченные до 30 дней недостаточно обеспеченные ссуды, а также просроченные от 30 до 60 дней обеспеченные ссуды. Таких ссуд насчитывается на 14 млн руб., или 10% от общего объема кредитов банка.

К третьей группе риска «Сомнительные ссуды» относят просроченные до 30 дней необеспеченные ссуды; просроченные от

30 до 60 дней недостаточно обеспеченные ссуды, а также просроченные от 60 до 90 дней обеспеченные ссуды. Эти ссуды составляют 14 млн руб., или 10% от общей суммы выданных кредитов в отчетном периоде.

В четвертую группу «Опасные ссуды» включаются просроченные от 30 до 60 дней необеспеченные ссуды, а также просроченные от 60 до 90 дней недостаточно обеспеченные ссуды. Такие ссуды в примере составляют 9 млн руб., или 6,4% от выданных банком кредитов.

К пятой группе риска «Безнадежные ссуды» относятся просроченные от 60 до 90 дней необеспеченные ссуды и все ссуды, просроченные свыше 90 дней. Сумма ссуд в этой группе риска, по приведенным данным, составила 5 млн руб., или 3,6% от общего объема выданных кредитов.

Поскольку для каждой группы риска устанавливается определенный процент резервирования средств (см. графу 2 табл. 3.2), приведенные данные в итоговой строке табл. 3.1. служат основой определения необходимого резерва на случай непогашения ссуд.

Таблица 3.2

Группа риска	Процент отчислений в резервный фонд	Размер ссуд (млн руб.)	Размер отчислений в резервный фонд (млн руб.)
1	2	3	4 (гр.2 x гр.3)
1	2	98	1, 96
2	5	14	0, 70
3	30	14	4, 20
4	75	9	6, 75
5	100	5	5,00
Итого:		140	18,61

В табл. 3.1 кроме группировок по качественным признакам была приведена и группировка по количественному признаку для выделения ссуд по числу дней просрочки. *Группировки по количественным признакам* чрезвычайно разнообразны. Таковы, например, группировки населения по возрасту, группировка рабочих по разрядам, по степени выполнения норм, предприятий - по численности работников, стоимости продукции и основных фондов, банков - по размеру активов и т.д. В ряде случаев группировки, на первый взгляд казалось бы качественные, в действительности основываются на количественных признаках. Например, при группировке кредитов по срокам, на которые они предоставляются, выделяют: 1) краткосрочные, 2) среднесрочные и 3) долгосрочные. При отнесении в ту или иную группу руководствуются тем, что первая группа включает кредиты на срок до 1 года, вторая - от 1 года до 5 лет и третья - кредиты на срок свыше 5 лет.

При группировке по количественному признаку, варьирующему в широких пределах, нередко возникает задача определения числа групп, на которые следует разбить весь диапазон изменения количественного признака, и в соответствии с установленным числом групп определить интервалы группировки.

Подытоживая вышеизложенное, можно перечислить те методологические проблемы, решение которых необходимо при практическом применении метода группировок: 1) выбор группировочного признака или их комбинации, 2) определение числа групп и величины интервалов группировки, 3) установление применительно к конкретной группировке состава тех показателей, которыми должны характеризоваться выделенные группы, 4) составление макета таблицы, в которой должны быть представлены результаты группировки.

3.2. Некоторые вопросы техники выполнения группировки

Остановимся подробнее на группировках по количественному признаку. Вопрос о количестве групп, а следовательно, и об интервалах группировки решается по-разному при типологических группировках и при выделении групп внутри типов. Изучая количественную сторону массовых общественных явлений, статистика, опираясь на конкретные положения экономической теории, должна в процессе группировки наметить точки перехода количества в новое качество: на основе анализа количественных изменений группировочных признаков обозначить точки перехода одного качества в другое. При типологической группировке интервалы должны намечаться таким образом, чтобы они отграничивали социально-экономические типы, установленные на основе экономической теории.

Теоретико-экономический анализ изучаемого явления должен быть предпосылкой научной статистической группировки, но вместе с тем использование аппарата современных статистических методов позволяет количественно оценить степень однородности выделенных групп, производить отбор существенных группировочных признаков, совершенствовать методику определения величины интервалов группировки. Количество выделяемых групп может зависеть и от характера вариации изучаемого показателя. Если в качестве группировочного используется *дискретный признак*, т.е. признак, способный принимать только некоторые определенные значения (например, целые), то число выделяемых групп соответствует количеству вариантов значений признака, если их число не очень велико. Например, распределение рабочих предприятия по тарифным разрядам, группировка семей по размеру, распределение студентов по успеваемости (по баллам) и т.д. В этом случае число групп определяется числом реально существующих вариантов значений изучаемого признака. Но дискретный признак может иметь и очень большое число вариантов значений, которые не всегда могут повторяться. В таких случаях варианты значений объединяются в группы. Например, при группировке предприятий по численности рабочих, по числу единиц установленного металлорежущего оборудования и т.д.

При непрерывном характере вариации группировочного признака, когда в определенных пределах признак может принимать любое значение (целое и дробное), весь диапазон изменения признака также разбивается на интервалы.

Приведем пример, когда для группировки использован непрерывный признак (см. табл. 3.3).

Таблица 3.3

Распределение коммерческих банков по объявленным уставным фондам

Группы коммерческих банков по объявленным уставным фондам, млн руб.	Количество коммерческих банков, в % к итогу	
	на 1.03.94	на 1.03.95
до 100	10,1	3,4
100–500	64,7	42,3
500–1000	14,1	14,8
1000–5000	9,3	31,4
5000 и более	1,8	8,1
Итого	100,0	100,0

Источник. Статистическое обозрение. 1995. № 4.

Группы банков по объявленным уставным фондам отмечены с помощью неравных интервалов, у которых разности между верхней и нижней границами неодинаковы в разных группах. Так, во второй группе величина интервала равна 400 млн руб., в третьей -

500 млн руб., в четвертой - 4000 млн руб., т.е. в приведенной группировке величина интервала постепенно увеличивается, а в последней группе верхняя граница не указывается совсем. Интервалы, в которых указана лишь одна граница (верхняя или нижняя) называются *открытыми*. В приводимом примере открытыми являются первый (указана верхняя граница) и последний интервал (указана нижняя граница), остальные интервалы являются закрытыми, так как в них указана и верхняя и нижняя граница. Неравные интервалы применяются при группировках, которые охватывают массу единиц неоднородной совокупности с неравномерными и значительными колебаниями признака. При этом учитывается, что небольшое изменение величины группировочного признака в низших группах отражает более существенные изменения в характере группы, чем такое же по абсолютной величине различие в высших группах. Так, нельзя считать однозначными различия банков с объявленным уставным фондом от 100 до 500 млн руб. и изменением уставного фонда на 400 млн руб. в группе с уставным фондом от 1000 до 5000 млн руб. – во втором случае банки гораздо меньше отличаются друг от друга. Кроме того, в экономических исследованиях, как правило, с увеличением значений признака уменьшается число единиц, варианты значений признака у которых находятся в интервале одной и той же длины. Поэтому применение равных интервалов может привести к тому, что число единиц в группах будет слишком мало, а поэтому закономерности в изменениях взаимосвязанных факторов могут отчетливо и не проявиться. Открытые интервалы используются обычно в тех случаях, когда признак в выделяемой группе единиц изменяется неравномерно и в широких пределах и когда отсутствуют качественные различия у отдельных единиц, включаемых в группу.

Границы интервалов проводимой с определенной целью группировки следует видоизменять в зависимости от особенностей изучаемой совокупности и временного периода, к которому относятся данные. Так, группируя предприятия по стоимости основных фондов для выделения мелких, средних и крупных предприятий, применяют различные интервалы в разных отраслях промышленности. Вместе с тем, развитие исследуемых явлений во времени требует изменения интервалов, т.е. предприятия, которые могли быть отнесены к крупным в изучаемой отрасли в 40-е и 50-е годы, в настоящее время скорее можно отнести к средним и даже мелким. Указанные обстоятельства находят отражение в изменении интервалов группировки.

Внутри типичных групп для характеристики количественных различий единиц, составляющих соответствующую группу, могут быть применены равные интервалы. В этом случае величина интервала определяется делением размаха вариации на принятое число групп. Более подробно эти вопросы рассмотрены в гл. 5.

При анализе разнородных данных, например, при анализе материала, собранного в различные периоды времени, относящегося к различным отраслям народного хозяйства, возникает необходимость применения **вторичной группировки**. Кроме того, методом вторичной группировки пользуются для того, чтобы показать интенсивность развития процессов и явлений в разнообразных условиях.

3.3. Статистические таблицы

Результаты группировок представляются в статистических таблицах. Статистическая таблица - форма рационального и наглядного изложения цифровых характеристик исследуемых явлений и его составных частей. Статистическое обобщение информации и представление ее в виде сводных статистических таблиц дает возможность характеризовать размеры, структуру и динамику изучаемых явлений. Часто к статистической таблице дается общий заголовок, в котором указывается содержание таблицы, место и время, к которым относятся приводимые в таблице данные, а также единицы измерения, если они одинаковы для всех приведенных сведений. Основные элементы статистической таблицы - подлежащее и сказуемое. *Подлежащим таблицы* являются единицы статистической совокупности или их группы. *Сказуемое таблицы* отражает то, что в ней говорится о подлежащем с помощью цифровых данных. В зависимости от строения подлежащего все статистические таблицы можно разделить на три группы:

1. Таблицы простые, или перечневые, в которых содержатся сводные показатели, относящиеся к перечню единиц наблюдения, или к перечню хронологических дат или территориальных подразделений. Соответственно таблицы могут быть названы простыми перечневыми, хронологическими или территориальными;

2. Таблицы групповые, в которых статистическая совокупность расчленяется на отдельные группы по какому-либо одному признаку, причем каждая из групп может быть охарактеризована рядом показателей;

3. Таблицы комбинационные, в которых совокупность разбита на группы не по одному, а по нескольким признакам.

Выбор типа таблицы зависит всегда от цели ее построения. Если таблицы используются для практических нужд планирования и управления, то в них должны содержаться сведения по тем частям, в разрезе которых ведется планирование и управление. Чаще всего этой задаче соответствуют простые таблицы, используются также и групповые. Если же ставится задача более глубокого познания исследуемого объекта, то используются групповые и комбинационные таблицы.

В простых таблицах помещаются данные по различного рода организациям: предприятиям, стройкам, учреждениям, министерствам и т.д., имеющие, как правило, познавательное значение.

Примером простой таблицы является табл. 3.4, где приведена динамика цен на муниципальные квартиры в разных районах Москвы за 5 месяцев 1995 г.

Таблица 3.4

Динамика цен 1 кв. м муниципального жилья в Москве в 1995 г. (млн руб.)

Местоположение жилья	Январь	Февраль	Март	Апрель	Май
Митино	3,4	3,7	4,0	4,8	5,2
Южное Бутово	3,2	3,5	3,8	4,5	4,9
Северное Бутово	3,5	3,8	4,1	4,8	5,8
Отрадное	4,5	4,9	5,3	6,0	6,5
Веерная улица	5,3	5,5	5,9	7,0	7,6
Жулебино	3,4	3,7	4,0	4,5	4,9

Сочетание территориальной группировки с данными за несколько месяцев 1995 г. позволяет получить весьма полезную информацию. По данным табл. 3.4, мы видим наличие вариации цен за 1 кв. м муниципального жилья по районам города Москвы и, если в январе цена варьировала от 3,2 млн руб. (Южное Бутово) до 5,3 млн руб. (ул. Веерная), то в мае границы вариации существенно сдвинулись вправо: от 4,9 млн руб. (Северное Бутово) до 7,6 млн руб. (ул. Веерная). Кроме того, можно сделать вывод о наличии общей для всех районов Москвы тенденции неуклонного роста цен муниципального жилья. Данные подобного рода таблиц могут использоваться для принятия оперативных решений, например оценки того, как изменятся затраты на приобретение жилья, если задержаться с принятием решения хотя бы на месяц.

В отличие от простых групповые и комбинационные таблицы обладают важными аналитическими свойствами: они позволяют производить наглядные сравнения и вскрывать существенные связи и различия в развитии явлений. Идея комбинационной таблицы состоит в том, что каждую из групп в групповой таблице разбивают на подгруппы по какому-либо признаку; выделенные подгруппы могут дальше расчленяться по следующему признаку и т.д. (пример комбинационной таблицы приведен в табл. 3.1).

Результаты комбинационной группировки по большому количеству признаков даже при небольшом числе интервалов группировки становятся трудно обозримыми, и таблица теряет свое важнейшее преимущество - наглядность. Поэтому нецелесообразно составлять комбинационные таблицы по сочетанию более чем трех признаков и при количестве интервалов более четырех. Использование комбинационных таблиц и системы взаимосвязанных группировок позволяет провести глубокий и всесторонний анализ сложных общественных явлений.

Группировка, осуществляемая не последовательно по отдельным признакам, как при комбинационной группировке, а одновременно по комплексу признаков, называется **многомерной**. Как уже сказано, характеристика одной и той же качественной стороны изучаемого явления может быть дана с помощью набора признаков. Например, для характеристики технического уровня развития предприятий могут быть использованы следующие показатели: удельный вес активной части промышленно-производственных основных фондов, удельный вес автоматических машин и оборудования в составе рабочих машин и оборудования; электровооруженность труда, машиновооруженность рабочих; степень охвата механизированным трудом, коэффициент обновления машин и оборудования и т.д.

Характеризуя таким образом каждую единицу совокупности набором признаков, можно рассматривать эту единицу как точку в m-мерном пространстве, а задача многомерной группировки будет состоять в выделении точек, составляющих однородные группы единиц. Мерой близости (сходства) между единицами могут служить различные критерии. В зависимости от выбранного критерия существуют различные методы многомерной группировки.

Применение методов многомерной группировки связано с большой вычислительной работой и требует использования электронной вычислительной техники. С помощью специальных алгоритмов на ЭВМ осуществляется формирование групп, в которых единицы совокупности объединяются на основании близости по всему комплексу признаков. В табл. 3.5 приведены результаты группировки предприятий отрасли по уровню технического развития и производительности труда.

Таблица 3.5

Распределение предприятий по уровню технической оснащенности и эффективности использования живого труда

Группы предприятий по техническому уровню	Группы предприятий по уровню производительности труда			Итого
	ниже среднего по отрасли	среднего по отрасли	выше среднего по отрасли	
А	1	2	3	4
Ниже среднего уровня	9	8	8	25
Среднего уровня	5	6	1	12
Выше среднего уровня	6	3	7	16
Итого	20	17	16	53

Примечание. Выделение однородных по техническому уровню групп предприятий было осуществлено с помощью метода кластерного анализа по восьми показателям технического уровня развития.

Анализ данных таблицы позволяет выделить группы предприятий, добившихся наибольшего эффекта в своей деятельности и группы предприятий, располагающих резервами роста производительности труда за счет лучшего использования технического потенциала. Это прежде всего те шесть предприятий (группа 3.1),* которые имеют высокий технический потенциал, но эффективность использования живого труда здесь ниже среднего по отрасли уровня. В то же время восемь предприятий с низким уровнем технического развития имеют уровень производительности труда выше среднего по отрасли, что позволяет говорить о высокоэффективной деятельности предприятий группы 1.3.

Такие группировки дают возможность, безусловно, лишь в общем оценить результаты деятельности предприятий соответствующих групп. Однако уже на их основе очевидна необходимость дифференцированного подхода к оценке результатов деятельности предприятий отрасли, располагающих примерно одинаковым техническим потенциалом, но различающихся уровнем эффективности использования труда.

* Первая цифра означает номер строки, вторая номер столбца.

3.4. Графическое представление статистических данных

Графики, наряду со статистическими таблицами, являются важным средством выражения и анализа статистических данных, поскольку наглядное представление облегчает восприятие информации. 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.

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», «Exel».

Для графического изображения статистических данных используются самые разнообразные виды графиков. Классификация графиков представлена на рис. 3.2. 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 соответствии 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».

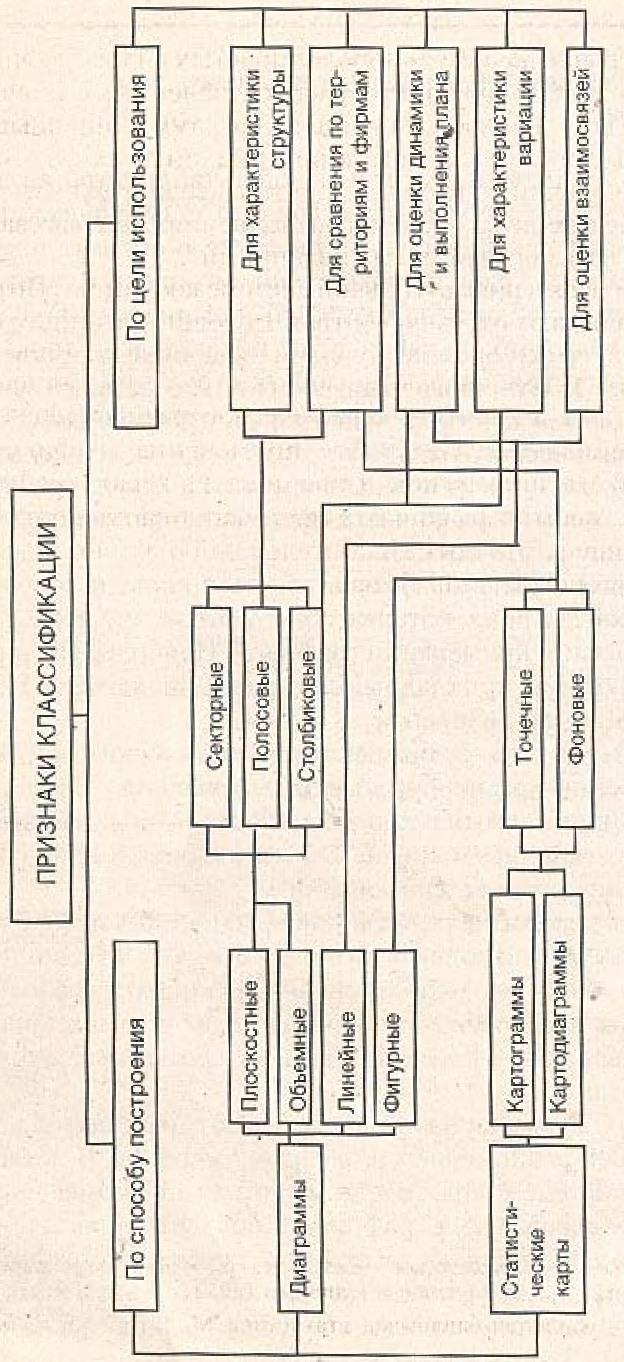

Рис. 3.2. Классификация графиков

Рассмотрение начнем с наиболее простых видов графиков и в то же время достаточно широко распространенных в экономико-статистическом анализе - линейных диаграмм. **Линейные диаграммы** применяются для характеристики динамики, т.е. оценки изменения явлений во времени; для характеристики вариации в рядах распределения; для оценки выполнения плановых заданий; для оценки взаимосвязи между явлениями*.

Они строятся в прямоугольной системе координат. По оси абсцисс откладывают отрезки, соответствующие датам или периодам времени, по оси ординат - уровни ряда динамики или темпы их изменения. Полученные точки соединяют отрезками в виде ломаной линии. Каждая точка линейной диаграммы соответствует уровню динамического ряда (или темпу его изменения) на определенный момент или за период времени. На одном графике может быть размещено несколько диаграмм, что позволяет сравнивать динамику различных показателей, либо одного показателя по разным регионам или странам.

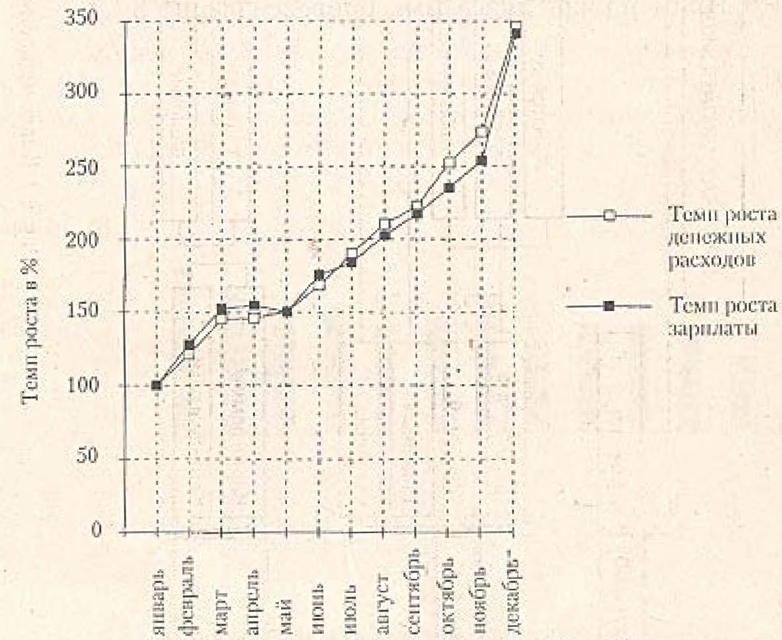

Рис. 3.3. Темпы роста денежных расходов и заработной платы населения России в % к январю 1994 г.

Источник. Бюллетень банковской статистики. М., 1995. № 20. С. 5.

* Подробнее см. в соответствующих главах.

Характер изменений в динамике заработной платы и денежных расходов населения Российской Федерации за 1994 г. по месяцам показан на рис. 3.3. Для анализа были использованы базисные темпы роста к январю 1994 г. Из графика видно, что в течение года наблюдалась тенденция к росту заработной платы и денежных расходов. В то же время с января по май рост зарплаты опережал рост денежных расходов. Начиная с июля заметно непрерывное ускорение роста денежных расходов по сравнению с ростом заработной платы.

Для тех же целей, а именно анализа динамики социально-экономических явлений, оценки выполнения плана и характеристики вариации в рядах распределений могут использоваться также **столбиковые диаграммы**. Столбики располагаются вплотную или раздельно на одинаковом расстоянии. Они имеют одинаковое основание, а их высота должна быть пропорциональна числовым значениям уровней признака. По высоте столбиков этой диаграммы определяют соотношение между уровнями изучаемых показателей. Пример такой диаграммы приведен на рис. 3.4.

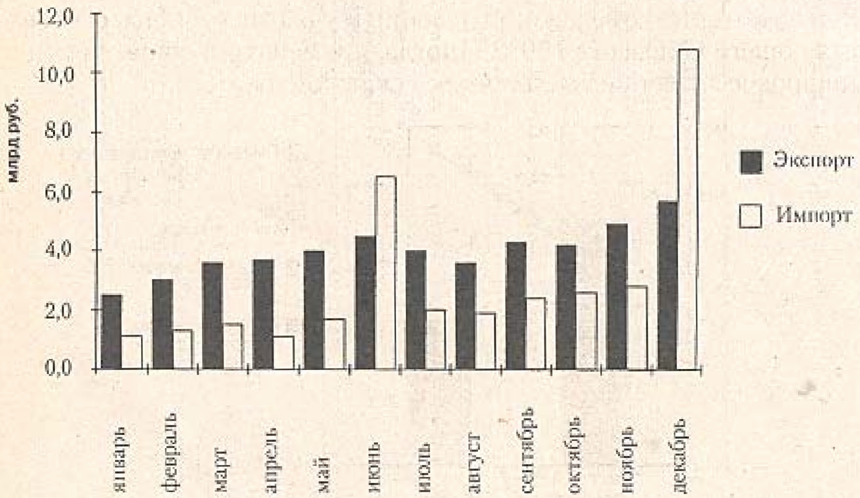

Рис. 3.4. **Динамика экспорта и импорта в Российской Федерации в 1994 г.**

Источник. Бюллетень банковской статистики. М., 1995. № 20. С.4.

Из этой диаграммы видно, что экспорт Российской Федерации в течение 1994 г. возрастал, причем наблюдалась тенденция более интенсивного его роста к концу каждого квартала. Что касается импорта, то он менялся очень неравномерно, и его весьма существенное увеличение отмечено в июне и декабре. Сравнивая динамику экспорта и импорта, можно заметить, что если в целом в течение года объем экспорта преобладал над объемом импорта, то в июне и, особенно, в декабре объем импорта значительно превышал объем экспорта.

Столбиковые диаграммы могут использоваться также для пространственных сопоставлений: сравнения по территориям, странам, фирмам, по различным видам продукции. Кроме того, столбиковые диаграммы широко используются для изучения структуры явлений. Примером использования такой диаграммы является распределение розничного товарооборота по предприятиям различных форм собственности в Российской Федерации в январе – сентябре 1993 г. (см. рис. 3.5).

На этой диаграмме разными штриховками отображается удельный вес различных частей исследуемого показателя, каждой из которых отведен определенный участок столбца, соответствующего по высоте 100%. Иногда, кроме штриховки приводят цифровое значение удельного веса каждой части.

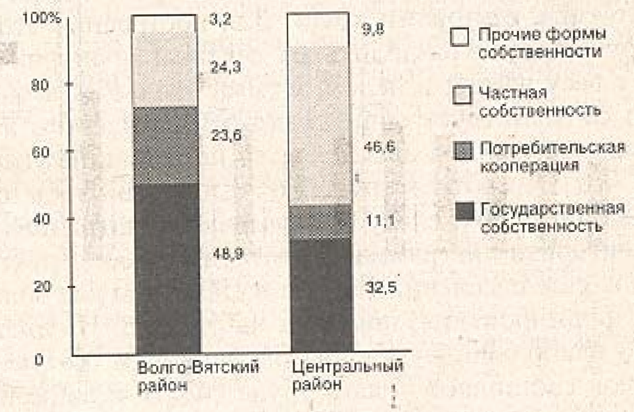

Рис. 3.5. **Распределение розничного товарооборота по предприятиям различных форм собственности в январе–сентябре 1993 г. в % к общему объему товарооборота**

Источник. О развитии экономических реформ в регионах Российской Федерации. М., Госкомстат РФ. 1994. С. 119.

Сравнивая данные по двум экономическим районам, можно отметить существенные различия в структуре розничного товарооборота по отдельным формам собственности в каждом из районов. Так, в Центральном районе преобладает розничный товарооборот предприятий частной формы собственности. На его долю приходится 46,6% всего розничного товарооборота. Также достаточно велика доля розничного товарооборота предприятий государственной формы собственности, составляющая 32,5% всего розничного товарооборота в районе. В то же время в Волго-Вятском районе наибольший удельный вес по розничному товарообороту сосредоточен на предприятиях с государственной формой собственности (48,9%), а доля розничного товарооборота на предприятиях с частной формой собственности составляет 24,3%.Таким образом, приведенная на рис. 3.5 столбиковая диаграмма показывает принципиальное различие в распределении розничного товарооборота по предприятиям различных форм собственности Центрального и Волго-Вятского районов.

Для характеристики структуры социально-экономических явлений достаточно широкое распространение получили **секторные диаграммы**. Анализ структуры проводится на основе сопоставления различных частей целого при помощи площадей, образуемых секторами круга. Для построения этой диаграммы круг следует разделить на секторы пропорционально удельному весу частей в целом. Сумма удельных весов равна 100%, что соответствует общему объему изучаемого явления. Размер каждого сектора определяется по величине угла с учетом того, что 1% соответствует 3,6°. Для того чтобы секторы были более наглядны, следует пользоваться штриховкой. Например, секторная диаграмма, характеризующая структуру денежных доходов населения России в 1994 г. по различным источникам формирования, показана на рис. 3.6. Из диаграммы видно, что наибольший удельный вес в общем объеме денежных доходов составляет оплата труда (48%), а наименьший - пенсии, пособия, стипендии (15 %).

Полосовые диаграммы состоят из прямоугольников, расположенных горизонтально (полосами, лентами). Масштабная шкала этих графиков находится на горизонтальной оси. Принцип построения полосовых диаграмм тот же, что и столбиковых диаграмм.

Рис. 3.6. **Структура денежных доходов населения России в 1994 г. в %**

Источник. Бюллетень банковской статистики. 1995. № 20. С. 5.

Иногда для целей сравнительного анализа по регионам, странам используют **квадратные, круговые, фигурные диаграммы** (диаграммы **фигур-знаков**). Диаграммы геометрических фигур отражают размер изучаемого объекта в соответствии с размером своей площади.

Для построения *квадратной диаграммы,* применяемой при сравнительном анализе, следует извлечь квадратные корни из сравниваемых величин статистических показателей, а затем построить квадраты со сторонами, пропорциональными полученным результатам. Пример квадратной диаграммы (см. рис. 3.7) показывает различие в ходе приватизации предприятий по трем районам России за январь - август 1993г.

Из этой диаграммы следует, что больше всего приватизированных предприятий сосредоточено в Центральном экономическом районе (5406), а наименьшее их число приходится на Волго-Вятский район (1550). Это свидетельствует о том, что число приватизированных предприятий распределяется по районам России крайне неравномерно.

Рис. 3.7. **Приватизация предприятий по районам России за январь - август 1993 г. единиц**

Источник. О развитии экономических реформ в регионах Российской Федерации. С. 98.

При построении *круговой диаграммы* значения показателей вначале делят на число π, т.е. 3,14, а затем из полученных величин извлекают квадратные корни и строят круги с радиусами, пропорциональными полученным результатам.

Диаграммы фигур-знаков представляют собой графические изображения в виде рисунков, силуэтов, фигур, соответствующих содержанию статистических данных. Рисунки отличаются друг от друга размером (соответственно величине показателя), либо величины статистических показателей изображаются на рисунках определенным количеством одинаковых по размеру и типу фигур. Например, грузооборот железнодорожного транспорта символически изображается в виде рисунков вагонов. Для таких диаграмм необходимы сопроводительные числовые надписи, так как зрительное сопоставление таких фигур довольно затруднительно.

Для оценки географического размещения явлений, сравнительного анализа по территориям применяются **статистические карты**. Они достаточно часто используются в публикациях ООН.

Статистические карты включают картограммы и картодиаграммы. *Картограмма* показывает территориальное распределение изучаемого признака по отдельным районам и используется для выявления закономерностей этого распределения. Картограммы бывают фоновые и точечные. *Фоновые картограммы* разной густотой цветовой окраски характеризуют распределение изучаемого признака на различных территориях. Например, данные об урожайности зерновых по нескольким районам наиболее наглядно могут быть представлены в виде фоновой картограммы, когда бледно-желтым цветом будут раскрашены районы с урожайностью до 20 ц с 1 га; ярко-желтым цветом - соответственно с 20 – 30 ц с 1 га; и желто-красным - выше 30 ц с 1 га.

На *точечной картограмме* каждой точке соответствует одно и то же принятое числовое значение, например, равное 100 т. Нанося на контур каждого района соответствующее количество точек, мы получаем точечную картограмму, характеризующую распределение изучаемого признака по районам. Как правило, фоновые картограммы используются при анализе статистических показателей в виде относительных и средних величин, в то время как точечные - для характеристики размещения абсолютных величин.

Картодиаграмма – представляет собой сочетание диаграммы с географической картой. Она позволяет отразить специфику каждого района в распределении изучаемого явления, его структурные особенности.

Контрольные вопросы к главе 3

1. В чем состоит значение метода группировок в анализе статистических данных?
2. Какие основные задачи решаются исследователем с помощью метода группировок?
3. Какова роль и значение классификаций? Приведите примеры важнейших классификаций.
4. Какие группировки называют комбинационными? Приведите пример.
5. В чем состоит отличие комбинационной и многомерной группировки?
6. Какие основные проблемы подлежат решению при группировке статистических данных?
7. Как выполняется группировка, если группировочный признак является дискретным?
8. В каких случаях необходимо определить интервалы группировки по количественным признакам?
9. Каковы функции статистических таблиц?
10. Какие могут быть выделены виды статистических таблиц?

11. С какой целью строятся графики в экономико-статистических исследованиях?

12.В каких целях используются секторные и столбиковые диаграммы? Приведите примеры.

13. Постройте график, характеризующий структуру студентов вашей группы по успеваемости в последнюю сессию.

Глава 4
АБСОЛЮТНЫЕ, ОТНОСИТЕЛЬНЫЕ И СРЕДНИЕ ВЕЛИЧИНЫ

4.1. Общие принципы построения статистических показателей

Специфика статистического изучения действительности состоит в том, что при конструировании показателей, отражающих количественную сторону массовых общественных явлений, статистика опирается на сформулированные экономической наукой понятия, отражающие сущность явлений. Процесс перехода от понятий и категорий экономической науки к системе экономико-статистических показателей представляет собой сложный взаимосвязанный процесс теоретического и эмпирического познания. Понятия экономической науки являются определяющими в этом процессе, ибо лишь после того как выяснена сущность и отличительные особенности той или иной формы, имеет смысл иллюстрировать ее развитие посредством обработанных надлежащим образом статистических данных.

Экономико-статистические показатели содержат количественную характеристику тех или иных свойств экономических явлений. В отличие от понятий их следует конкретизировать применительно к тем реальным явлениям, которые отражаются этим понятием. Таким образом, понятие надо свести к простым и однородным сторонам, свойствам и признакам, которые допускают операции сравнения или измерения.

Показатель формализует содержание изучаемых сторон социально-экономических явлений и представляет собой модель их количественной характеристики. Чем сложнее исследуемое явление, тем труднее оно поддается формализации и моделированию. С помощью показателей определяется, что, где, когда и каким образом следует численно измерить. Каждый *статистический показатель* с возможной точностью должен соответствовать сущности того явления, которое должно быть измерено с его помощью. Например, измерение объема продукции промышленности требует предварительного установления тех видов деятельности промышленного предприятия, которые будут учтены в составе промышленной продукции, и определения тех результатов этой деятельности, которые могут быть включены в ее объем.

Необходимость рассмотрения исследуемого объекта во всех его связях и отношениях приводит к тому, что для получения целостной статистической характеристики изучаемых явлений применяются *системы статистических показателей.* Важнейшая особенность системы показателей - их содержательное единство, связанное с характеристиками единого объекта исследования. Так, система стоимостных показателей продукции промышленного предприятия включает следующие показатели: товарная продукция, отгруженная продукция, реализованная продукция, чистая продукция, стоимость, добавленная обработкой и др. Совершенствование управления экономикой обусловило переход от преимущественной ориентации на показатель валовой продукции к таким показателям, оценивающим вклад непосредственно данного предприятия в производство продукции, как стоимость, добавленная обработкой, или чистая продукция.

С одной стороны, переход к количественному описанию явлений действительности обогащает и уточняет понятие о явлениях. С другой, - при построении показателей почти всегда приходится дополнительно упрощать, схематизировать реальные явления, а потому статистические показатели лишь с известной степенью приближения отражают объективную реальность. Все это делает очень важной постоянную работу по совершенствованию системы показателей. Развитие систем статистических показателей происходит в соответствии с развитием отражаемой объективной действительности и в результате углубления процесса познания реальных систем.

При формировании систем статистических показателей должны быть четко сформулированы значение и область применения соответствующих показателей, определены функции, которые им надлежит выполнять.

Статистические показатели - это величины, адекватно характеризующие отображаемое явление в конкретных условиях времени и места. В этом определении прежде всего охарактеризована учетная функция статистических показателей, реализация которой связана с отражением в них объективных свойств изучаемых явлений. Изучая отчетные данные по действующей системе показателей, руководители разного уровня управления должны получить объективную информацию о реальном состоянии и тенденциях развития организаций, отраслей и экономики в целом. Руководствуясь действующей системой показателей, планирование может сознательно направлять развитие предприятий, фирм, корпораций. *Плановые показатели*, осуществляя директивную функцию, призваны ориентировать руководителей и работников предприятий на выполнение поставленных задач. Составы систем *плановых и учетно-статистических показателей* различаются и довольно значительно, что связано с необходимостью проведения комплексного и всестороннего анализа деятельности предприятий. С помощью анализа выявляются неиспользованные резервы, определяются пути устранения отмеченных недостатков в работе и т.д. При этом нужно иметь в иду, что отчетные показатели должны быть сопоставимы с плановыми по методологии их определения.

Статистические показатели призваны способствовать усилению воздействия плановых показателей на деятельность производственных коллективов, т.е. в определенной степени выполнять также стимулирующую функцию. Таким образом система экономико-статистических показателей в управлении предприятиями призвана выполнять три основные функции: *директивную, учетную и стимулирующую.* В практике хозяйственной деятельности некоторые из применяемых показателей выполняют одновременно все названые функции. Вместе с тем необходимо отметить, что статистические показатели не только играют важную роль в формировании информационного обеспечения управления разных уровней, но и используются в повышении информированности всего населения страны о процессах, происходящих в экономическом и социальном развитии общества. С этих позиций очень важна *познавательная* функция статистических показателей.

4.2. Абсолютные величины

В процессе статистического наблюдения получают данные о значениях тех или иных признаков, характеризующих каждую единицу исследуемой совокупности. Для характеристики совокупности в целом или отдельных ее частей данные по отдельным единицам совокупности подвергают сводке и получают обобщающие показатели, в которых отражаются результаты познания количественной стороны изучаемых явлений.

Обобщающие показатели могут быть представлены *абсолютными, относительными и средними величинами*. Они широко используются в планировании и анализе деятельности предприятий, фирм, концернов, отраслей и экономики в целом. Многообразная характеристика всех сторон исследуемых социально-экономических явлений может быть дана лишь с помощью всех видов обобщающих показателей. Вместе с тем, каждый вид показателей имеет определенное значение и занимает различное место в процессе познания реальной действительности.

Путем непосредственного суммирования первичных данных получают обобщающие **абсолютные показатели**, которые характеризуют численность совокупности и объем (размер) изучаемого явления в конкретных границах времени и места. Например, численность населения России по переписи на 12 января 1989 г. составила 147,4 млн человек, валовый внутренний продукт в 1994 г. - 630 трлн руб. Первый из приведенных показателей характеризует численность населения по состоянию на определенный момент, и получают такие абсолютные величины по результатам сводки данных единовременного наблюдения. Второй из приведенных показателей характеризует размер явления за определенный период и он получен как результат сводки данных текущего наблюдения.

Абсолютные показатели являются всегда именованными числами, т.е. имеют какую-либо единицу измерения. *Натуральные единицы измерения* применяют в тех случаях, когда единица измерения соответствует потребительским свойствам продукта. Например, производство цемента оценивается в тоннах, тканей - в квадратных метрах, автомобилей - в штуках и т.д.

Натуральные единицы могут быть и составными (сложными). Например, отработанное рабочими время учитывается в человеко-днях и человеко-часах, грузооборот автомобильного и железнодорожного транспорта - в тонно-километрах и т.п. Составные единицы отражают сочетание двух различных сторон явления. Так, при учете затрат рабочего времени отражается совместное измерение численности рабочих и времени их работы, при определении объема работы транспорта - измерение объема грузов и расстояния перевозок.

При учете продукции в натуральном выражении нередки случаи, когда применяются различные единицы измерения для одного и того же вида продукции. Это делается для того, чтобы полнее охарактеризовать потребительское назначение продукции и

изменение ее состава. Например, производство электродвигателей учитывается в штуках и киловаттах мощности, бумаги - в тоннах и квадратных метрах, стальных труб - в тоннах и погонных метрах и т.п.

Если некоторые разновидности продукции обладают общностью основного потребительного свойства, обобщенные итоги по выпуску этих разновидностей продукции можно получить, используя *условно-натуральные единицы*. В этом случае одна из разновидностей принимается в качестве единого измерителя, а другие приводятся к этому измерителю с помощью соответствующих коэффициентов пересчета. Например, в тоннах условного топлива определяется общий объем потребленного на предприятии топлива, при определении объема производства минеральных удобрений пересчет производится на стандартное или 100%-ное содержание питательного вещества и т.п.

При обобщении учетных данных даже на уровне предприятий, а тем более на уровне отраслей и народного хозяйства широко используются *стоимостные (денежные) единицы измерения*. Для получения общего объема продукции в денежном выражении количество единиц каждого вида продукции в натуральном выражении умножается на цену соответствующего вида, а затем полученные произведения суммируют во всем видам. При определении стоимостных показателей объема продукции абсолютные величины получаются расчетным путем. Тем более это касается определения таких обобщающих показателей, как чистая продукция промышленности, прибыль, объем валового национального продукта и др.

Так, объем чистой продукции определяют вычитанием из стоимости объема продукции стоимости материальных затрат на ее производство. Прибыль от реализации продукции промышленного предприятия получают как разность между выручкой от реализации продукции по оптовым ценам предприятий и ее полной себестоимостью. Расчетный метод определения абсолютных величин может опираться также на балансовый метод. Например, наличие оборудования на конец года определяется прибавлением к количеству оборудования на начало года числа единиц введенного оборудования и вычитанием числа единиц выбывшего оборудования. Таким образом, абсолютные величины получают непосредственным подсчетом данных статистического наблюдения или расчетным путем.

Абсолютные статистические показатели могут быть измерены с различной степенью точности. С переходом к более высоким ступеням обобщения применяются и более укрупненные единицы измерения. Например, производство металлорежущих станков на промышленном предприятии учитывается в штуках, а в масштабах страны - в миллионах штук; государственные закупки зерна в хозяйствах измеряются в тоннах, а по России в целом - в тысячах тонн.

Соблюдение одноименных единиц измерения исследуемых показателей является важным исходным условием при международных сравнениях. Вместе с тем одноименные единицы измерения объема производства однотипной продукции в разных странах могут быть неодинаковыми по величине или методологии учета. Это касается, например, показателей производства электроэнергии, выпуска строительного кирпича и ряда других продуктов.

4.3. Относительные величины

Хотя абсолютные величины играют важную роль в практической и познавательной деятельности человека, анализ фактов обязательно приводит к необходимости различного рода сопоставлений. И тогда абсолютные показатели, характеризующие те или иные изучаемые явления, рассматриваются не только самостоятельно, но и в сравнении с другим показателем, который принимается за масштаб оценки или иначе за базу сравнения.

Сопоставление статистических данных осуществляется в различных формах и по разным направлениям. В соответствии с различными задачами и направлениями сопоставления статистических данных применяются различные виды **относительных величин**, классификация которых представлена на рис. 4.1.

Как видно из приведенной классификации, сопоставлять можно одноименные показатели, относящиеся к различным периодам, различным объектам или разным территориям. Результат такого сопоставления может быть представлен коэффициентом (база сравнения принята за единицу) или выражен в процентах и показывает, во сколько раз или на сколько процентов сравниваемый показатель больше или меньше базисного.

В результате соотношения одноименных показателей получают следующие относительные величины: относительные величины динамики; относительные величины выполнения плана и

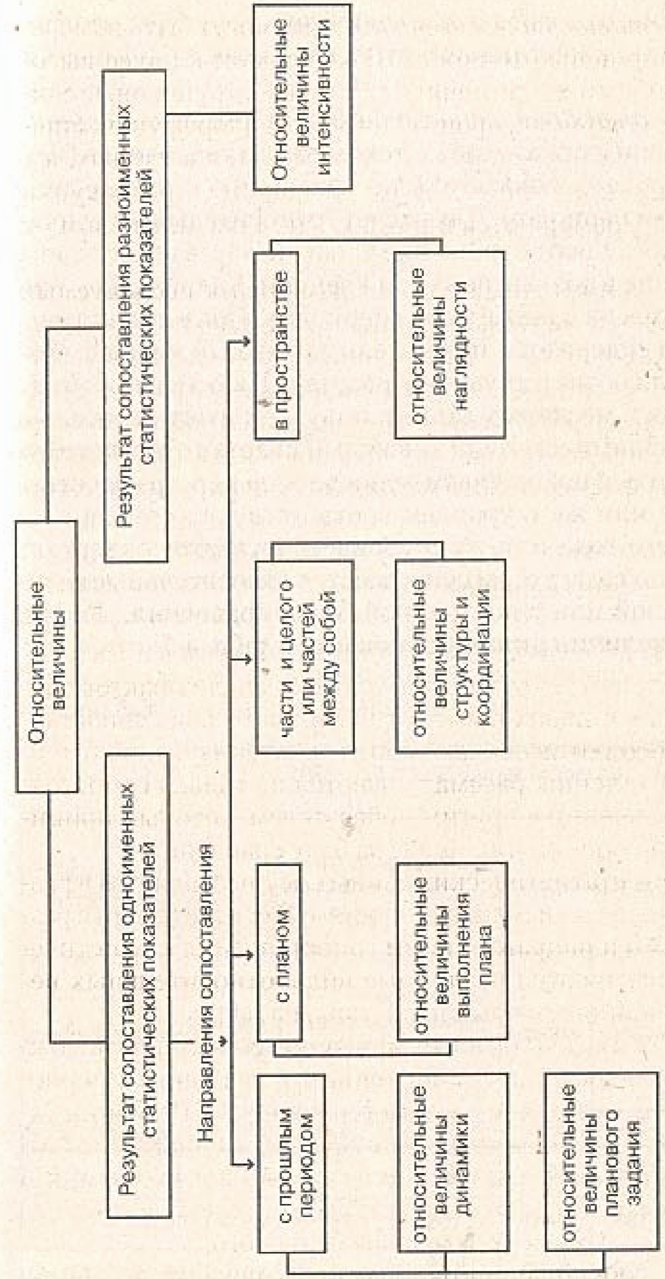

Рис. 4.1. Классификация относительных величин

планового задания; относительные величины структуры; относительные величины координации; относительные величины наглядности.

Относительная величина динамики характеризует изменение явления во времени и показывает, во сколько раз увеличился (или уменьшился) уровень показателя по сравнению с каким-либо предшествующим периодом. Для расчета относительной величины динамики определяют отношение уровней, характеризующих изучаемое явление в разные периоды времени. Как правило, при анализе опираются на данные за ряд периодов. Так, в статистической отчетности содержатся показатели за каждый месяц отчетного года или за соответствующий квартал. На основе обобщения квартальных и месячных данных получаем итоги за каждый год. Статистические показатели за каждый квартал отчетного года можно сравнить с показателем непосредственно предшествующего квартала или же с уровнем соответствующего квартала предшествующего года или же с уровнем четвертого квартала предшествующего года, т.е. рассчитываются относительные величины с переменной или с постоянной базой сравнения. Расчет относительных величин динамики показан в табл. 4.1.

Таблица 4.1

Внешнеторговый оборот России со странами дальнего зарубежья в I кв. 1995 г.

Месяц и год	Внешнеторговый оборот России, млн. долл. США (со странами дальнего зарубежья)	В том числе		Коэффициент роста внешнеторгового оборота, %			Удельный вес, %	
		экспорт	импорт	к соотв. периоду прошлого года	к пред. месяцу	к дек. 1994 г.	экспорта	импорта
1	2	3	4	5	6	7	8	9
Дек. 1994 г.	8640	4730	3910	92	-	100	54,75	45,25
Янв. 1995 г.	5629	3188	2441	106	65	65	56,64	43,36
Фев. 1995 г.	6902	3800	3102	117	123	80	55,06	44,94
Март 1995 г.	7741	4536	3205	121	112	90	58,60	41,40

Относительные величины имеют форму коэффициентов, если они исчисляются делением сравниваемой величины на базу сравнения. Если коэффициент умножить на 100, то получим результат сопоставления в процентах. Таковы относительные величины, характеризующие динамику внешнеторгового оборота (см. гр. 5, 6, 7, табл. 4.1). Рассчитанная относительная величина динамики показывает, например, что в январе 1995 г. по сравнению с декабрем 1994 г. оборот внешней торговли России снизился на 35%. Эта величина исчисляется как разность между величиной темпа роста (коэффициент роста, выраженный в процентах) и 100% и называется относительным приростом (в приводимом примере – относительное снижение). В таблице 4.1 расчет коэффициентов роста с постоянной базой ведется в сравнении с декабрем 1994 г. Выбор базы сравнения нередко имеет существенное значение. Так, в ряде случаев в качестве базы сравнения принимаются годы, являющиеся исторически обусловленной границей отдельных периодов времени.

В планах предприятий задания устанавливаются как в абсолютных показателях, так и в виде относительных величин, которые показывают, во сколько раз или на сколько процентов должна увеличиться (уменьшиться) величина показателя по плану в сравнении с его уровнем в предшествующем периоде. Например, по плану темп роста экспорта продукции предприятия должен был составить в 1995 г. (в % к предшествующему году) 106,1%, относительная величина динамики объема экспорта продукции составила 107,3%. Сравнение двух относительных величин - планового задания и динамики – свидетельствует о перевыполнении плана по экспорту продукции предприятия в 1995 г. Степень выполнения плана оценивается с помощью относительной величины выполнения плана, которую получают отношением фактического уровня показателя в отчетном периоде к его уровню, запланированному на этот же период.

Рассмотрим, как связаны между собой относительные величины планового задания, выполнения плана и динамики.

Обозначим y^0 - фактический уровень показателя в базисном периоде; $y^{пл}$ - планируемый уровень показателя на отчетный период; y^1 - фактический уровень показателя в отчетном периоде. Тогда в принятых обозначениях относительные величины могут быть представлены следующими соотношениями:

Относительная величина планового задания – $\dfrac{y^{пл}}{y^0}$.

Относительная величина выполнения плана – $\frac{y^1}{y^{пл}}$.

Относительная величина динамики – $\frac{y^1}{y^0}$.

Следовательно, относительная величина динамики может быть получена произведением относительных величин планового задания и выполнения плана, т.е.

$$\frac{y^1}{y^0} = \frac{y^{пл}}{y^0} \cdot \frac{y^1}{y^{пл}}. \tag{4.1}$$

Таким образом, в приводимом примере на основе относительных величин динамики и планового задания может быть определена относительная величина выполнения плана по росту экспорта продукции предприятия:

$$\frac{y^1}{y^{пл}} = \frac{y^1}{y^0} : \frac{y^{пл}}{y^0}, \text{ т.е. } \frac{y^1}{y^{пл}} = 1{,}073 : 1{,}061 = 1{,}0113,$$

или относительная величина выполнения плана составила 101,13%, а это означает, что план по росту экспорта продукции предприятия был перевыполнен в 1995 г. на 1,13 %.

В ряде случаев расчет относительной величины выполнения плана может выполняться по методу нарастающего итога. Так, оценка выполнения квартального плана по объему продукции производится по данным, взятым нарастающим итогом с начала квартала, оценка выполнения годовых планов – нарастающим итогом с начала отчетного года. В табл. 4.2 приведен расчет процента выполнения плана по объему продукции нарастающим итогом:

$$\text{процент выполнения плана по объему продукции} = \frac{\sum_{i=1}^{n} П_i^1}{\sum_{i=1}^{n} П_i^{пл}} \cdot 100\%, \tag{4.2}$$

где $П_i^1$ и $П_i^{пл}$ – соответственно фактический и запланированный объем продукции по месяцам года;

n - число месяцев, за которые ведется расчет процента выполнения плана.

Таблица 4.2

Объем продукции предприятия за I-е полугодие 1995 г. (млн руб.)

Месяц	Объем продукции, млн руб.		Объем продукции нарастающим итогом, млн руб.		% выполнения плана по продукции (гр. 5 : : гр.4)
	план	факт	план	факт	
1	2	3	4	5	6
Январь	1800	1770	1800	1770	98, 33
Февраль	1850	1865	3650	3635	99,59
Март	1920	2010	5570	5645	101, 35
Апрель	1910	1920	7480	7565	101,14
Май	1870	1890	9350	9455	101,12
Июнь	2100	2070	11 450	11 525	100,66

По данным примера, невыполнение плана (см. гр. 6 табл. 4.2) имело место в январе и феврале, а за I-й квартал и I-е полугодие план был перевыполнен соответственно на 1,35 % и 0, 66%.

Относительные величины структуры характеризуют долю отдельных частей в общем объеме совокупности, их рассчитывают как отношение числа единиц (или объема признака) в отдельных частях совокупности к общей численности единиц (или объему признака) по всей совокупности. Относительные величины структуры рассчитываются по сгруппированным данным. Например, в составе промышленно-производственного персонала выделяются четыре категории: рабочие, руководители, специалисты и служащие. Анализ показателей доли (удельного веса) каждой категории в составе промышленно-производственного персонала позволяет сопоставлять состав работников по категориям на различных предприятиях отрасли, в различных отраслях и т.д.

Расчет относительных величин структуры за несколько периодов позволяет выявить структурные сдвиги. В гр. 8 и 9 табл. 4.1 приведен расчет относительных величин структуры. В них показана доля экспорта и импорта в объеме внешнеторгового оборота за период с декабря 1994 г. по март 1995 г.

В целях анализа наряду с соотношением части и целого (показателями структуры) определяют соотношение между двумя частями одного целого. Относительные величины, характеризующие соотношение между частями одного целого, называют **относительными величинами координации**. К таким величинам относятся, например, показатели, характеризующие соотношение между численностью городского и сельского населения, между численностью рабочих и служащих, между численностью мужчин и женщин, между величиной заемного и собственного капитала банка и т.д. Относительные величины координации нередко характеризуются числом единиц одной части на 100 или 1000 единиц другой части. Например, в табл. 4.3 показано, сколько безработных приходится на 1000 человек занятых в экономике России. Данные этой таблицы свидетельствуют об устойчивой тенденции увеличения соотношения численности безработных и занятых в экономике России за шесть кварталов 1994 и 1995 гг.

Таблица 4.3

Соотношение численности безработных и занятых в экономике России

	1994 г.				1995 г.	
	I кв.	II кв.	III кв.	IV кв.	I кв.	II кв.
Численность безработных на 1000 занятых в экономике	70	72	75	77	82	83

Относительные величины наглядности отражают результаты сопоставления одноименных показателей, относящихся к одному и тому же периоду (или моменту) времени, но к разным объектам или территориям. Этот вид относительных величин применяется для сравнительной оценки уровня развития стран и регионов, а также при оценке результатов деятельности отдельных предприятий отрасли. Обычно их исчисляют в процентах или кратных отношениях, показывающих во сколько раз одна из сравниваемых величин больше (или меньше) другой.

Таблица 4.4

Соотношение производства некоторых основных видов промышленной продукции в России и США (к уровню производства США)

Год	Вид продукции, % к США			
	электроэнергия	нефть (включая газовый конденсат)	цемент	сталь
1990	33,81	139,08	116,08	99,89
1991	32,87	123,53	119,05	96,74
1992	30,84	109,92	87,15	79,48
1993	28,31	97,46	67,62	67,01

Данные этой таблицы свидетельствуют об увеличении разрыва в производстве всех приведенных в таблице видов промышленной продукции, а по таким видам продукции, как нефть и цемент, по которым Россия опережала уровень американского производства, наметилось отставание (с 1992 г. по цементу, а с 1993 г. по нефти) и довольно значительное, например по производству цемента.

Относительные величины наглядности находят широкое применение не только при международных сопоставлениях, но и при сравнительной оценке показателей деятельности предприятий и различных регионов. Так, могут сравниваться результаты работы предприятий различных форм собственности. Безусловный интерес представляет сопоставление цен на продукцию государственных и частных предприятий, средней заработной платы работников этих предприятий и т.д., принимая при этом цену и другие показатели государственных предприятий за базу сравнения.

Отношения между разноименными абсолютными величинами называют **относительными величинами интенсивности**. В их числе можно назвать показатели жизненного уровня населения, к которым относятся показатели потребления продуктов питания и непродовольственных товаров на душу населения, показатели обеспеченности населения предметами культурно-бытового и хозяйственного назначения длительного пользования в расчете на 100 семей или на 1000 человек населения; обеспеченности населения жильем и т.д. Учитывая экономическую сущность относительных величин интенсивности, их можно было бы назвать показателями уровня экономического и социального развития.

Так, в практике статистической работы находят широкое применение показатели технической оснащенности труда: фондо-, машино- и энерговооруженность труда. Показатель энерговооруженности труда рабочих рассчитывается делением суммарной энергетической мощности, используемой в производственном процессе, на число рабочих в наиболее заполненной смене. Коэффициент фактической электровооруженности труда исчисляется делением потребленной за определенный период электроэнергии на число отработанных за этот период человеко-часов. Таким образом, с помощью относительных величин интенсивности оценивается уровень технического развития производства.

В отличие от относительных величин, являющихся результатом сопоставления одноименных показателей и представляемых с помощью коэффициентов или процентов, относительные величины интенсивности являются именованными числами (число аварий на 10 тыс. человек населения, производство хлопчатобумажных тканей в квадратных метрах на душу населения и т.д.).

Итак, относительные величины - один из важнейших способов обобщения и анализа статистической информации. Цели и направления исследования определяют выбор вида относительных величин. Вместе с тем для полной характеристики различных сторон изучаемых явлений необходима система относительных величин, рассчитанных по ряду существенных признаков. Например, состав рабочих промышленного предприятия изучается по профессиям, полу, возрасту, уровню образования, стажу работы, квалификации, уровню механизации их труда, т.е. относительные величины структуры должны быть рассчитаны по группам, сформированным на основе вышеуказанных признаков. С другой стороны, важно выявить основные тенденции в изменении состава рабочих. Это требует сопоставления данных за ряд лет с помощью относительных величин динамики. Определенный интерес представляет сопоставление численности основных и вспомогательных рабочих и других относительных величин, позволяющих всесторонне характеризовать состав и движение рабочих.

Важно отметить, что в процессе экономико-статистического анализа абсолютные и относительные величины должны рассматриваться во взаимосвязи, т.е. пользоваться относительными величинами нужно не формально, а представлять, какая абсолютная величина скрывается за каждым относительным показателем. Особенно важно соблюдать это положение при расчете относительных величин динамики.

Одно из главных требований, которое предъявляется при исчислении относительных величин, заключается в необходимости обеспечения сопоставимости сравниваемой величины и величины, принятой за базу сравнения. Прежде всего должна быть обеспечена сопоставимость по методологии расчета сравниваемых показателей, по степени охвата объектов исследуемой совокупности и другим существенным обстоятельствам (подробнее см. гл. 8).

4.4. Виды средних величин и их значение в социально-экономических исследованиях

В процессе обработки и обобщения статистических данных возникает необходимость определения средних величин. Как правило, индивидуальные значения одного и того же признака у различных единиц совокупности неодинаковы. *Средняя величина - обобщающая характеристика изучаемого признака в исследуемой совокупности. Она отражает его типичный уровень в расчете на единицу совокупности в конкретных условиях места и времени.*

Например, при изучении доходов рабочих концерна обобщающей характеристикой служит средний доход одного рабочего. Для его определения общую сумму средств, направленных на потребление, в виде заработной платы, социальных и трудовых льгот, материальной помощи, дивидендов по акциям и процентов по вкладам в имущество концерна за рассматриваемый период (год, квартал, месяц) делят на численность рабочих концерна.

Естественно, индивидуальные значения дохода отдельных рабочих отличаются от среднего уровня в силу ряда причин (квалификации, стажа работы, наличия акций, суммы вклада в имущество концерна и др.). Средний доход в свою очередь характеризует то общее, что свойственно всей совокупности рабочих предприятия, т.е. уровень дохода массы рабочих в конкретных условиях функционирования данного концерна в рассматриваемом периоде.

Важный вклад в обоснование и развитие теории средних величин внес крупнейший ученый XIX в. Адольф Кетле (1796-1874), член Бельгийской академии наук, член-корреспондент Петербургской академии наук.

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 совокупности, а следовательно, их отклонения от типичного уровня. Эти причины не связаны с природой изучаемого явления, их называют случайными причинами.

При исчислении средней величины по массе 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.

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.

Основным условием научного использования средней величины является качественная однородность совокупности, по которой исчислена средняя.

Подумайте, можно ли назвать величину 500 долл. средней заработной платой трех лиц, индивидуальные заработки которых соответственно равны 1200, 200 и 100 долл. Ясно, что по уровню своей заработной платы обследованные лица относятся к разным категориям работников, и некорректным будет использование данной величины для характеристики средней заработной платы обследованных лиц.

Или другой пример некорректного использования средней величины. Акционерный капитал компании с ограниченной ответственностью равен 1000 млн руб., количество акционеров компании 100 человек. Средний показатель участия в акционерном капитале - средняя величина пакета акций - равняется 10 млн руб. Эта средняя величина - 10 млн руб. показывает, что капитал компании находится преимущественно в руках

мелких держателей акций. В действительности положение может быть следующим: 1 акционер имеет 1010 акций на сумму 505 млн руб.; а 99 акционеров имеют по 10 акций на общую сумму 495 млн руб.

Как видим, существует две категории акционеров, к первой из них относится один акционер с величиной пакета акций, равной 505 млн руб.; ко второй - 99 акционеров со средней величиной пакета акций, равной 5 млн руб.

Таким образом, один из акционеров владеет более чем 50% капитала и осуществляет контроль над всей компанией. Полученная же средняя, равная 10 млн руб., не может считаться надежной оценкой свойств данной совокупности, так как она в два раза больше по своей величине, чем индивидуальные пакеты акций 99% акционеров компании. 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. 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 бывшего СССР проводилось в региональном аспекте (по союзным республикам). Традиционно более высокая рождаемость была в Средней Азии и Закавказье по сравнению с Центральными районами России. Среднее количество детей в семье, исчисленное по каждому региону - это групповые средние, а соответственно исчисленное по всей территории СССР - 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).

Существуют две категории средних величин: степенные средние (к ним относятся средняя арифметическая, средняя гармоническая, средняя геометрическая и др.), а также структурные средние (мода и медиана). Выбор того или иного вида средней производится в зависимости от цели исследования, экономической сущности усредняемого показателя и характера имеющихся исходных данных.

Рассмотрим пример. Известны значения месячной заработной платы рабочих бригады за октябрь 1995 г.

Табельный номер рабочего	15	16	27	30	20	41	25	32	18	49	Всего
Месячная заработная плата рабочего, тыс. руб.	493	561	609	718	850	894	901	1070	1203	1251	8550

Требуется определить среднюю месячную заработную плату рабочих бригады ($\overline{x}$).

Общая сумма заработной платы всех рабочих $\sum_{i=1}^{n} x_i = 8550$ тыс. руб., это *определяющий показатель* исчисленный как сумма индивидуальных значений заработной платы x_i каждого рабочего, другими словами - это фонд оплаты их труда, который может быть записан алгебраически: $x_1+x_2+x_3+\ldots+x_n=\sum_{i=1}^{n} x_i$.

Определяющий показатель, выраженный математически, называется *определяющей функцией*.

Определяющей функции соответствует уравнение средних, где индивидуальная заработная плата каждого рабочего замене-

на средней заработной платой, поскольку такая замена не сказывается на общей сумме оплаты труда всех рабочих бригады - определяющего показателя: $\overline{X}+\overline{X}+\overline{X}+...+\overline{X}=\sum_{i=1}^{n} X_i$. Зная определяющую функцию и уравнение средних $\overline{X}+\overline{X}+\overline{X}+...+\overline{X}=X_1+X_2+X_3+...+X_n$, или $n\overline{X}=\sum_{i=1}^{n} X_i$ получаем формулу **простой средней арифметической**

$$\overline{X}=\frac{\sum_{i=1}^{n} X_i}{n}, \qquad (4.3)$$

где X_i - индивидуальное значение признака каждой единицы совокупности;

n - число единиц совокупности.

Таким образом, средняя месячная заработная плата одного рабочего бригады вычисляется по формуле:

$\overline{X}_{\text{арифм.}}=\frac{\sum_{i=1}^{n} X_i}{n}=\frac{8550}{10}=855$ тыс. руб. Если бы все единицы изучаемой совокупности развивались под действием одних общих условий и на них не действовали никакие «случайные» факторы, то величина признака у каждой единицы - индивидуальное значение месячной заработной платы - была бы одинаковой, равной 855 тыс. руб. и обеспечивала величину итогового показателя: 855 тыс.руб. x 10 чел. = 8550 тыс. руб.

Итак, при выборе вида средней величины обычно исходят из логической сущности осредняемого признака и его взаимосвязи с итоговым (определяющим) показателем. Величина итогового показателя не должна изменяться при замене индивидуальных значений признака средней величиной.

Способность средних величин сохранять свойства статистических совокупностей называют *определяющим свойством*.

Общая формула степенной средней записывается следующим образом:

$$\overline{X}=\left(\frac{\sum_{i=1}^{n} X_i^k}{n}\right)^{\frac{1}{k}}=\sqrt[k]{\frac{\sum_{i=1}^{n} X_i^k}{n}}=\sqrt[k]{\frac{X_1^k+X_2^k+\ldots+X_n^k}{n}} \qquad (4.4)$$

С изменением показателя степени k выражение данной функции меняется, и в каждом отдельном случае приходим к определенному виду средней.

Запишем формулы различных видов степенных средних, придавая k значения: -1, 0, 1, 2.

При $k=-1$ получим **среднюю гармоническую величину:**

$$\overline{X}_{\text{гарм}}=\sqrt[-1]{\frac{\sum_{i=1}^{n} X_i^{-1}}{n}}=\frac{n}{\sum_{i=1}^{n}\frac{1}{X_i}}. \qquad (4.5)$$

При $k=0$ получим **среднюю геометрическую величину:**

$$\overline{X}_{\text{геом.}}=\sqrt[0]{\frac{\sum_{i=1}^{n} X_i^0}{n}}=\left(\frac{\sum(1_1+1_2+\ldots+1_n)}{n}\right)^{\frac{1}{0}}=\left(\frac{n}{n}\right)^{\infty}=1^{\infty} \qquad (4.6)$$

Для раскрытия неопределенности прологарифмируем обе части степенной средней:

$$\ln\overline{X}=\frac{1}{k}\left(\ln\sum_{i=1}^{n} x_i^k-\ln n\right)$$

и, подставив $k=0$, получим

$$\ln\overline{X}=\frac{1}{0}(\ln\sum 1-\ln n)=\frac{1}{0}\left(\ln(1_1+1_2+\ldots+1_n)-\ln n\right)=\frac{0}{0}, \qquad (4.7)$$

т.е. неопределенность типа 0/0.

Для ее раскрытия используем правило Лопиталя и найдем $\lim\ln X$ как предел отношения производных по k числителя и знаменателя в правой части равенства.

При $k\longrightarrow 0$

$$\lim\ln\overline{X}=\lim_{k\to 0}\frac{\ln\sum X^k-\ln n}{k}=\lim_{k\to 0}\frac{\frac{1}{\sum X^k}\sum X^k\ln X}{1}=\frac{\sum\ln X}{n}. \qquad (4.7а)$$

Таким образом, при $k=0$, $\ln\overline{X}_{\text{геом.}}=\frac{\sum\ln X}{n}$, после потенцирования $\overline{X}_{\text{геом.}}=\sqrt[n]{X_1\cdot X_2\cdot\ldots\cdot X_n}=\sqrt[n]{\Pi X_i}$.

При $k=1$ получим **среднюю арифметическую:**

$$\overline{X}_{\text{арифм.}}=\sqrt[1]{\frac{\sum_{i=1}^{n} X_i}{n}}=\frac{\sum_{i=1}^{n} X_i}{n} \qquad (4.7б)$$

При $k=2$ - **среднюю квадратическую:**

$$\overline{X}_{\text{кв.}}=\sqrt{\frac{\sum_{i=1}^{n} X_i^2}{n}} \qquad (4.7в)$$

и т.д. для любой степени.

Приведенные выше формулы *простых* средних применяются в случае, если индивидуальные значения усредняемого признака не повторяются.

Однако, когда в практических исследованиях отдельные значения изучаемого признака встречаются несколько раз у единиц исследуемой совокупности, тогда частота повторения индивидуальных значений признака (вес) присутствует в расчетных формулах степенных средних. В этом случае они называются формулами *взвешенных* средних и имеют следующий вид:

средняя гармоническая $\overline{X}_{\text{гарм.}}=\dfrac{\sum_{i=1}^{n} f_i}{\sum_{i=1}^{n}\frac{1}{X_i}f_i}$, (4.8)

средняя геометрическая $\overline{X}_{\text{геом.}}=\sqrt[\sum f]{X_1^{f_1}\cdot X_2^{f_2}\cdot\ldots}=\sqrt[\sum f_i]{\Pi X_i^{f_i}}$, (4.9)

средняя арифметическая $\overline{X}_{\text{арифм.}}=\dfrac{\sum_{i=1}^{n} X_i f_i}{\sum_{i=1}^{n} f_i}$, (4.10) средняя квадратическая $$\bar{x}_{кв.} = \sqrt{\frac{\sum_{i=1}^{n} x_i^2 f_i}{\sum_{i=1}^{n} f_i}}, \qquad (4.11)$$

где f_i - частота повторения индивидуального значения признака (его вес).

Весом может быть и *частость*, т.е. отношение частоты повторения индивидуального значения признака к сумме частот:

$$d_i = \frac{f_i}{\sum f_i}.$$

Для наглядности представим в табл. 4.5 формулы вычисления различных видов степенных средних величин, наиболее часто применяемых в практических исследованиях.

Таблица 4.5

Формулы расчета различных видов степенных средних величин

Значение k	Наименование средней	Формулы средней: простая	Формулы средней: взвешенная
-1	Средняя гармоническая	$\bar{x} = \frac{n}{\sum_{i=1}^{n} \frac{1}{x_i}}$	$\bar{x} = \frac{\sum_{i=1}^{n} f_i}{\sum_{i=1}^{n} \frac{1}{x_i} f_i}$
0	Средняя геометрическая	$\bar{x} = \sqrt[n]{\Pi x_i}$	$\bar{x} = \sqrt[\sum f_i]{\Pi x_i^{f_i}}$
1	Средняя арифметическая	$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$	$\bar{x} = \frac{\sum_{i=1}^{n} x_i f_i}{\sum_{i=1}^{n} f_i}$
2	Средняя квадратическая	$\bar{x} = \sqrt{\frac{\sum_{i=1}^{n} x_i^2}{n}}$	$\bar{x} = \sqrt{\frac{\sum_{i=1}^{n} x_i^2 f_i}{\sum_{i=1}^{n} f_i}}$

Известно, что степенные средние разных видов, исчисленные по одной и той же совокупности, имеют различные количественные значения. И чем больше показатель степени k, тем больше и величина соответствующей средней:

$$\bar{x}_{гарм.} < \bar{x}_{геом.} < \bar{x}_{арифм.} < \bar{x}_{кв.}.$$

Это свойство степенных средних возрастать с повышением показателя степени определяющей функции называется **мажорантностью средних**.

К средним величинам, кроме степенных средних, относят также **моду** и **медиану**.

Для вычисления степенных средних необходимо использовать все имеющиеся значения признака. Мода и медиана определяются лишь структурой распределения. Поэтому их именуют структурными позиционными средними. Медиану и моду часто используют как среднюю характеристику в тех совокупностях, где расчет средней степенной невозможен или нецелесообразен.

Например, выборочное обследование в одном из районов Москвы 12 коммерческих пунктов обмена валюты позволило зафиксировать различные цены за доллар при его продаже (данные на 10 октября 1995 г. при биржевом курсе долл. - 4493 руб.).

Таблица 4.6

№ пункта обмена валюты	1	2	3	4	5	6	7	8	9	10	11	12
Цена за 1 долл., руб.	4500	4560	4540	4535	4550	4500	4560	4570	4560	4560	4570	4500

В силу того, что исследователь не располагает данными об объеме продаж в каждом обменном пункте, расчет средней арифметической с целью определения средней цены за доллар нецелесообразен. Однако можно определить то значение признака, которое делит единицы ранжированного ряда на две части. Такое значение носит название **медианы**. *Медиана лежит в середине ранжированного ряда и делит его пополам.*

Расчет медианы по несгруппированным данным производится следующим образом:

а) расположим индивидуальные значения признака в возрастающем порядке:

X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}	X_{12}
4500	4500	4500	4535	4540	4550	4560	4560	4560	4560	4570	4570

б) определим порядковый номер медианы по формуле:

$$№ Me = \frac{n+1}{2},$$

в нашем случае $№ Me = \frac{12+1}{2} = 6{,}5$.

Это означает, что медиана в данном случае расположена между шестым и седьмым значениями признака в ранжированном ряду, так как ряд имеет четное число индивидуальных значений. Таким образом, Me равна средней арифметической из соседних значений: 4550, 4560.

$$Me = \frac{4550+4560}{2} = 4555 \text{ руб.}$$

в) рассмотрим порядок вычисления медианы в случае нечетного числа индивидуальных значений.

Допустим, мы наблюдали не 12, а 11 пунктов обмена валюты, тогда ранжированный ряд будет выглядеть следующим образом (отбрасываем 12-й пункт):

X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}
4500	45 00	4535	4540	4550	4560	4560	4560	4560	4570	4570

Находим номер медианы: $№ Me = \frac{11+1}{2} = 6$;

на шестом месте стоит $X_6 = 4560$, который и является медианой: $Me = 4560$ руб.

Мода – это наиболее часто встречающееся значение признака у единиц данной совокупности. Она соответствует определенному значению признака.

В нашем случае модальной ценой за доллар можно назвать 4560 руб.: это значение повторяется 4 раза, чаще, чем все другие. На практике моду находят, как правило, по сгруппированным данным. Определить величину моды в первичном ряду в точном соответствии с данным правилом возможно только при достаточно большом количестве наблюдений и при условии, что одно из индивидуальных значений изучаемого признака у отдельных единиц совокупности повторяется значительно чаще, чем все другие значения.

Методология расчета моды и медианы по сгруппированным данным будет рассмотрена в гл. 5.

4.5. Средняя арифметическая, ее свойства и другие степенные средние

В статистической практике из всех перечисленных выше видов средних чаще всего используется средняя арифметическая. Ее расчет осуществляется по-разному для несгруппированных и сгруппированных данных. Рассмотрим пример.

Требуется вычислить средний стаж работы 12 работников рекламного агентства. При этом известны индивидуальные значения признака (стажа) в годах: 6, 4, 5, 3, 3, 5, 5, 6, 3, 7, 4, 5.

$$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n} = \frac{6+4+5+3+3+5+5+6+3+7+4+5}{12} = \frac{56}{12} = 4{,}7 \text{ года}$$

Как видим, средняя арифметическая может оказаться дробным числом, если даже индивидуальные значения признака заданы только целыми числами. Это вытекает из сущности средней арифметической, которая есть величина абстрактная (теоретическая), т.е. она может принимать такое числовое значение, которое не встречается в представленной совокупности индивидуальных значений признака.

Под средней арифметической понимается такое значение признака, которое имела бы каждая единица совокупности, если бы общий итог всех значений признака был распределен равномерно между всеми единицами совокупности.

Заметим, что в нашем примере одно и то же значение признака встречается несколько раз. Объединив данные по величине признака и подсчитав число случаев повторения каждого из них, проведем расчет среднего стажа по сгруппированным данным с помощью формулы средней взвешенной арифметической.

Стаж работы, годы	3	4	5	6	7	Итого
Количество работников, человек	3	2	4	2	1	12

$$\bar{X} = \frac{\sum_{i=1}^{n} X_i f_i}{\sum_{i=1}^{n} f_i} = \frac{3\cdot3+4\cdot2+5\cdot4+6\cdot2+7\cdot1}{12} = \frac{56}{12} = 4{,}7 \text{ года}.$$

Легко заметить, что средняя арифметическая взвешенная, по которой производился расчет в рассмотренном примере, не имеет принципиальных отличий от простой средней арифметической (средние, рассчитанные по разным формулам совпадают), просто суммирование f раз одного и того же значения признака (варианта) заменено в ней умножением варианта на f.

Однако естественно, что при этом величина средней зависит уже не только от величины индивидуальных значений признака (как в простой средней арифметической), но и от соотношения их весов (частот). Чем большие веса имеют малые значения вариантов, тем меньше величина средней и наоборот.

При расчете средних по сгруппированным данным следует учитывать, что важное значение имеет обоснование и выбор веса при расчете средней арифметической взвешенной. Проиллюстрируем сказанное на следующем условном примере. Имеются данные о доле экспорта в стоимости товарной продукции предприятий, выпускающих минеральные удобрения.

Таблица 4.7

Доля экспорта в товарной продукции	Число предприятий	Товарная продукция предприятий группы, млн руб.
0,15	5	200
0,2	7	460
0,3	4	600
Итого:	16	1260

Средняя доля экспорта, исчисленная как средняя арифметическая взвешенная по числу предприятий, является формальной средней.

$$\bar{X}=\frac{0,15\cdot5+0,2\cdot7+0,3\cdot4}{16}=\frac{3,35}{16}=0,209 \quad (20,9\%)$$

Логически обоснованным можно считать выбор в качестве весов объемов товарной продукции в каждой группе предприятий с определенной долей экспорта, поскольку доля экспорта получается делением объема экспорта на товарную продукцию предприятия.

$$\bar{X}=\frac{0,15\cdot200+0,2\cdot460+0,3\cdot600}{1260}=0,24 \quad (24\%)$$

Теперь, в числителе мы получили общую стоимость экспортной продукции, а в знаменателе - общую стоимость всей товарной продукции (16 предприятий). Таким образом, в результате расчета определена средняя доля экспорта предприятий исследуемой совокупности, равная 0, 24 (24%).

Средняя арифметическая взвешенная применяется также при вычислении общей средней для всей совокупности из частных (групповых) средних. Например, одним из современных индикаторов качества жизни населения являются его вклады на счета государственных и коммерческих банков с целью получения дополнительных доходов. Располагая данными о числе вкладчиков и размере вклада за I-й кв. 1995 г. по трем филиалам Сбербанка одного района города, определим средний размер вклада (на 30.03.95).

Таблица 4.8

№ филиала Сбербанка	Число вкладчиков, человек (ω)	Средний остаток по вкладу, млн руб. (X)
589/082	1350	1,50
578/080	1290	1,81
534/085	2050	2,05

Для определения среднего остатка вклада по трем филиалам в целом следует общую сумму остатков по вкладу для всех вкладчиков разделить на общее число вкладчиков. Используя условные обозначения признаков в таблице, имеем формулу:

$$\bar{X}=\frac{\sum_{i=1}^{3}x_i\,\omega_i}{\sum_{i=1}^{3}\omega_i}, \qquad (4.12)$$

где X_i среднее значение признака по каждой группе (в нашем примере - средний остаток по вкладу отдельного филиала); ω_i - веса средней (число вкладчиков по каждому филиалу).

$$\bar{X}=\frac{1,5\cdot1350+1,81\cdot1290+2,05\cdot2050}{1350+1290+2050}=\frac{8562,4}{4690}=1,83 \text{ млн руб.}$$

Средняя арифметическая обладает некоторыми свойствами, которые определяют ее широкое применение в экономических расчетах и в практике статистического исследования.

Свойство 1. Средняя арифметическая постоянной величины равна этой постоянной: $\overline{A} = A$ при $A = \text{const}$.

Свойство 2 (нулевое). Алгебраическая сумма линейных отклонений (разностей) индивидуальных значений признака от средней арифметической равна нулю: $\sum_{i=1}^{n}(X_i - \overline{X}) = \sum d_i = 0$ для первичного ряда и $\sum_{i=1}^{n}(X_i - \overline{X})\, f_i = \sum d_i f_i = 0$ для сгруппированных данных (d_i - линейные (индивидуальные) отклонения от средней, т.е. $X_i - \overline{X}$). Это свойство можно сформулировать следующим образом: сумма положительных отклонений от средней равна сумме отрицательных отклонений. Логически оно означает, что все отклонения от средней в ту и в другую сторону, обусловленные случайными причинами, взаимно погашаются.

Свойство 3 (минимальное). Сумма квадратов отклонений индивидуальных значений признака от средней арифметической есть число минимальное: $\sum_{i=1}^{n}(X_i - \overline{X})^2 = \sum d_i^2 = \min$ или $\sum_{i=1}^{n}(X_i - \overline{X})^2 < \sum_{i=1}^{n}(X_i - A)^2$, где $A = \overline{X} \pm \varepsilon$, что означает: сумма квадратов отклонений индивидуальных значений признака каждой единицы совокупности от средней арифметической всегда меньше суммы квадратов отклонений вариантов признака от любого значения (A), сколь угодно мало отличающегося от средней у выбранной единицы исследуемой совокупности.

Для сгруппированных данных имеем: $\sum_{i=1}^{n}(X_i - \overline{X})^2 f_i = \sum d_i^2 f_i = \min$ или $\sum_{i=1}^{n}(X_i - \overline{X})^2 f_i < \sum_{i=1}^{n}(X_i - A)^2 f_i$.

Минимальное и нулевое свойства средней арифметической применяются для проверки правильности расчета среднего уровня признака; при изучении закономерностей изменения уровней ряда динамики; для нахождения параметров уравнения регрессии при изучении корреляционной связи между признаками.

Рассмотренные свойства выражают сущностные черты средней арифметической. Существуют также расчетные (вычислительные) свойства средней арифметической, имеющие прикладное значение:

если значения признака каждой единицы совокупности (все усредняемые варианты) уменьшить или увеличить на одну и ту же величину A, то и со средней арифметической произойдут аналогичные изменения;

если значения признака каждой единицы совокупности разделить или умножить на какое-либо постоянное число A, то средняя арифметическая уменьшится или увеличится в A раз;

если вес (частоту) каждого значения признака разделить на какое-либо постоянное число A, то средняя арифметическая не изменится.

В настоящее время вычислительные свойства средней арифметической потеряли свою актуальность в связи с использованием ЭВТ при расчете обобщающих статистических показателей.

Средняя гармоническая величина как и средняя арифметическая может быть простой и взвешенной.

Если веса у каждого значения признака равны, то можно использовать среднюю гармоническую простую:

$$\overline{X}_{гарм.} = \frac{n}{\sum_{i=1}^{n} \frac{1}{X_i}}.$$

Однако в статистической практике чаще применяется средняя гармоническая взвешенная. Она используется, как правило, при расчете общей средней из средних групповых.

На основе имеющихся данных по трем филиалам Сбербанка города за II-й кв. 1995 г. имеем (на 30.06.95):

Таблица 4.9

№ филиала Сбербанка	Средний остаток по срочному вкладу, млн руб. (X)	Общая сумма остатков по срочному вкладу всех вкладчиков, млн руб. (ω)
589/082	1,67	1897,8
578/080	2,80	5040,0
534/085	3,25	6987,5

Для определения среднего остатка вклада по трем филиалам в целом необходимо общую сумму остатков по вкладам разделить на общее число вкладчиков. Число вкладчиков по каждому филиалу вычисляется делением общей суммы остатков по вкладам на средний остаток по вкладу. Используя условные обозначения, проставленные в табл. 4.9, расчет среднего остатка по вкладу в целом для всей совокупности банков выполним по формуле:

$$\bar{X} = \frac{\sum_{i=1}^{n} \omega_i}{\sum_{i=1}^{n} \frac{\omega_i}{X_i}} = \frac{1897{,}8+5040+6987{,}5}{\frac{1897{,}8}{1{,}67}+\frac{5040}{2{,}8}+\frac{6987{,}5}{3{,}25}} = \frac{13925{,}3}{5086{,}4} = 2{,}74 \text{ млн руб.}$$

Следовательно, если имеется ряд данных по двум взаимосвязанным показателям, для одного из которых требуется вычислить среднюю величину, и при этом известен итог числителя, а итог знаменателя не известен, но может быть определен как сумма частных от деления численных значений одного показателя на другой, средняя должна вычисляться по формуле средней гармонической взвешенной.

Так как наблюдались одни и те же филиалы банков, можно проследить динамику среднего остатка по вкладу (или среднего вклада) во втором квартале по сравнению с первым. Средний остаток по срочному вкладу с ежемесячной выплатой дохода увеличился на 49, 7% [(2,74/1,83) x 100 – 100%], что составило 910 тыс. руб. Причины, которые могли повлиять на это изменение, прежде всего количество вкладчиков, увеличение суммы вкладов, а также процентные ставки банков.

Итак, при расчете одного и того же показателя - среднего остатка по вкладу в целом по совокупности - в первом случае (табл. 4.7) использовалась средняя взвешенная арифметическая, во втором (табл. 4.8) - средняя гармоническая взвешенная. Это обусловлено прежде всего одной и той же логической формулой для вычисления искомого показателя, но вместе с тем различными исходными данными, которые были представлены в вышеупомянутых таблицах.

Логическая формула вытекает из сущности средней, ее социально-экономического содержания. Средняя величина признака - это отношение. Поэтому прежде чем оперировать цифрами, нужно выяснить, соотношением каких показателей является средняя в данном конкретном случае. Это исходное соотношение необходимо записать словами в виде формулы, которую и называют *логической формулой средней.*

Для нашего случая:

$$\text{Средняя величина остатка по вкладу в целом по трем филиалам} = \frac{\text{Общая сумма остатков по вкладу всех вкладчиков}}{\text{Общее число вкладчиков}}$$

После того как записана логическая формула средней, которую нужно вычислить, необходимо внимательно рассмотреть имеющиеся для вычисления данные и заменить словесные обозначения числителя и знаменателя логической формулы средней соответствующими цифровыми данными, после чего остается только провести необходимые вычисления.

Этот принцип обеспечивает правильный выбор формы средней, а следовательно, и правильное определение величины средней. И еще одно важное свойство принципа логической формулы в том, что здесь не возникает проблема выбора весов средней[1].

При применении **средней геометрической** индивидуальные значения признака представляют собой, как правило, относительные величины динамики, построенные в виде цепных величин, как отношение к предыдущему уровню каждого уровня в ряду динамики. (Причем временные отрезки ряда динамики одинаковые.) Средняя характеризует, таким образом, средний коэффициент роста (см. подробно в гл. 8).

Средняя геометрическая величина используется также для определения равноудаленной величины от максимального и минимального значений признака. Например, страховая фирма заключает договоры на оказание клиентам различных услуг медицинского страхования. В зависимости от категории медицинского учреждения, ассортимента услуг, конкретного рискового случая страховая сумма выплат может изменяться от 100 до 10 000 долл. в год. Средняя сумма выплат по страховке $\sqrt{100 \cdot 10\ 000} = 1000$ долл.

Формула **средней квадратической** используется для измерения степени колеблемости индивидуальных значений признака вокруг средней арифметической в рядах распределения. Так, при расчете показателей вариации среднюю вычисляют из квадратов отклонений индивидуальных значений признака от средней арифметической величины. Пример расчета будет рассмотрен в следующей главе.

[1] См. *Овсиенко В.Е.* Выбор формы средней и о некоторых ошибках, допускаемых в этом вопросе// Вестник статистики. 1989. № 2. С. 16-24.

Контрольные вопросы к главе 4

1. Какова роль статистических показателей в управлении экономикой?
2. Какие могут быть выделены группы обобщающих статистических показателей?
3. Назовите виды относительных величин и охарактеризуйте их значение.
4. Как связаны между собой относительные величины выполнения плана, планового задания и динамики?
5. Охарактеризуйте состав студентов группы по полу и возрасту и рассчитайте относительные величины структуры.
6. Для чего рассчитывают относительные величины координации?
7. Каково значение относительных величин интенсивности. Приведите примеры их использования в анализе уровня экономического и социального развития страны.
8. Почему важно анализировать абсолютные и относительные показатели во взаимосвязи?
9. Что представляет собой средняя величина и в чем состоит ее определяющее свойство?
10. Напишите формулу средней арифметической и приведите пример исчисления средней по формуле:

а) средней арифметической простой,

б) средней арифметической взвешенной.

11. Назовите основные свойства средней арифметической.
12. Как обосновывается выбор весов при расчете взвешенных средних?
13. Для каких целей используется формула средней геометрической?
14. В чем различие между степенными и структурными средними?
15. Использование моды и медианы и их расчет по несгруппированным данным.

Глава 5

СТАТИСТИЧЕСКИЕ РАСПРЕДЕЛЕНИЯ И ИХ ОСНОВНЫЕ ХАРАКТЕРИСТИКИ

5.1. Вариация признака в совокупности и значение ее изучения

Составной частью сводной обработки данных статистического наблюдения является построение рядов распределения. Цель его - выявление основных свойств и закономерностей исследуемой статистической совокупности. В зависимости от того, является ли признак, взятый за основу группировки, качественным или количественным, различают соответственно два типа рядов распределения – атрибутивные и вариационные. Ряды распределения, построенные по качественным признакам, называют *атрибутивными*. Примером атрибутивных рядов может служить распределение населения по полу, характеру труда, национальности, профессии и т.д. Ряды распределения, построенные по количественному признаку, называют *вариационными*. Величины того или иного количественного признака у отдельных единиц совокупности более или менее различаются между собой. Такое различие в величине признака носит название вариации.

В таблице 3.4 приводились различные цены на муниципальное жилье в разных районах города Москвы. Так, в мае 1995 г. цена 1 кв. м муниципального жилья составила по районам (в млн руб): в Митино - 5,2; Южном Бутово - 4,9; северном Бутово - 5,8; Отрадном - 6,5; на Веерной улице - 7,6; в Жулебино - 4,9.

Числовые значения признака, встречающиеся в данной совокупности, называют вариантами значений. В данном примере встречается пять вариантов значений признака, которые колеблются в пределах от 4,9 до 7,6 млн руб. за 1 кв. м. Наличие вариации у отдельных единиц совокупности обусловлено влиянием большого числа факторов на формирование уровня признака. Часто эти факторы могут оказывать разнонаправленное воздействие на уровень анализируемого показателя. Например, снижение цен на строительные материалы может привести к уменьшению цен на 1 кв. м жилья, а увеличение спроса на жилье может повысить эту цену и т.д. В результате совместного влияния различных факторов и складывается цена 1 кв. м муниципального жилья в определенное время. Но есть еще одна группа факторов - экологическая обстановка в разных районах, обусловливающая вариацию рассматриваемых цен в разных районах города. Поэтому при изучении вариации различных показателей можно выделить две группы факторов, формирующих уровень признака в исследуемой совокупности единиц и обусловливающих наличие различий в величине признака у отдельных единиц. *Первую группу составляют факторы, общие для всех единиц изучаемой совокупности. Во вторую группу входят факторы, свойственные конкретным единицам совокупности и определяющие их индивидуальные особенности.*

Изучение характера и степени вариации признаков у отдельных единиц, составляющих изучаемую совокупность, является важнейшим вопросом всякого статистического исследования. Управление процессом развития в желаемом направлении требует определения роли не только каждой из выше названных групп факторов в вариации тех или иных признаков, но и роли отдельных факторов соответствующих групп. Для решения такой задачи в статистике применяются специальные методы исследования вариации, основанные на использовании системы показателей, с помощью которой измеряется вариация.

Представленные выше данные о цене 1 кв. м муниципального жилья по шести районам г. Москвы без какой-либо систематизации образуют так называемый *первичный ряд* данных. При наличии достаточно большого количества вариантов значений признака первичный ряд становится трудно обозримым и непосредственное рассмотрение его не дает представления о распределении единиц по величине признака в совокупности. Первым шагом в упорядочении первичного ряда является его *ранжирование*, т.е. расположение всех вариантов ряда в возрастающем (или убывающем) порядке. В мае 1995 г. по районам Москвы мы располагали такой информацией о цене 1 кв. м муниципального жилья:

Таблица 5.1

Район Москвы	Южное Бутово	Жулебино	Митино	Северное Бутово	Отрадное	Веерная улица
Цена 1 кв. м жилья, млн руб.	4,9	4,9	5.2	5,8	6,5	7,6

Ранжированный ряд данных позволяет сразу увидеть наименьшее и наибольшее значение признака в совокупности, определить расстояние между крайними значениями признака, а также выделить наиболее часто повторяющиеся значения в обследуемой совокупности. Использование ранжированного ряда также позволяет легко разделить все данные по группам. В нашем примере цена 1 кв. м жилья варьирует по районам г. Москвы от 4,9 млн до 7,6 млн руб., т.е. размах вариации составляет 2,7 млн руб. за 1 кв. м, можно также видеть, что в двух районах встречается одна и та же цена - 4,9 млн руб.

Число повторений отдельных вариантов значений признаков называют **частотой повторения**. В дальнейшем частоту повторения значения признака будем обозначать f_i, а сумму частот, равную объему изучаемой совокупности, - $\sum_{i=1}^{k} f_i$, или $n(\sum_{i=1}^{k} f_i = n)$,

где k - число вариантов значений признака.

По характеру вариации различают дискретные и непрерывные признаки. **Дискретные признаки** отличаются друг от друга на некоторую конечную величину, т.е. даны в виде прерывных чисел. Например, тарифный разряд рабочих, число детей в семье, число рабочих на предприятии и т.д. **Непрерывные признаки** могут отличаться один от другого на сколь угодно малую величину и в определенных границах принимать любые значения. Например, заработная плата рабочих, размер среднедушевого денежного дохода, стоимость основных фондов предприятия и т.д.

Способы построения вариационного ряда для этих видов признаков различны. Для построения дискретного ряда с небольшим числом вариантов достаточно перечислить все встречающиеся варианты значений признака, обозначаемые через x_i, а затем подсчитать частоту повторения каждого варианта f_i (например, распределение рабочих по разрядам, студентов по успеваемости и т.п.). Ряд распределения принято оформлять в виде таблицы. Например, распределение рабочих участка по квалификации характеризуется данными табл. 5.2.

Таблица 5.2

Тарифный разряд рабочего x_i	Число рабочих, имеющих этот разряд f_i	Частость w_i	Накопленная частота s_i
2	1	0,05	1
3	5	0,25	6
4	8	0,40	14
5	4	0,20	18
6	2	0,10	20
Итого	20	1,00	

Таким образом, ряд первичных данных, характеризующих квалификацию 20 рабочих, заменен коротким рядом, состоящим из пяти групп. Вместо абсолютного числа рабочих, имеющих определенный разряд, можно установить долю рабочих этого разряда. Частоты, представленные в относительном выражении, называют **частостями** и обозначают w_i (гр. 3 табл. 5.2):

$$W_i = \frac{f_i}{\sum\limits_{i=1}^{k} f_i}.$$

Частости могут быть выражены в долях единицы или в процентах. Замена частот частостями позволяет сопоставлять вариационные ряды с различным числом наблюдений.

В тех случаях, когда число вариантов дискретного признака достаточно велико, а также при анализе вариации непрерывного признака, когда значения признака у отдельных единиц могут вообще не повторяться, строятся **интервальные ряды распределения**. Интервал указывает определенные пределы значений варьирующего признака и обозначается нижней и верхней границами интервала. Такие распределения наиболее распространены в практике статистической работы.

При построении интервальных рядов распределения необходимо прежде всего установить число групп (интервалов), на которые следует разбить все единицы изучаемой совокупности. При

группировке внутри однородных совокупностей появляется возможность применения равных интервалов, величина которых зависит от вариации признака в совокупности и от количества обследованных единиц.

Определение величины интервала h для построения вариационного ряда с равными интервалами производится следующим образом:

1) вычисляется разность между максимальным и минимальным значениями признака первичного ряда (определяется размах вариации, R):

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

2) размах вариации делится на число групп k, т.е. $h = \frac{R}{k}$.

Число групп приближенно определяется по формуле Стэрджесса: $k \approx 1+3{,}322 \lg n$, где n - общее число изучаемых единиц совокупности.

Указанное выражение почти всегда оказывается дробной величиной, которую округляют до целого числа, поскольку количество групп не может быть дробным.

Величина интервала должна определяться в соответствии с точностью данных наблюдения: если исходные данные представлены целыми числами, то рассчитанная величина интервала округляется до ближайшего целого числа; если данные представлены с точностью до 0,1, то величина интервала округляется до целых с десятыми и т.д. (здесь округление производится в большую сторону).

Значение величины интервала позволяет определить границы всех интервалов ряда распределения. Нижнюю границу первого интервала целесообразно принимать равной минимальному значению признака. Имеются и предложения иного рода. Так, нижнюю границу первого интервала рекомендуют определять вычитанием из минимального значения признака половины величины интервала. В каждый интервал включаются варианты, числовые значения которых больше или равны нижней границе и меньше или равны верхней границе. При построении интервальных рядов для непрерывных признаков имеет место совпадение верхних границ предшествующих интервалов и нижних границ следующих за ними интервалов. Здесь должны даваться пояснения, в какой интервал относить единицы совокупности, числовые значения признака у которых совпадают с одной из этих границ.

В интервальных рядах распределения дискретных признаков отнесение единиц совокупности в ту или иную группу не вызывает затруднений, так как между верхней границей одного интервала и нижней границей смежного интервала существует разрыв.

Рассмотрим построение ряда распределения по первичным данным о размере прибыли 20 коммерческих банков за год (млрд руб.):

x_i - 3,7; 4,3; 6,7; 5,6; 5,1; 8,1; 4,6; 5,7; 6,4; 5,9; 5,2; 6,2; 6,3; 7,2; 7,9; 5,8; 4,9; 7,6; 7,0; 6,9.

Определяем количество групп интервального вариационного ряда: $k \approx 1{+}3{,}322 \lg 20 = 1{+}3{,}322 \cdot 1{,}301 = 5{,}32$. Округляя, получим число групп, равное 5. Величина интервала составит 0,9 млрд руб. $(\frac{8{,}1-3{,}7}{5})$. В результате группировки получим ряд распределения банков по величине полученной прибыли за год (см. графы 1, 2 табл. 5.3). Значения признака у отдельных единиц совпали с границами интервалов (например, значения 4,6 и 6,4). Так как минимальное значение признака (3,7 млрд руб.) совпадает с нижней границей первого интервала и включается в этот интервал, эти значения следует включать в тот интервал, нижняя граница которого совпадает с указанным значением (например, значение признака 4,6 млрд руб. включается во второй интервал, а значение признака 6,4 млрд руб. - в четвертый).

Для пояснения в первой строке графы 1 табл. 5.3 поставлен знак (−). Этот знак соответствует принципу «исключительно» и означает, что значения признака, совпадающие с верхней границей интервала в этот интервал не включаются, а попадают в следующий интервал. Если ставится знак (+), это соответствует принципу «включительно» и означает, что значения признака, совпадающие с верхней границей интервала, включаются в эту группу.

Таблица 5.3

Размер прибыли, млрд руб.	Число банков	Накопленная частота
1	2	3
3,7 - 4,6 (−)	2	2
4,6 - 5,5	4	6
5,5 - 6,4	6	12
6,4 - 7,3	5	17
7,3 - 8,1	3	20
Итого	20	

Если приведен вариационный ряд с неравными интервалами, то для правильного представления о характере распределения необходимо рассчитать *абсолютную* или *относительную плотности распределения,* для определения которых вычисляют отношение частот или частостей к величине интервала: абсолютная плотность $m_i^{(a)} = \frac{f_i}{h_i}$; относительная плотность $m_i^{(o)} = \frac{w_i}{h_i}$. Эти показатели используют для преобразования интервалов, что бывает необходимо при сравнительной оценке данных, собранных по различным совокупностям и по-разному обработанных. Например, по двум группам банков одного из штатов США, различающихся своими размерами, известно их распределение по уровню рентабельности активов (табл. 5.4).

Таблица 5.4

Группы банков по стоимости активов

От 50 до 100 млн долл.		От 100 до 300 млн долл.	
группы банков по рентабельности активов	число банков, % к итогу	группы банков по рентабельности активов	число банков, % к итогу
1	2	3	4
0,8 - 1,0	20	0,6 - 0,8	10
1,0 - 1,5	40	0,8 - 1,0	30
1,5 - 2,0	30	1,0 - 1,2	10
		1,2 - 1,4	15
2,0 и более	10	1,4 - 1,8	25
Итого	100	1,8 - 2,0	10
		Итого	100

Для перегруппировки данных с целью получения сопоставимых групп и их анализа можно воспользоваться следующей группировкой банков по уровню рентабельности активов (см. графу 1 табл. 5.4а).

Для подсчета частот в такой группировке возникает необходимость расщепления интервалов для группы банков со стоимостью активов от 50 до 100 млн долл.

Если известна относительная плотность распределения, то частости соответствующего интервала можно определить умножением плотности на величину интервала, т.е. $w_i = m_i^{(o)} \cdot h_i$.

По данным табл. 5.4, определяем плотности распределения группы банков со стоимостью активов от 50 до 100 млн долл. для интервалов по уровню рентабельности активов (см. графы 1 и 2 табл. 5.4):

второй интервал (1,0 - 1,5) $m_2^{(о)} = 80\% \left(\frac{40}{0,5}\right)$,

третий интервал (1,5 - 2,0) $m_3^{(о)} = 60\% \left(\frac{30}{0,5}\right)$.

Тогда количество банков этой группы с рентабельностью активов от 1,0 до 1,4 составит 32% (80•0,4). В группу с рентабельностью 1,4 - 1,8 попадает 26% от общего количества банков (80•0,1 + 60•0,3), а в группу с рентабельностью 1,8 - 2,0 - 12% банков (60•0,2).

Для второй группы банков с размером активов от 100 до 300 млн долл. число банков в группе 1,0 - 1,4 по рентабельности активов определяется суммированием частостей в третьей и четвертой строках табл. 5.4.

Таблица 5.4а

Группировка банков штата по уровню рентабельности активов

Группы банков по рентабельности активов	Количество банков со стоимостью активов, в % к итогу	
	от 50 до 100 млн долл.	от 100 до 300 млн долл.
1	2	3
0,6 - 0,8	-	10
0,8 - 1,0	20	30
1,0 - 1,4	32	25
1,4 - 1,8	26	25
1,8 - 2,0	12	10
2,0 и более	10	-
Итого	100	100

5.2. Основные характеристики и графическое изображение вариационного ряда

Для целей анализа и сравнительной характеристики различных рядов распределения применяются обобщающие показатели вариационного ряда. Система таких показателей может быть наглядно представлена при сравнении особенностей нескольких рядов распределения.

Так, на рис. 5.1 кривые распределения 1 и 2 имеют одинаковый размах вариации и характер распределения частот, но отличаются величиной варьирующего признака, являющегося центром группирования (отмечена на оси X).

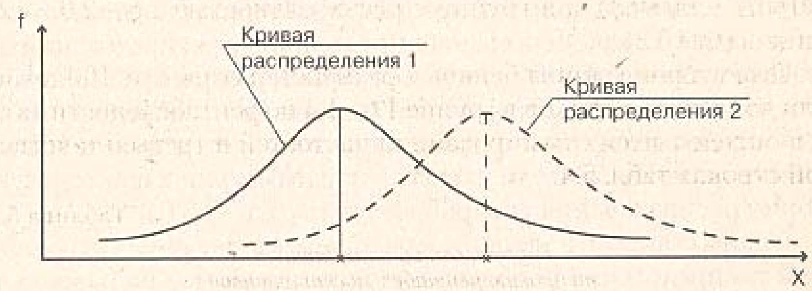

Рис. 5.1. **Кривые распределения с разными центрами группирования**

Характеристики центра группирования, таким образом, составляют одну из групп обобщающих показателей. В качестве них используют среднюю арифметическую, моду и медиану.

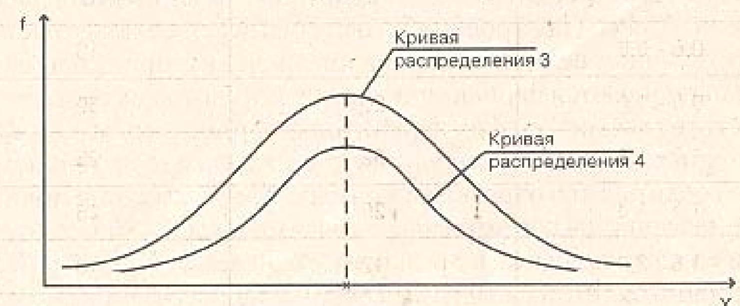

Рис. 5.2. **Кривые распределения с различной вариацией признака**

Представленные на рисунке 5.2 кривые распределения 3 и 4 имеют один и тот же центр группирования и симметричное расположение частот вокруг него, но отличаются пределами вариации. Следовательно, кроме показателей центра группирования, для характеристики особенностей распределения необходимо иметь представление о степени колеблемости признака и знать показатели степени вариации. Эти две группы показателей - показатели центра распределения и вариации признака - имеют особое значение при принятии решений в управлении.

Ряды распределения могут иметь один и тот же центр группирования, одинаковые пределы варьирования признака, симметричный характер расположения частот, но разную степень вытянутости вдоль оси ординат, которая характеризуется показателями эксцесса. И наконец, сравнение различных распределений показывает, что они могут отличаться характером распределения частот относительно центра; степень отклонения распределения частот от симметричной формы характеризуется показателями асимметрии. Показатели эксцесса и асимметрии характеризуют форму распределения (подробнее см. параграф 5.6).

Таким образом, в зависимости от характеризуемых особенностей распределения обобщающие показатели можно разбить на три группы:

1) показатели центра распределения (центра группирования);
2) показатели степени вариации;
3) показатели формы распределения.

Графическое изображение рядов распределения облегчает их анализ и позволяет судить о форме распределения. Для графического изображения дискретного ряда применяют **полигон распределения**. Для его построения на оси абсцисс отмечают точки, соответствующие величине вариантов значений признака, из них восстанавливаются перпендикуляры, длина которых соответствует частоте (частости) этих вариантов по принятому масштабу на оси ординат. Вершины перпендикуляров в последовательном порядке соединяются отрезками прямых. Для замыкания полигона крайние вершины соединяются с точками на оси абсцисс, отстоящими на одно значение в принятом масштабе от X_{max} и X_{min}. Такое построение полигона облегчает восприятие его графического изображения.

Воспользуемся данными, приведенными в табл. 5.2, для построения полигона (рис. 5.3).

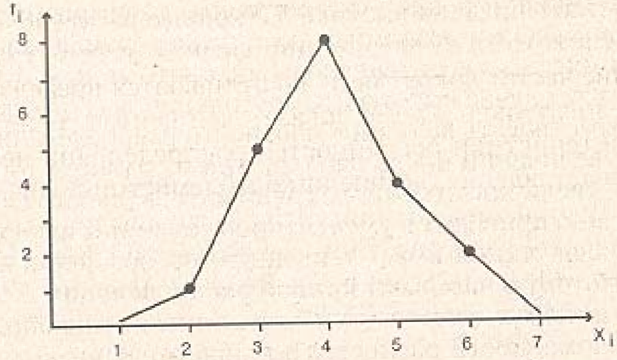

Рис. 5.3. **Распределение рабочих участка по квалификации: X_i - тарифный разряд; f_i - число рабочих**

Для графического изображения интервальных вариационных рядов применяется **гистограмма**. Она строится так: на оси абсцисс откладываются равные отрезки, которые в принятом масштабе соответствуют величине интервалов вариационного ряда. На отрезках строят прямоугольники, площади которых пропорциональны частотам (или частостям) интервала.

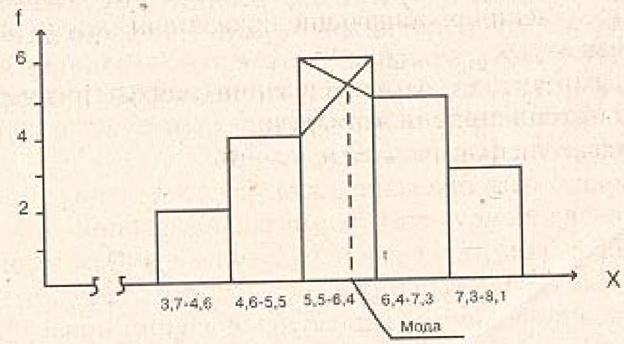

Рис. 5.4. **Распределение коммерческих банков по размеру прибыли: X - размер прибыли, млрд руб., f - число банков**

На рис. 5.4. изображена гистрограмма ряда распределения банков по размеру прибыли (по данным табл. 5.3). Гистрограмма может быть преобразована в полигон распределения, если середины верхних сторон прямоугольников соединяются отрезками прямых. Две крайние точки прямоугольников замыкаются по оси абсцисс на середины интервалов, в которых частоты (частости) равны нулю. При построении гистрограммы для вариационного ряда с неравными интервалами следует по оси ординат наносить показатели плотности интервалов (абсолютные или относительные). В этом случае высоты прямоугольников гистрограммы будут соответствовать величине плотности распределения.

При увеличении числа наблюдений из одной и той же совокупности увеличивается число групп интервального ряда, что соответственно приводит к уменьшению величины интервала. При этом ломаная линия имеет тенденцию превращения в плавную кривую, которую называют **кривой распределения**. Кривая распределения характеризует в обобщенном виде вариацию признака и закономерности распределения частот внутри однокачественной совокупности.

В ряде случаев для изображения вариационных рядов используется **кумулятивная кривая** (кумулята). Для ее построения надо рассчитать накопленные частоты и частости. Накопленные частоты показывают, сколько единиц совокупности имеют значения признака не больше, чем рассматриваемое значение, и определяются последовательным суммированием частот интервалов. Построим кумулятивную кривую по данным табл. 5.3 о распределении банков по размеру прибыли. (Накопленные частоты рассчитаны в гр. 3 табл. 5.3.) При построении кумуляты интервального ряда распределения нижней границе первого интервала соответствует частота, равная нулю, а верхней границе - вся частота данного интервала. Верхней границе второго интервала соответствует накопленная частота, равная сумме частот первых двух интервалов, и т.д. (см. рис. 5.5).

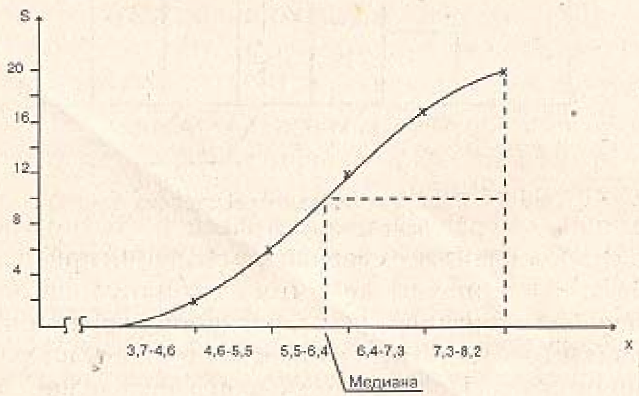

Рис. 5.5. **Кумулята ряда распределения банков по размеру прибыли:**
***X* - размер прибыли, млрд руб., *S* - накопленные частоты**

Изображение вариационного ряда в виде кумуляты особенно удобно при сравнении вариационных рядов, а также в экономических исследованиях, в частности для анализа концентрации производства.

5.3. Показатели центра распределения

Для характеристики среднего значения признака в вариационном ряду используются средняя арифметическая, мода и медиана. Общие понятия о средних величинах и их свойствах даны в гл. 4. В данном параграфе рассмотрим расчет показателей центра распределения для вариационных рядов.

Средняя арифметическая для дискретного ряда распределения рассчитывается по формуле

$$\bar{x} = \frac{\sum_{i=1}^{k} x_i f_i}{\sum_{i=1}^{k} f_i}, \qquad (5.1)$$

где x_i - варианты значений признака;
f_i - частота повторения данного варианта.

В интервальном вариационном ряду средняя арифметическая определяется по формуле

$$\bar{x} = \frac{\sum_{i=1}^{k} x'_i f_i}{\sum_{i=1}^{k} f_i}, \qquad (5.1а)$$

где x'_i - середина соответствующего интервала.

Пример расчета средней арифметической для интервального вариационного ряда приведен далее, в табл. 5.5.

В отличие от средней арифметической, рассчитываемой на основе использования всех вариантов значений признака, мода и медиана характеризуют величину варианта, занимающего определенное положение в ранжированном вариационном ряду. **Медиана (Me)** соответствует варианту, стоящему в середине ранжированного ряда. Положение медианы определяется ее номером $N_{Me} = \frac{n+1}{2}$, где n - число единиц в совокупности.

Используем данные примера, приведенные в табл. 5.2, для определения медианы и моды. $N_{Me}=\frac{20+1}{2}=10{,}5$, т.е. медиана равна средней арифметической из 10-го и 11-го значений признака. По накопленным частотам определяем, что 10-я и 11-я единицы ряда имеют величину признака, равную четвертому разряду, т.е. медиана равна четвертому разряду.

Мода (Мо) - наиболее часто встречающееся значение признака в совокупности - для данного ряда распределения также равна четвертому разряду (этому разряду соответствует максимальная частота, равная 8). В интервальном ряду распределения сразу можно указать только интервал, в котором будут находиться мода или медиана. Для определения их величины используются следующие формулы[1]:

$$Me = x_{Me} + h\,\frac{\frac{n+1}{2} - S_{(-1)}}{f_{Me}}, \qquad (5.2)$$

где x_{Me} - нижняя граница медианного интервала;
h - величина интервала;
$S_{(-1)}$ - накопленная частота интервала, предшествующего медианному;
f_{Me} - частота медианного интервала.

Используя данные примера, приведенные в табл. 5.3, рассчитаем медиану. По накопленным частотам (гр. 3) определяем, что медиана находится в интервале 5,5 - 6,4. Тогда

$$Me = 5{,}5 + 0{,}9\,\frac{\frac{20+1}{2} - 6}{6} = 6{,}175 \text{ млрд руб.}$$

Таким образом, 50% банков имеют прибыль менее 6,175 млрд руб., а 50% банков - более 6,175 млрд руб.

Наибольшая частота соответствует также интервалу 5,5 - 6,4, т.е. мода должна находиться в этом интервале. Ее величину определяем по формуле:

$$M_o = x_{Mo} + h\,\frac{f_{Mo} - f_{(-1)}}{\left[f_{Mo} - f_{(+1)}\right] + \left[f_{Mo} - f_{(-1)}\right]}, \qquad (5.3)$$

[1] Формула медианы выведена исходя из предположения о том, что плотность внутри интервала остается постоянной.

где X_{Mo} - начало модального интервала;
f_{Mo} - частота, соответствующая модальному интервалу;
$f_{(-1)}$ - предмодальная частота;
$f_{(+1)}$ - послемодальная частота.

Приведенная формула моды может быть использована в вариационных рядах с равными интервалами.

$$M_o = 5{,}5 + 0{,}9\,\frac{6-4}{(6-4)+(6-5)} = 6{,}10 \text{ млрд руб.}$$

Таким образом, в данной совокупности наиболее часто встречается размер прибыли 6,10 млрд руб.

Медиану и моду можно определить графически. Медиана определяется по кумуляте (см. рис. 5.5). Для ее определения высоту наибольшей ординаты, которая соответствует общей численности совокупности, делят пополам. Через полученную точку проводят прямую, параллельную оси абсцисс, до пересечения ее с кумулятой. Абсцисса точки пересечения является медианой.

Мода определяется по гистрограмме распределения. Для этого правую вершину модального прямоугольника соединяем с правым верхним углом предыдущего прямоугольника, а левую вершину модального прямоугольника - с левым верхним углом последующего прямоугольника. Абсцисса точки пересечения этих прямых и будет модой распределения (см. рис. 5.4).

Основной характеристикой центра распределения является средняя арифметическая, опирающаяся на всю информацию об изучаемой совокупности единиц. Однако в ряде случаев средняя арифметическая должна быть дополнена и даже заменена модальным значением или медианой. Например, в статистическом контроле качества продукции удобнее пользоваться медианой, а не средней арифметической, так как определение медианы для ранжированного ряда данных не требует специального расчета. Кроме того, она не чувствительна к крайним значениям взятой контрольной пробы. В рядах с открытыми интервалами также целесообразнее пользоваться в качестве характеристики центра распределения модой и медианой. Мода применяется при изучении спроса населения на товары народного потребления (например, на обувь, одежду и т.д.), когда интерес представляет определение модального размера, т.е. размера, пользующегося наибольшим спросом.

Если сформулировать общие правила для выбора средней арифметической, моды или медианы в качестве показателя центра распределения, то можно сказать, что в симметричных рядах все названные показатели равноправны, поскольку $\overline{x} = Me = Mo$, но предпочтение отдается средней арифметической.

Для асимметричных рядов распределения медиана часто является предпочтительной характеристикой центра распределения, поскольку занимает положение между средней арифметической и модой.

5.4. Показатели вариации (колеблемости) признака

Средняя величина дает обобщающую характеристику всей совокупности изучаемого явления. Однако два ряда распределения, имеющие одинаковую среднюю арифметическую величину, могут значительно отличаться друг от друга по степени колеблемости (вариации) величины изучаемого признака. Если индивидуальные значения признака ряда мало отличаются друг от друга, то средняя арифметическая будет достаточно показательной характеристикой данной совокупности. Если же ряд распределения характеризуется значительным рассеиванием индивидуальных значений признака, то средняя арифметическая будет ненадежной характеристикой этой совокупности и ее практическое применение будет ограничено.

Для измерения вариации признака применяются различные абсолютные и относительные показатели. К абсолютным показателям вариации относятся размах колебаний, среднее линейное отклонение, дисперсия, среднее квадратическое отклонение и квартильное отклонение.

Размах колебаний, или **размах вариации**, представляет собой разность между максимальным и минимальным значениями признака в изучаемой совокупности:

$$R = x_{max} - x_{min}\,. \tag{5.4}$$

Безусловным достоинством этого показателя является простота расчета. Однако размах вариации зависит от величины только крайних значений признака, поэтому область его применения ограничена достаточно однородными совокупностями. В частности, на практике он находит применение в предупредительном контроле качества продукции.

Точнее характеризует вариацию признака показатель, основанный на учете колеблемости всех значений признака. Поскольку средняя арифметическая является обобщающей характеристи-

кой свойств совокупности, большинство показателей вариации основано на рассмотрении отклонений значений признака отдельных единиц совокупности от этой величины. К таким показателям относятся среднее линейное отклонение, дисперсия и среднее квадратическое отклонение, представляющие собой среднюю арифметическую из отклонений индивидуальных значений признака от средней арифметической. Среднее линейное отклонение рассчитывается из отклонений в первой степени, дисперсия и среднее квадратическое - из отклонений во второй степени. Так как алгебраическая сумма отклонений индивидуальных значений признака от средней арифметической (согласно нулевому свойству) всегда равна нулю, то для расчета среднего линейного отклонения используется арифметическая сумма отклонений, т.е. суммируются абсолютные значения индивидуальных отклонений значений признака независимо от знака.

Среднее линейное отклонение $\overline{d}$ вычисляется по следующим формулам:

для несгруппированных данных

$$\overline{d} = \frac{\sum_{i=1}^{n} \left| x_i - \overline{x} \right|}{n}\,, \tag{5.5}$$

для сгруппированных данных

$$\overline{d} = \frac{\sum_{i=1}^{k} \left| x_i - \overline{x} \right| f_i}{\sum_{i=1}^{k} f_i}\,. \tag{5.5a}$$
[1]

Дисперсия σ^2 - средняя из квадратов отклонений вариантов значений признака от их средней величины. Дисперсия рассчитывается по следующим формулам:

для несгруппированных данных

$$\sigma^2 = \frac{\sum_{i=1}^{n} (x_i - \overline{x})^2}{n}\,, \tag{5.6}$$

[1]При расчете среднего линейного (квадратического) отклонения для интервального вариационного ряда рассчитывают отклонения центральных значений интервала от средней арифметической, т.е. величины $x_i' - \overline{x}$.

для сгруппированных данных (вариационного ряда)

$$\sigma^2 = \frac{\sum_{i=1}^{k}(x_i - \bar{x})^2 f_i}{\sum_{i=1}^{k} f_i} \qquad (5.6а)$$

Формулу для расчета дисперсии можно преобразовать следующим образом:

$$\sigma^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n} = \frac{\sum_{i=1}^{n}\left[x_i^2 - 2x_i\bar{x} + (\bar{x})^2\right]}{n} = \frac{\sum_{i=1}^{n}x_i^2 - 2\bar{x}\sum_{i=1}^{n}x_i + n(\bar{x})^2}{n} = \frac{\sum_{i=1}^{n}x_i^2}{n} - (\bar{x})^2,$$

т.е. дисперсия равна разности средней из квадратов значений признака и квадрата средней арифметической. При пользовании этой формулой исключается дополнительная процедура по расчету отклонений индивидуальных значений признака от $\bar{x}$ и исключается ошибка в расчете, связанная с округлением отклонений $(x_i - \bar{x})$.

Дисперсия обладает рядом свойств, некоторые из них позволяют упростить ее вычисления:

1) дисперсия постоянной величины равна нулю;

2) если все варианты значений признака уменьшить на одно и то же число, то дисперсия не уменьшится;

3) если все варианты значений признака уменьшить в одно и то же число раз (k раз), то дисперсия уменьшится в k^2 раз.

Среднее квадратическое отклонение σ представляет собой корень квадратный из дисперсии:

для несгруппированных данных

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n}}, \qquad (5.7)$$

для вариационного ряда

$$\sigma = \sqrt{\frac{\sum_{i=1}^{k}(x_i - \bar{x})^2 f_i}{\sum_{i=1}^{k} f_i}}. \qquad (5.7а)$$

Размах вариации, среднее линейное и среднее квадратическое отклонение являются величинами именованными. Они

имеют те же единицы измерения, что и индивидуальные значения признака.

Дисперсия и среднее квадратическое отклонение - наиболее широко применяемые показатели вариации. Объясняется это тем, что они входят в большинство теорем теории вероятностей, служащих фундаментом математической статистики. Кроме того, дисперсия может быть разложена на составные элементы, позволяющие оценить влияние различных факторов, обусловливающих вариацию признака. В последующих главах будет показано, как дисперсия используется для построения показателей тесноты корреляционной связи, при оценке результатов выборочных наблюдений, в дисперсионном анализе и т.д.

Расчет показателей вариации для банков, сгруппированных по размеру прибыли, показан в табл. 5.5.

Таблица 5.5

Размер прибыли, млрд руб.	Число банков	Расчетные показатели				
		x'_i	$x'_i f_i$	$x'_i - \bar{x}$	$\lvert x'_i - \bar{x}\rvert f_i$	$(x_i - \bar{x})^2 f_i$
1	2	3	4	5	6	7
3,7 - 4,6 (–)	2	4,15	8,30	-1,935	3,870	7,489
4,6 - 5,5	4	5,05	20,20	-1,035	4,140	4,285
5,5 - 6,4	6	5,95	35,70	-0,135	0,810	0,109
6,4 - 7,3	5	6,85	34,25	+0,765	3,825	2,926
7,3 - 8,2	3	7,75	23,25	+1,665	4,995	8,317
Итого	20		121,70		17,640	23,126

Примечание. Знак (–) в первом интервале указывает, что значения признака, совпадающие с верхней границей интервала, включаются в следующий интервал. Это связано с тем, что верхняя граница последнего интервала больше X_{max} в данной совокупности.

$$\bar{x} = \frac{121,70}{20} = 6,085 \text{ млрд руб.}$$

$$\bar{d} = \frac{17,640}{20} = 0,882 \text{ млрд руб.}$$

$$\sigma^2 = \frac{23,126}{20} = 1,156$$

$$\sigma = \sqrt{1,156} = 1,075 \text{ млрд руб.}$$

Среднее линейное и среднее квадратическое отклонения показывают на сколько в среднем колеблется величина признака у единиц исследуемой совокупности. Так, в данном примере средняя величина колеблемости размера прибыли составляет: по среднему линейному отклонению 0,882 млрд руб., по среднему квадратическому отклонению - 1,075 млрд руб.

По свойству мажорантности средних величин (см. гл. 4) среднее квадратическое отклонение всегда больше среднего линейного отклонения. Если распределение признака близко к нормальному или симметричному распределению, то между σ и $\bar{d}$ существует взаимосвязь: $\sigma = 1{,}25\bar{d}$, или $\bar{d} = 0{,}8\sigma$.

Среднее квадратическое отклонение показывает как расположена основная масса единиц совокупности относительно средней арифметической. В соответствии с теоремой П.Л. Чебышева (1821-1894) можно утверждать, что независимо от формы распределения 75% значений признака попадают в интервал $\bar{x} \pm 2\sigma$, а по крайней мере 89% всех значений попадают в интервал $\bar{x} \pm 3\sigma$.

Для нормального распределения вероятности попадания значений признака в интервалы, отклоняющиеся от средней на то или иное число средних квадратических отклонений, будут рассмотрены в параграфе 5.6.

При вычислении обобщающих статистических характеристик для интервальных рядов распределений действительные значения признака заменяются центральными значениями интервалов, которые более или менее отличаются от средней арифметической из значений, включенных в соответствующий интервал. Так, в первый интервал попали значения 3,7 и 4,3, их средняя арифметическая $\bar{x}$ составит 4,0 млрд руб. $\left(\frac{3{,}7+4{,}3}{2}\right)$, а центральное значение первого интервала – 4,15 млрд руб. (см. графу 3 табл. 5.5). Во втором интервале средняя прибыль составит 4,95 млрд руб. $\left(\frac{4{,}6+4{,}9+5{,}1+5{,}2}{4}\right)$, в третьем - 5,92 млрд, в четвертом - 6,84 млрд руб. и в пятом - 7,87 млрд руб. Таким образом, расхождение в величине $\bar{x}_i$ и x'_i имеет место во всех интервалах, причем наименьшие отклонения $(x'_i - \bar{x}_i)$ соответственно равные +0,03 и +0,01 имеют место в третьем и четвертом интервалах, а наибольшее отклонение, равное +0,15, наблюдается в первом интервале. Можно видеть, что чем больше частота соответствующего интервала, тем меньше расхождение в величине средней арифметической из значе-

ний признака, включенных в интервал, и центральным значением интервала. Этим объясняется требование достаточно большого числа наблюдений в каждом интервале. Однако, чем далее удалены от центра распределения значения признака, тем больше расхождения между $\bar{x}_i$ и x'_i.

Если при расчете средней арифметической для достаточно симметричного ряда распределения расхождения между $\bar{x}_i$ и x'_i не оказывают существенного влияния на ее отклонение от средней арифметической, рассчитанной по первичным данным, то при расчете дисперсии этот факт приведет к появлению систематической ошибки. В.Ф. Шеппард установил, что ошибка в дисперсии, вызванная применением сгруппированных данных, при расчете составляет 1/12 квадрата величины интервала, т.е. скорректированная дисперсия равна: $\sigma^2 - \frac{1}{12}h^2$.

Поправка Шеппарда должна применяться при условиях, если распределение:

1) относится к признаку с непрерывным характером вариации;

2) характеризуется тесной близостью с осью абсцисс на концах кривой;

3) построено по большому числу исходных данных (n>500).

Если в качестве показателя центра распределения используется медиана, то для характеристики вариации признаков в совокупности можно применить так называемое **квартильное отклонение** Q. Этот показатель также можно применить вместо размаха вариации, чтобы избежать недостатков, связанных с использованием крайних значений.

$$Q = \frac{Q_3 - Q_1}{2}, \tag{5.8}$$

где Q_1 и Q_3 - соответственно первая и третья квартили распределения.

Квартили - это значения признака в ранжированном ряду распределения, выбранные таким образом, что 25% единиц совокупности будут меньше по величине Q_1; 25% единиц будут заключены между Q_1 и Q_2; 25% - между Q_2 и Q_3 и остальные 25% превосходят Q_3. Квартили определяются по формулам, аналогичным приведенной выше формуле для расчета медианы:

$$Q_1 = x_{Q_1} + h\frac{\frac{n+1}{4} - S_{(-1)}}{f_{Q_1}}, \tag{5.9}$$

где x_{Q1} - нижняя граница интервала, в котором находится первая квартиль; $S_{(-1)}$ - сумма накопленных частот интервалов, предшествующих интервалу, в котором находится первая квартиль; f_{Q1} - частота интервала, в котором находится первая квартиль;

$$Q_3 = x_{Q_3} + h\frac{0{,}75(n+1) - S_{(-1)}}{f_{Q_3}} \qquad (5.10)$$

(условные обозначения те же, что и для величины Q_1). По данным табл. 5.5, вычислим:

$$Q_1 = 4{,}6 + 0{,}9\frac{21/4 - 2}{4} = 5{,}331 \text{ млрд руб.}$$

$$Q_2 = 5{,}5 + 0{,}9\frac{21/2 - 6}{6} = 6{,}175 \text{ млрд руб.}$$

$$Q_3 = 6{,}4 + 0{,}9\frac{21 \cdot 3/4 - 12}{5} = 7{,}075 \text{ млрд руб.}$$

Используя результаты приведенных расчетов, получим квартильное отклонение:

$$Q = \frac{Q_3 - Q_1}{2} = \frac{7{,}075 - 5{,}331}{2} = 0{,}872 \text{ млрд руб.}$$

В симметричных или умеренно асимметричных распределениях $Q \approx 2/3\sigma$. Так как на квартильное отклонение не влияют отклонения всех значений признака, то его использование следует ограничить случаями, когда определение среднего квадратического отклонения затруднительно или невозможно. В частности, этот показатель может быть рекомендован для рядов распределения с открытыми интервалами.

При сравнении колеблемости различных признаков в одной и той же совокупности или же при сравнении колеблемости одного и того же признака в нескольких совокупностях с различной величиной средней арифметической пользуются *относительными показателями вариации*. Эти показатели вычисляются как отношение абсолютных показателей вариации к средней арифметической (или медиане). Используя в качестве абсолютного показателя вариации размах вариации, среднее линейное отклонение, среднее квадратическое отклонение и квартильное отклонение, получим относительные показатели колеблемости (чаще всего они выражаются в процентах):

коэффициент осцилляции $K_R = \frac{R}{\bar{x}} \cdot 100\%$, (5.11)

относительное линейное отклонение $K_{\bar{d}} = \frac{\bar{d}}{\bar{x}} \cdot 100\%$, (5.12)

коэффициент вариации $v = \frac{\sigma}{\bar{x}} \cdot 100\%$, (5.13)

относительный показатель квартильной вариации

$$K_Q = \frac{Q}{Me} \cdot 100\%, \text{ или } K_Q = \frac{Q_3 - Q_1}{2Q_2} \cdot 100\%. \qquad (5.14)$$

Наиболее часто применяемый показатель относительной колеблемости - **коэффициент вариации**. Его используют не только для сравнительной оценки вариации, но и для характеристики однородности совокупности. Совокупность считается однородной, если коэффициент вариации не превышает 33% (для распределений, близких к нормальному). Для примера, приведенного в табл. 5.5 относительные показатели вариации получились следующими:

$$K_R = \frac{8{,}1 - 3{,}7}{6{,}085} \cdot 100\% = 72{,}31\%$$

$$K_{\bar{d}} = \frac{0{,}882}{6{,}085} \cdot 100\% = 14{,}5\%$$

$$v = \frac{1{,}075}{6{,}085} \cdot 100\% = 17{,}6\%$$

$$K_Q = \frac{0{,}872}{6{,}175} \cdot 100\% = 14{,}12\%$$

Основываясь на коэффициенте вариации, можно сделать вывод, что по размеру прибыли совокупность банков является однородной.

Применим коэффициент вариации для сравнительной оценки колеблемости одного признака в различных совокупностях. Используя данные примера, представленного в табл. 3.4,рассчитаем среднее квадратическое отклонение в уровне цен на муниципальное жилье в Москве в январе и мае 1995 г. За январь среднее квадратическое отклонение составило 0,7603 млн руб., а в мае - 0,9754 млн руб. Означает ли это, что в мае возросла вариация цен на муниципальное жилье в Москве? Если принять во внимание что σ рассчитывается по отклонениям от соответствующей средней арифметической, нужно сопоставить средний уровень цен на муниципальное жилье в Москве за январь и май 1995 г. В январе средняя цена 1 кв. м муниципального жилья составляла 3,88 млн руб., а в мае - 5,82 млн руб. Следовательно, вывод об увеличении или уменьшении вариации цен на муниципальное жилье в мае 1995 г. может быть сделан после расчета показателя относительной колеблемости уровня признака, т.е. коэффициента вариации.

Расчет коэффициента вариации цен 1 кв. м муниципального жилья в Москве приводит к следующим результатам:

в январе 1995 г. $v_1 = \frac{0{,}7603}{3{,}88} \cdot 100\% = 19{,}59\%$,

в мае 1995 г. $v_2 = \frac{0{,}9754}{5{,}82} \cdot 100\% = 16{,}77\%$.

Следовательно, в мае 1995 г. вариация цен 1 кв. м муниципального жилья в Москве снизилась по сравнению с январем 1995 г., поскольку $v_2 < v_1$.

Если статистическая совокупность разбита на группы по какому-либо признаку, то для оценки влияния различных факторов, определяющих колеблемость индивидуальных значений признака, можно воспользоваться разложением дисперсии на составляющие: на межгрупповую и внутригрупповую дисперсии.

Если рассчитать дисперсию признака по всей изучаемой совокупности, т.е. общую дисперсию σ_0^2, то полученный показатель будет характеризовать вариацию признака как результат влияния всех факторов, определяющих индивидуальные различия единиц совокупности. Если же поставить дальнейшую задачу - выделить в составе общей дисперсии ту ее часть, которая обусловлена влиянием какого-либо определенного фактора, то следует разбить изучаемую совокупность на группы, положив в основу группировки интересующий нас фактор. Затем нужно изучить раздельно вариацию признака внутри однородных в отношении данного фактора групп и изменения в величине признака от группы к группе. Выполнение такой группировки позволяет разложить общую дисперсию признака на две дисперсии, одна из которых будет характеризовать часть вариации, обусловленную влиянием фактора, положенного в основу группировки, а вторая - вариацию, происходящую под влиянием прочих факторов (кроме фактора, положенного в основу группировки).

Отклонение индивидуальных значений признака от общей средней $\bar{x}_0$ можно представить так:

$$x_{ij} - \bar{x}_0 = (x_{ij} - \bar{x}_j) + (\bar{x}_j - \bar{x}_0).$$

Вариацию, обусловленную влиянием фактора, положенного в основу группировки, характеризует **межгрупповая дисперсия** δ^2, которая является мерой колеблемости частных средних по группам $\bar{x}_j$ вокруг общей средней $\bar{x}_0$ и исчисляется по формуле

$$\delta^2 = \frac{\sum_{j=1}^{k}(\bar{x}_j - \bar{x}_0)^2 n_j}{\sum_{j=1}^{k} n_j}, \qquad (5.15)$$

где k - число групп; n_j - число единиц в j-й группе; $\bar{x}_j$ - частная средняя по j-й группе; $\bar{x}_0$ - общая средняя по совокупности единиц.

Вариацию, обусловленную влиянием прочих факторов, характеризует в каждой группе **внутригрупповая дисперсия**, σ_j^2.

$$\sigma_j^2 = \frac{\sum_{i=1}^{n_j}(x_{ij} - \bar{x}_j)^2}{n_j}. \qquad (5.16)$$

По совокупности в целом вариация значений признака под влиянием прочих факторов характеризуется средней из внутригрупповых дисперсий:

$$\overline{\sigma^2} = \frac{\sum_{j=1}^{k}\sum_{i=1}^{n_j}(x_{ij} - \bar{x}_j)^2}{\sum_{j=1}^{k} n_j}, \text{ или } \overline{\sigma^2} = \frac{\sum_{j=1}^{k}\sigma_j^2 n_j}{\sum_{j=1}^{k} n_j}. \qquad (5.17)$$

Между общей дисперсией σ_0^2, средней из внутригрупповых дисперсий $\overline{\sigma^2}$ и межгрупповой δ^2 дисперсиями существует соотношение, определяемое правилом сложения дисперсий. Согласно этому правилу, *общая дисперсия равна сумме средней из внутригрупповых и межгрупповой дисперсий*:

$$\sigma_0^2 = \overline{\sigma^2} + \delta^2 \qquad (5.18)$$

Рассмотрим правило сложения дисперсий на следующем примере. По результатам маркетингового обследования туристических фирм, организующих недельные туры в Испанию в различные курортные города, получены следующие данные о вариации стоимости туров в сентябре 1995 г. (см. табл. 5.6).

Таблица 5.6

Группировка туристических фирм по сегментам рынка в Испании в сентябре 1995 г.

Местоположение курорта	Число туристических фирм n_j	Средняя цена недельного тура, долл. $\bar{x}_j$	Дисперсия цен тура в группе σ_j^2
Коста-Брава	7	528,57	2728,04
Коста-дель-Соль	6	588,33	8851,14
Итого	13	556,16	5554,09

Вариация цен в обследованной группе туристических фирм, обусловленная различием в местоположении курорта, будет характеризоваться величиной межгрупповой дисперсии.

Средняя цена недельного тура по всем фирмам составляет 556,16 долл.: $\bar{x}_o = \frac{528{,}57 \cdot 7 + 588{,}33 \cdot 6}{13} = 556{,}16$ долл.

Тогда межгрупповая дисперсия будет равна 412,05:

$$\delta^2 = \frac{(528{,}57-556{,}16)^2 \cdot 7 + (558{,}33-556{,}16)^2 \cdot 6}{13} = 412{,}05.$$

Вариация цен под влиянием всех прочих факторов, кроме местоположения курорта, будет характеризоваться величиной средней из внутригрупповых дисперсий:

$$\overline{\sigma^2} = \frac{2728{,}04 \cdot 7 + 8851{,}14 \cdot 6}{13} = 5554{,}09.$$

Вариация цен на недельные туры в Испанию, обусловленная влиянием всех факторов, формирующих уровень цен в данной группе, состоящей из 13 туристических фирм, определяется величиной общей дисперсии:

$$\sigma_o^2 = 412{,}05 + 5554{,}09 = 5966{,}14.$$

Отсюда можно сделать вывод, что на 6,91% ($\frac{412{,}05}{5966{,}14} \cdot 100\%$) дисперсия цен на недельные туры объясняется различиями в местоположении курорта, а на 93,09% ($\frac{5554{,}09}{5966{,}14} \cdot 100\%$) - влиянием прочих факторов. Таким образом, преобладающие влияние на вариацию цен недельных туров в Испанию оказывают прочие факторы.

Выше был рассмотрен расчет показателей вариации для количественных признаков. Но наряду с вариацией количественных признаков может ставиться задача оценки вариации качественных признаков. При наличии двух взаимоисключающих вариантов значений признака говорят о наличии альтернативной изменчивости качественных признаков. Например, при изучении качества изготовленной продукции можно разделить ее на две группы: годную и бракованную. В таком случае будем иметь дело с альтернативным признаком. Можно считать, что эквивалентом качественного признака будет переменная, которая принимает значение 1 или 0, причем значение 1 она принимает в том случае, когда обследуемая единица обладает данным признаком, а значение 0, когда не обладает им.

Допустим, общее число единиц совокупности равно n. Число единиц, обладающих данным признаком - f, тогда число единиц, не обладающих данным признаком, будет равно $n-f$. Учитывая изложенное, построим ряд распределения по качественному признаку:

Значение переменной	Частота повторений
1 0	f n-f
Итого	n

Средняя арифметическая такого ряда равна:

$$\bar{x} = \frac{1 \cdot f + 0(n-f)}{n} = \frac{f}{n},$$

т.е. равна относительной частоте (частости), которую можно обозначить через p, тогда $\bar{x} = p$.

Таким образом, доля единиц, обладающих данным признаком, равна p; соответственно доля единиц, не обладающих данным признаком, равна q; $p+q = 1$. Тогда дисперсия альтернативного признака определяется по формуле

$$\sigma^2 = \frac{(1-p)^2 p + (0-p)^2 q}{p+q} = \frac{q^2 p + p^2 q}{p+q} = pq. \tag{5.19}$$

Среднее квадратическое отклонение альтернативного признака:

$$\sigma = \sqrt{pq}. \tag{5.19а}$$

Например, в результате контроля качества при приемке из 1000 готовых изделий 20 оказались бракованными. Применяя вышеуказанную символику, строим ряд распределения (1-й – соответствует бракованным изделиям, а 0 – годной продукции).

Значение переменной	Частота повторений
1 0	20 980
Итого	1000

Доля брака по данным примера составляет 2% $(\frac{20}{1000}\cdot 100\%)$. Тогда величина дисперсии составит 0,0196 (σ_p^2=0,02 • 0,98), а среднее квадратическое отклонение σ_p= 0,14.

Анализ вариации в рядах распределения целесообразно дополнить **показателями дифференциации**. По первичным данным может быть рассчитан так называемый *коэффициент фондовой дифференциации* $K_ф$, который рассчитывают соотношением двух средних, полученных из 10% наибольших и наименьших значений признака. Так, по данным о размере прибыли по 20 коммерческим банкам (см. с. 112) находим, что два банка (10% от общего числа) с наименьшим размером прибыли 3,7 и 4,3 млрд руб. имеют $\bar{x}_{наим.}$ = 4,0 млрд руб. Для двух банков с наибольшими значениями 7,9 и 8,1 млрд руб. средняя арифметическая $\bar{x}_{наиб.}$ составит 8,0 млрд руб. Таким образом, коэффициент фондовой дифференциации будет равен 2,0: $K_ф = \frac{\bar{x}_{наиб.}}{\bar{x}_{наим.}} = \frac{8,0}{4,0} = 2,0$.

Это означает, что размер прибыли у 10% банков с наивысшими доходами в два раза превышает размер прибыли 10% коммерческих банков с наименьшими доходами[1].

Если представлены сгруппированные данные, то для характеристики дифференциации можно воспользоваться соотношением десятой и первой децили (децили делят все число единиц в совокупности на 10 равных частей). Для определения децилей используются формулы, аналогичные тем, что приведены выше для расчета квартилей. Общая схема их расчета такова:

1) определяется номер децили N_{D_i} для первой децили $N_{D_1} = \frac{n+1}{10}$; для девятой $N_{D_9} = \frac{9(n+1)}{10}$, где D_i – i-тая дециль;

[1] Этот пример имеет чисто методический характер, так как рассчитывать коэффициенты дифференциации при малом объеме совокупности нецелесообразно.

2) устанавливается интервал, где должны будут находиться децили;

3) рассчитывается значение децилей при предположении равномерного наращения величины интервала на каждую единицу частоты (частости);

4) определяется *коэффициент децильной дифференциации*

$$K_D = \frac{D_9}{D_1}.$$

Покажем расчет этого коэффициента по данным о распределении населения России по размеру среднедушевого денежного дохода в I кв. 1995 г. (см. табл. 5.7).

Таблица 5.7

Среднедушевой доход в месяц, тыс.руб.	Человек, млн f_i	В процентах к итогу	Накопленные частоты S_i
20,1 - 40	0,2	0,1	0,2
40,1 - 60	1,0	0,7	1,2
60,1 - 120	12,8	8,6	14,0
120,1 - 180	23,0	15,5	37,0
180,1 - 240	24,0	16,2	61,0
240,1 - 300	20,7	13,9	81,7
300,1 - 360	16,3	11,0	98,0
360,1 - 420	12,4	8,4	110,4
420,1 - 480	9,3	6,3	119,7
480,1 - 540	6,9	4,7	126,6
540,1 - 600	5,1	3,5	131,7
600,1 - 700	5,9	4,0	137,6
700,1 - 800	3,6	2,4	141,2
800,1 - 900	2,3	1,6	143,5
900,1 - 1000	1,5	1,0	145,0
свыше 1000	3,2	2,1	148,2
Итого	148,2	100,0	

Источник. Статистическое обозрение. 1995. № 4. С. 58.

$$N_{D_1} = \frac{148,2}{10} = 14,8; \quad N_{D_9} = \frac{9\cdot 148,2}{10} = 133,4.$$

Тогда D_1 будет находиться в интервале 120,1 - 180 тыс. руб., а D_9 - в интервале 600,1 - 700 тыс. руб.

Величины D_1 и D_9 определятся так:

$$D_1 = 120{,}1 + 60\,\frac{14{,}8 - 14{,}0}{23} = 122{,}2 \text{ тыс. руб.},$$

$$D_9 = 600{,}1 + 100\,\frac{133{,}4 - 131{,}7}{5{,}9} = 628{,}9 \text{ тыс. руб.}$$

$K_D = \frac{628{,}9}{122{,}2} = 5{,}15$, т.е. наименьший уровень среднедушевого денежного дохода 10% наиболее обеспеченного населения и наивысший уровень среднедушевого денежного дохода 10% наименее обеспеченного населения в I кв. 1995 г. отличались в 5,15 раза.

5.5. Моменты распределения

Моментом распределения называется средняя арифметическая тех или иных степеней отклонений индивидуальных значений признака от определенной исходной величины. Рассчитывается по формуле:

$$\frac{\sum_{i=1}^{k}(x_i - A)^{\alpha} f_i}{\sum_{i=1}^{k} f_i} \qquad (5.20)$$

где A - величина, от которой определяются отклонения, α - степень отклонения (порядок момента).

В зависимости от того, что принимается за величину A, различают три вида моментов:

начальные моменты M_α получают при $A = 0$:

$$M_\alpha = \frac{\sum_{i=1}^{k} x_i^{\alpha} f_i}{\sum_{i=1}^{k} f_i} \qquad (5.20а)$$

центральные моменты μ_α получают при $A = \bar{x}$:

$$\mu_\alpha = \frac{\sum_{i=1}^{k}(x_i - \bar{x})^{\alpha} f_i}{\sum_{i=1}^{k} f_i}, \qquad (5.20б)$$

условные моменты m_α получают при A, не равной средней арифметической и отличной от нуля:

$$m_\alpha = \frac{\sum_{i=1}^{k}(x_i - A)^{\alpha} f_i}{\sum_{i=1}^{k} f_i}. \qquad (5.20в)$$

В статистической практике пользуются моментами первого, второго, третьего и четвертого порядков. Представим моменты первых четырех порядков в табл. 5.8.

Рассматривая формулы моментов, видим, что начальный момент первого порядка представляет собой среднюю арифметическую и используется как показатель центра распределения. Центральный момент первого порядка (в соответствии с нулевым свойством средней арифметической) всегда равен нулю. Центральный момент второго порядка представляет собой дисперсию и служит основной мерой колеблемости признака. Центральный момент третьего порядка равен нулю в симметричном распределении и используется для определения показателя асимметрии. Центральный момент четвертого порядка применяется при вычислении показателя эксцесса. Начальные моменты второго, третьего и четвертого порядков так же, как и условные моменты, самостоятельного значения не имеют, а используются для упрощения вычислений центральных моментов. Например, используя начальные моменты первого и второго порядка, можно получить дисперсию по такой формуле:

$$\sigma^2 = \mu_2 = M_2 - M_1^{\,2}. \qquad (5.21)$$

Таблица 5.8

Моменты распределения порядка	Начальные	Центральные	Условные
Первого	$M_1 = \frac{\sum x_i f_i}{\sum f_i}$	$\mu_1 = \frac{\sum (x_i - \bar{x}) f_i}{\sum f_i}$	$m_1 = \frac{\sum (x_i - A) f_i}{\sum f_i}$
Второго	$M_2 = \frac{\sum x_i^2 f_i}{\sum f_i}$	$\mu_2 = \frac{\sum (x_i - \bar{x})^2 f_i}{\sum f_i}$	$m_2 = \frac{\sum (x_i - A)^2 f_i}{\sum f_i}$
Третьего	$M_3 = \frac{\sum x_i^3 f_i}{\sum f_i}$	$\mu_3 = \frac{\sum (x_i - \bar{x})^3 f_i}{\sum f_i}$	$m_3 = \frac{\sum (x_i - A)^3 f_i}{\sum f_i}$
Четвертого	$M_4 = \frac{\sum x_i^4 f_i}{\sum f_i}$	$\mu_4 = \frac{\sum (x_i - \bar{x})^4 f_i}{\sum f_i}$	$m_4 = \frac{\sum (x_i - A)^4 f_i}{\sum f_i}$

Для упрощения вычислений центральных моментов распределения сначала определяют условные моменты относительно величины A, а для нахождения центральных моментов используют формулу перехода от условных моментов к центральным:

$$\mu_\alpha = \mu_\alpha - c_\alpha^1 m_{\alpha-1} \cdot m_1 + c_\alpha^2 m_{\alpha-2} \cdot m_1^2 - c_\alpha^3 m_{\alpha-3} \cdot m_1^3 + ... + (-m_1)^\alpha,$$

где $c_\alpha^1, c_\alpha^2, c_\alpha^3, ...$ - числа сочетаний из α по 1, по 2, по 3 и т.д.

5.6. Изучение формы распределения

Для получения приблизительного представления о форме распределения строят графики распределения (полигон и гистограмму). Число наблюдений, по которому строится *эмпирическое распределение*, обычно невелико и представляет собой выборку из исследуемой генеральной совокупности. Поэтому эмпирические данные в определенной степени связаны со случайными ошибками наблюдения, величина которых неизвестна. Влияние этих случайностей затемняет основную закономерность изменения величины признака. С увеличением числа наблюдений и одновременным уменьшением величины интервала зигзаги полигона начи-

нают сглаживаться, и в пределе мы приходим к плавной кривой, которая называется **кривой распределения**.

Кривая распределения характеризует *теоретическое распределение*, т.е. то распределение, которое получилось бы при полном погашении всех случайных причин, затемняющих основную закономерность. Исследование закономерности (или формы) распределения включает решение трех последовательных задач: 1) выяснение общего характера распределения; 2) выравнивание эмпирического распределения, которое состоит в том, что на основании эмпирического распределения строится кривая $y = f(x)$ с заданной формой; 3) проверку соответствия найденного теоретического распределения эмпирическому.

В практике статистического исследования приходится встречаться с самыми различными распределениями. Однородные совокупности характеризуются, как правило, одновершинными распределениями. Многовершинность свидетельствует о неоднородности изучаемой совокупности. Появление двух и более вершин говорит о необходимости перегруппировки данных с целью выделения более однородных групп. Выяснение общего характера распределения предполагает оценку степени его однородности, а также вычисление показателей асимметрии и эксцесса. *Симметричным является распределение, в котором частоты любых двух вариантов, равноотстоящих в обе стороны от центра распределения, равны между собой.* Для симметричных распределений имеет место равенство средней арифметической, моды и медианы. В связи с этим *простейший показатель асимметрии основан на соотношении показателей центра распределения: чем больше разница между средними ($\bar{x} - M_0$), тем больше асимметрия ряда.*

Для сравнительного анализа степени асимметрии нескольких распределений рассчитывают относительный показатель A_s:

$$As = \frac{\bar{x} - Mo}{\sigma} \qquad (5.22)$$

Величина показателя асимметрии As может быть положительной и отрицательной. Положительная величина показателя асимметрии указывает на наличие правосторонней асимметрии (правая ветвь относительно максимальной ординаты вытянута больше, чем левая, рис. 5.6). При правосторонней асимметрии между показателями центра распределения существует соотношение: $Mo < Me < \bar{x}$. Отрицательный знак показателя асимметрии свидетельствует о наличии левосторонней асимметрии (рис. 5.6). Между показателями центра распределения в этом случае имеется такое соотношение: $Mo > Me > \bar{x}$.

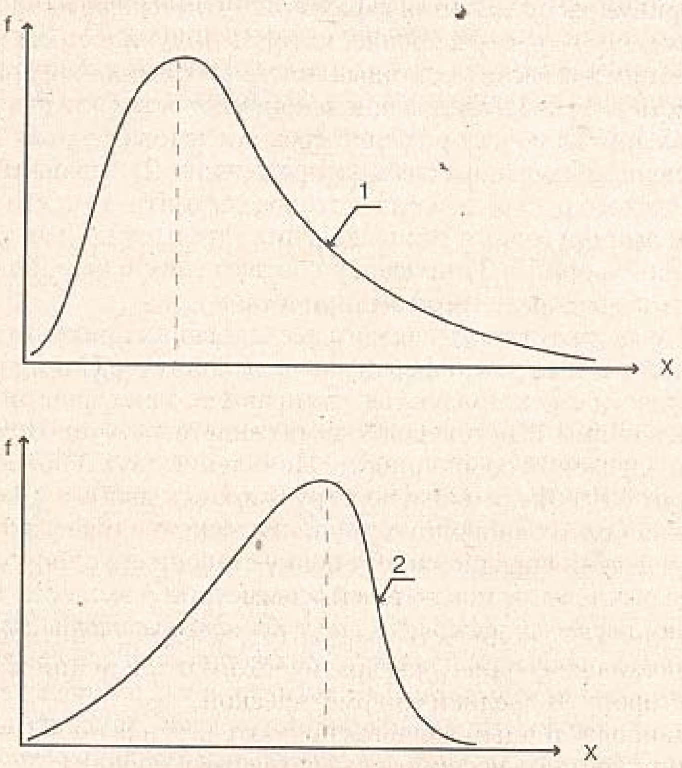

Рис. 5.6. **Асимметричные ряды распределения:**
1 - с правосторонней асимметрией;
2 - с левосторонней асимметрией.

Другой показатель асимметрии, предложенный шведским математиком Линдбергом, рассчитывают по формуле:

$$As = П - 50, \qquad (5.23)$$

где $П$ - 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):

$$As = \frac{\mu_3}{\sigma^3} \qquad (5.24)$$

Применение этого показателя дает возможность не только определить степень асимметрии, но и ответить на вопрос о наличии 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 и рассчитывается по формуле:

$$\sigma_{As} = \sqrt{\frac{6(n-1)}{(n+1)(n+3)}} \qquad (5.24а)$$

Если отношение $\frac{|As|}{\sigma_{As}} > 3$, асимметрия существенна и распределение признака в генеральной совокупности не является симметричным. Если отношение $\frac{|As|}{\sigma_{As}} < 3$, асимметрия несущественна, ее наличие может быть объяснено влиянием различных случайных обстоятельств.

Для симметричных распределений рассчитывается показатель **эксцесса** (островершинности). Линдбергом предложен следующий показатель для оценки эксцесса:

$$Ex = П - 38{,}29, \qquad (5.25)$$

где $П$ - доля (%) количества вариантов, лежащих в интервале, равном половине среднего квадратического отклонения в ту и другую сторону от средней арифметической.

Наиболее точным является показатель основанный на использовании центрального момента четвертого порядка:

$$Ex = \frac{\mu_4}{\sigma^4} - 3 \qquad (5.26)$$

Рис. 5.7. **Ряды распределения с положительным (I) и отрицательным (II) эксцессом.**

На рис. 5.7 представлены два распределения: одно - островершинное (величина эксцесса положительная), второе - плосковершинное (величина эксцесса отрицательная). *Эксцесс представляет собой выпад вершины эмпирического распределения вверх или вниз от вершины кривой нормального распределения.* В нормальном распределении отношение $\frac{\mu_4}{\sigma^4} = 3$.

Средняя квадратическая ошибка эксцесса рассчитывается по формуле:

$$\sigma_{Ex} = \sqrt{\frac{24n(n-2)(n-3)}{(n-1)^2(n+3)(n+5)}}, \qquad (5.27)$$

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

$$f(x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(x-\bar{x})^2}{2\sigma^2}}, \qquad (5.28)$$

то она подчиняется закону нормального распределения. Для построения кривой нормального распределения надо знать два параметра: $\bar{x}$ и σ.

Если средняя арифметическая не меняется, но растет величина среднего квардатического отклонения, распределение имеет более плосковершинный характер (см. рис. 5.8).

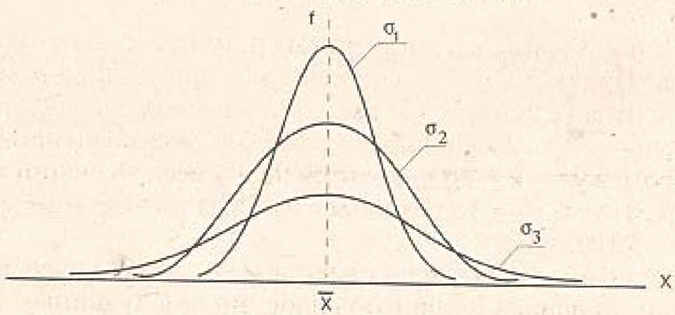

Рис. 5.8. **Кривые нормального распределения с одинаковой $\bar{x}$, но разными σ ($\sigma_1 < \sigma_2 < \sigma_3$)**

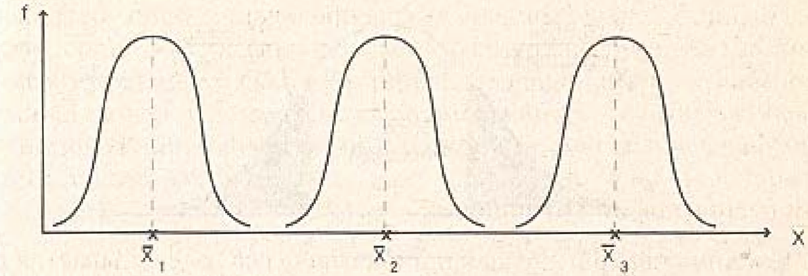

Рис. 5.9. **Кривые нормального распределения с одинаковым σ, но разными $\bar{x}$ ($\bar{x}_1 < \bar{x}_2 < \bar{x}_3$)**

На рис. 5.9. приведено «семейство» кривых нормального распределения с одной и той же величиной среднего квадратического отклонения, но с разными средними ($\bar{x}_1 < \bar{x}_2 < \bar{x}_3$). В этом случае кривая, не меняя своей формы, сдвигается вправо вдоль оси абсцисс.

Укажем особенности кривой нормального распределения.

1. Кривая симметрична относительно максимальной ординаты. Максимальная ордината соответствует значению $x = Mo = Me$, ее величина равна $\frac{1}{\sqrt{2\pi}\cdot\sigma}$.

2. Кривая асимптотически приближается к оси абсцисс, продолжаясь в обе стороны до бесконечности. Следовательно, чем больше значения отклоняются от $\bar{x}$, тем реже они встречаются. Одинаковые по абсолютному значению, но противоположные по знаку отклонения значений переменной x от $\bar{x}$ равновероятны.

3.Кривая имеет две точки перегиба,находящиеся на расстоянии $\pm\sigma$ от $\bar{x}$.

4. 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.

5. 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 (см. рис. 5.10).

Нормальное распределение возможно в том случае, когда на величину признака влияет большое число случайных причин. Действие этих причин независимо, и ни одна из причин не имеет преобладающего влияния над другими.

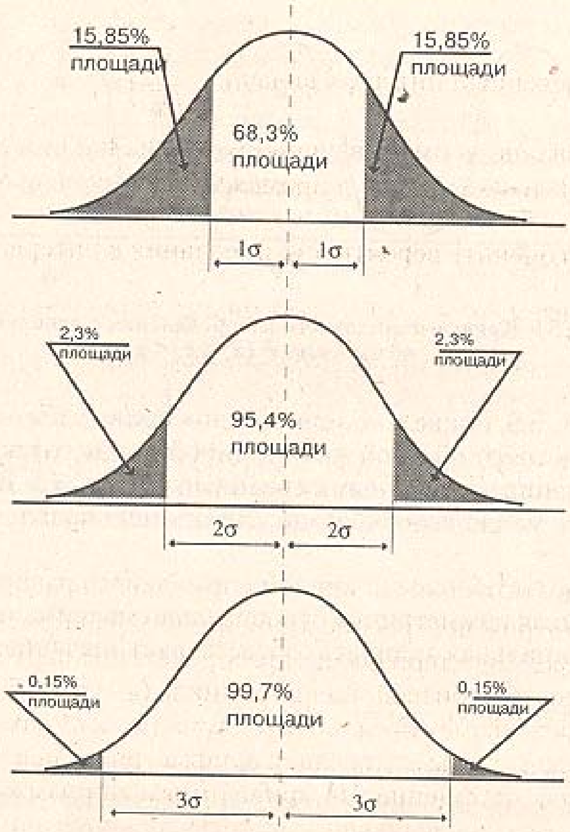

Рис. 5.10. **Соотношение площади под кривой нормального распределения в зависимости от расстояния от средней арифметической**

Для удобства вычислений вероятностей случайные величины нормируются, а затем используются заранее табулированные значения плотности функции распределения нормированной случайной величины. Если обозначим $\frac{x-\bar{x}}{\sigma}$ через t, то величину

$\frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}$ назовем нормированной функцией, эта функция табулирована. Для нормированной случайной величины математическое ожидание равно нулю, а дисперсия равна единице (см. Приложение 1).

Определенный интеграл вида $F(t) = \frac{1}{\sqrt{2p}} \int_0^t e^{-\frac{t^2}{2}} dt$

носит название нормированной функции Лапласа и характеризует площадь под кривой в промежутке от нуля до t (см. Приложение II).

Чтобы оценить вероятность попадания в интервал от $-\infty$ до x, рассчитываем $F(x) = \frac{1}{2} + F(t)$. Для определения вероятности попадания нормально распределенной случайной величины x в заданный интервал $(x_1; x_2)$ находим разность $F(x_2) - F(x_1)$, т.е.

$$P(x_1 \le x \le x_2) = F(x_2) - F(x_1) = \left[\frac{1}{2} + F(t_2)\right] - \left[\frac{1}{2} + F(t_1)\right] = F(t_2) - F(t_1),$$

где $t_1 = \frac{x_1 - \bar{x}}{\sigma}$; $t_2 = \frac{x_2 - \bar{x}}{\sigma}$.

Рассмотрим построение кривой нормального распределения на примере, характеризующем распределение партий деталей по длительности производственного цикла (см. табл. 5.9). Нормальное распределение определяется двумя параметрами, оценки которых нужно знать: это средняя арифметическая и среднее квадратическое отклонение. По приводимым данным $\bar{x} = 331$ ч., $\sigma =$ 157,25 ч. Все последующие расчеты для определения теоретических частот представлены в гр. 3-8 табл. 5.9. (Условные обозначения в табл.: $x_i^н$ и $x_i^в$ - нижние и верхние границы интервалов; $t_i^н$ и $t_i^в$ - нормированные отклонения для нижней и верхней границ интервала; $F(t_i^н)$ и $F(t_i^в)$ - значения интегральной функции Лапласа для $t_i^н$ и $t_i^в$; P_i - оценка вероятности попадания в интервал; f_i' - частота теоретического распределения).

Значения $F(t_i^н)$ и $F(t_i^в)$ (гр. 5 и 6) определяются по таблицам интегральной функции Лапласа (см. Приложение II). Оценка вероятности попадания случайной величины в интервал P_i (гр. 7) определяется разностью $[F(t_i^в)$ и $F(t_i^н)]$. Теоретическая частота $f_i' = p_i \cdot n$. Например, для первого интервала $(-\infty \div 28)$ $f_i' = 0{,}0268 \cdot 71 =$

$=1{,}9$ и т.д. (см. гр. 8), а $\sum_{i=1}^{k} f_i = \sum_{i=1}^{k} f_i'$.

Таблица 5.9

Границы интервала, час	Наблюдаемая частота, f_i	$t_i^{н} = \frac{x_i^{н} - \bar{x}}{\sigma}$	$t_i^{в} = \frac{x_i^{в} - \bar{x}}{\sigma}$	$F(t_i^{н})$	$F(t_i^{в})$	P_i	f_i'
1	2	3	4	5	6	7	8
-∞ - 28	0	-∞	-1,927	-0,5000	-0,4732	0,0268	1,9
28 - 113	5	-1,927	-1,393	-0,4732	-0,4177	0,0555	3,94
113 - 198	12	-1,393	-0,852	-0,4177	-0,3023	0,1154	8,19
198 - 283	12	-0,852	-0,312	-0,3023	-0,1217	0,1806	12,82
283 - 368	15	-0,312	+0,229	-0,1217	-0,0910	0,2127	15,11
368 - 453	9	+0,229	+0,769	+0,0910	+0,2791	0,1884	13,40
453 - 538	9	+0,769	+1,31	+0,2791	+0,4049	0,1258	8,93
538 - 623	7	+1,31	+1,86	+0,4049	+0,4686	0,0637	4,52
623 - 708	2	+1,86	+2,39	+0,4686	+0,4915	0,0229	1,63
708 - +∞	0	+2,39	+∞	+0,4915	+0,5000	0,0085	0,59
Итого	71						71

При рассмотрении маловероятных событий, имеющих место в большой серии независимых испытаний некоторое (конечное) число раз, вероятности появления этих событий подчиняются закону Пуассона, или закону редких событий $P_m = \frac{\lambda^m e^{-\lambda}}{m!}$, где λ равна среднему числу появлений событий A в n одинаковых независимых испытаниях, т.е. $\lambda = np$, где p - вероятность события при одном испытании; $e = 2{,}71828$; m - частота данного события. Математическое ожидание m равно λ.

Закон Пуассона можно применять для совокупностей, достаточно больших по объему ($n \geqslant 100$) и имеющих достаточно малую долю единиц, обладающих данным признаком ($p \leqslant 0{,}1$). Например, количество бракованных деталей; количество отказов автоматических линий и т.д. (см. табл. 5.10).

Таблица 5.10

Количество бракованных деталей	Наблюдаемая частота	Частота теоретического распределения
0	604	606
1	306	303
2	77	76
3	12	13
4	1	2
Итого	1000	1000

Сопоставление наблюдаемых и теоретических частот свидетельствует о достаточном соответствии эмпирического распределения распределению Пуассона. Степень расхождения теоретических и эмпирических частот оценивается с помощью особых показателей - критериев согласия, с помощью которых проверяется гипотеза о законе распределения.

5.7. Критерии согласия

Критерии согласия основаны на использовании различных мер расстояний между анализируемым эмпирическим распределением и функцией распределения признака в генеральной совокупности.

Одним из наиболее часто употребляемых критериев согласия является *критерий «хи-квадрат» (χ^2)*, предложенный К. Пирсоном,

$$\chi^2 = \sum_{i=1}^{k} \frac{(f_i - f_i')^2}{f_i'} . \tag{5.29}$$

где f_i и f_i' - соответственно частоты эмпирического и теоретического распределений в i-том интервале.

Чем больше разность между наблюдаемыми и теоретическими частотами, тем больше величина критерия Пирсона. Чтобы отличить существенные значения χ^2 от значений, которые могут возникнуть в результате случайностей выборки, рассчитанное значение критерия сравнивается с табличным значением $\chi^2_{табл}$ при соответствующем числе степеней свободы и заданном уровне значимости. Уровень значимости выбираем таким образом, что $P(\chi^2_{расч.} > \chi^2_{табл.}) = \alpha$ (величина α принимается равный 0,05 или 0,01).

Определив значение критерия Пирсона по данным конкретной выборки, можно встретиться с такими вариантами:

1) $\chi^2_{расч.} > \chi^2_{табл.}$, т.е. χ^2 попадает в критическую область. Это означает, что расхождение между эмпирическими и теоретическими частотами существенно и его нельзя объяснить случайными колебаниями выборочных данных. В таком случае гипотеза о близости эмпирического распределения к нормальному отвергается.

2) $\chi^2_{расч.} \leqslant \chi^2_{табл.}$, т.е. рассчитанный критерий не превышает максимально возможную величину расхождений эмпирических и теоретических частот, которая может возникнуть в силу случайных колебаний выборочных данных. В этом случае гипотеза о близости эмпирического распределения к нормальному не отвергается.

Как уже указывалось, табличное значение критерия Пирсона определяется при фиксированном уровне значимости и соответствующем числе степеней свободы.

Число степеней свободы равно $k - l - 1$, где l - число условий, которые предполагаются выполненными при вычислении теоретических частот, k - число групп.

Так как при вычислении теоретических частот нормального распределения в качестве оценок генеральной средней и дисперсии используются соответствующие выборочные характеристики, то для проверки гипотезы о нормальности распределения число степеней свободы равно $(k - 3)$. Расчет теоретических частот нормального распределения рассмотрен в параграфе 5.6.

При расчете критерия Пирсона нужно соблюдать следующие условия: 1) число наблюдений должно быть достаточно велико, во всяком случае $n \geqslant 50$; 2) если теоретические частоты в некоторых интервалах меньше 5, то такие интервалы объединяют так, чтобы частоты были более 5.

Воспользуемся данными примера, приведенного в табл. 5.9 для расчета критерия «хи-квадрат», предварительно округлив теоретические частоты в графе 8 табл. 5.9, а также объединив частоты первых двух и трех последних интервалов, выполняя требование $f_i' \geqslant 5$. После таких предварительных расчетов мы получим частоты эмпирического и теоретического распределений, приведенные в табл. 5.11.

Таблица 5.11

Номер интервала	Эмпирические частоты	Теоретические частоты	$(f_i - f_i')^2$	$\frac{(f_i - f_i')^2}{f_i'}$
1	5	6	1	0,17
2	12	8	16	2,00
3	12	13	1	0,08
4	15	15	0	0,00
5	9	13	16	1,23
6	9	9	0	0,00
7	9	7	4	0,57
Итого	71	71		4,05

$$\chi^2_{\text{расч.}} = 4,05$$

Число степеней свободы равно: 7 - 3 = 4.

При уровне значимости 0,05 и числе степеней свободы 4 $\chi^2_{\text{табл.}} = 9,5$ (см. Приложение V).

Таким образом, расчетное значение критерия Пирсона не превышает табличное значение при $\alpha = 0,05$, т.е. проведенный расчет дает право не отвергать гипотезу о нормальном характере эмпирического распределения.

Используя величину χ^2, В.И. Романовский предложил оценку близости эмпирического распределения кривой нормального распределения производить по отношению:

$$\frac{\chi^2 - (k - 3)}{\sqrt{2(k - 3)}}, \qquad (5.30)$$

где k - число групп, а величина $(k-3)$ равна числу степеней свободы при исчислении частот нормального распределения.

Если отношение $\frac{\chi^2 - (k - 3)}{\sqrt{2(k - 3)}} > 3$, то расхождение частот эмпирического распределения и рассчитанных частот нормального распределения нельзя признать случайным, и гипотезу о нормальном законе распределения следует отвергнуть.

Если же отношение $\frac{\chi^2 - (k - 3)}{\sqrt{2(k - 3)}} < 3$, то возможно принять гипотезу о нормальном характере эмпирического распределения.

Продолжая приведенный выше пример, рассчитаем *критерий В.И. Романовского*:

$$\frac{\chi^2 - (k - 3)}{\sqrt{2(k - 3)}} = \frac{4,05 - 4}{\sqrt{8}} = \frac{0,05}{2,83} = 0,0177$$

Так как рассчитанное отношение значительно меньше трех, следует принять гипотезу о нормальности эмпирического распределения.

Следующий весьма распространенный критерий согласия - *критерий А.Н. Колмогорова*. Для определения соответствия между теоретическим и эмпирическим распределениями академик А.Н. Колмогоров предложил рассматривать максимальную разность между значениями интегральной функции теоретического распределения $F(x)$ и значениями интегральной функции эмпирического распределения $F'(x)$.

Обозначим символом d_n абсолютную величину максимальной разности $F(x)$ и $F'(x)$, т.е. $d_n = \max |F(x) - F'(x)|$.

А.Н. Колмогоров установил, что когда n неограниченно возрастает, вероятность того, что d_n будет меньше величины $\frac{\lambda}{\sqrt{n}}$, приближается к значениям функции.

$$K(\lambda) = \sum_{-\infty}^{+\infty} (-1)^K e^{-2k^2\lambda^2}, \text{ т.е. } \lim P(d_n \le \frac{\lambda}{\sqrt{n}}) = K(\lambda) \qquad (5.31)$$

По таблицам вероятностей $K(\lambda)$ можно найти величину λ, соответствующую данной величине вероятности $K(\lambda)$. Если известен объем выборки, можно рассчитать соотношение $\frac{\lambda}{\sqrt{n}}$. Величина d_n, рассчитанная по данным выборки, сравнивается с величиной $\frac{\lambda}{\sqrt{n}}$.

Если $d_n < \frac{\lambda}{\sqrt{n}}$, то с вероятностью $K(\lambda)$ можно считать, что рассматриваемое распределение следует закону нормального распределения.

Если же $d_n > \frac{\lambda}{\sqrt{n}}$, то вероятность того, что гипотеза о соответствии эмпирического и теоретического распределения верна, определяется величиной $1 - K(\lambda)$.

Так как величину $K(\lambda)$ выбирают равной 0,95 или 0,99, то $P(d_n > \frac{\lambda}{\sqrt{n}}) \le 0{,}05$. В силу малой вероятности такого события гипотеза о нормальном характере эмпирического распределения отвергается.

Приведем краткую выдержку из таблицы значений функции $K(\lambda)$ А.Н. Колмогорова:

λ	1,23	1,36	1,63	1,80	2,00
$K(\lambda)$	0,9030	0,9505	0,9902	0,9970	0,9993

Покажем расчет критерия А.Н. Колмогорова на том же примере, результаты расчетов приведены в табл. 5.12.

Таблица 5.12

Номер интервала	Частоты		Частоты абсолютные накопленные		Частоты относительные накопленные			
	эмпирические f_i	теоретические f'_i	эмпирические Σf_i	теоретические $\Sigma f'_i$	эмпирические $F(x)$	теоретические $F'(x)$	$\lvert F(x) - F'(x)\rvert$	$\lvert\Sigma f_i - \Sigma f'_i\rvert$
1	2	3	4	5	6	7	8	9
1	5	6	5	6	0,0704	0,0845	0,0141	1
2	12	8	17	14	0,2394	0,1972	0,0422	3
3	12	13	29	27	0,4085	0,3803	0,0282	2
4	15	15	44	42	0,6197	0,5915	0,0282	2
5	9	13	53	55	0,7465	0,7746	0,0281	2
6	9	9	62	64	0,8732	0,9014	0,0282	2
7	9	7	71	71	1,0000	1,0000	0,0000	0

Наибольшая по абсолютной величине разность накопленных относительных частот $d_n = 0{,}0422$ соответствует второму интервалу (гр. 8).

Принимая уровень значимости $\alpha = 0{,}05$, т.е. при доверительной вероятности $K(\lambda) = 0{,}95$, по таблицам значений функций А.Н. Колмогорова находим $\lambda = 1{,}36$, и, следовательно, для $n = 71$ величина $\frac{\lambda}{\sqrt{n}}$ равна 0,1614. Вероятность того, что $d_n > \frac{\lambda}{\sqrt{n}}$ была установлена нами равной 0,05 (λ=0,05) . Такой результат является практически невозможным, когда проверяемая гипотеза о соответствии эмпирического распределения нормальному закону верна. Поскольку величина d_n, рассчитанная по данным приводимого примера, получилась меньше величины $\frac{\lambda}{\sqrt{n}}$, с вероятностью 95% принимается гипотеза о нормальности эмпирического распределения.

Для расчета функции $K(\lambda)$ (критерия А.Н. Колмогорова) можно использовать и максимальную разность абсолютных накопленных частот, т.е. величину $D_n = \max|\Sigma f_i - \Sigma f_i'|$. Тогда между D_n и λ будет соотношение: $D_n = \lambda\sqrt{n}$, т.е. $\lambda = \frac{D_n}{\sqrt{n}}$. Максимальная величина разности абсолютных накопленных частот равна 3 (гр. 9 табл. 5.12). Величина λ, соответствующая $K(\lambda) = 0{,}9505$, равна 1,36. Следовательно, для $n = 71$ величина $\lambda\sqrt{n}$ равна 11,4594. Так как значение D_n=3 меньше $\lambda\sqrt{n} = 11{,}4594$, то можно считать, что результаты выборки не противоречат предположению о нормальном характере распределения исследуемого признака.

5.8. Вариационный ряд и группировка

Знакомство со свойствами и назначением основных показателей, характеризующих вариационные ряды, а также с особенностями кривой нормального распределения позволяет более подробно остановиться на принципах группировки исходных данных. Как уже указывалось, в статистических исследованиях должно обеспечиваться единство анализа качественной и количественной сторон изучаемых явлений. Одно из важнейших требований теории статистики к научно обоснованному анализу - соблюдение

принципа однородности исследуемых единиц. *Необходимое условие однородности - однотипность явлений*, устанавливаемая уже при первых шагах статистического исследования. Затем в рамках выполнимости необходимого условия проверяется и количественная однородность как достаточное условие. Обычно при построении вариационного ряда речь идет не о выделении типов, а ставится цель количественной характеристики вариации. Однако в процессе такого анализа может возникнуть проблема выделения в составе исследуемой совокупности достаточно различающихся между собой групп. При построении рядов распределений важное значение имеет решение вопроса о выборе *величины интервала*, что связано с характеристикой формы распределения и точностью вычисляемых оценок. Так, Р. Фишер подчеркивает значение выбора интервала для характеристики формы распределения: «В тех весьма частых случаях, когда переменная изменяется непрерывно, так что возможны все промежуточные значения, выбор размера интервалов и их границ произволен, а это может привести к заметным различиям во внешнем виде диаграмм»[1].

При расчете эмпирических характеристик по сгруппированным данным в расчетные формулы вместо фактических значений подставляются центральные значения интервалов. И здесь естественно поставить вопрос, в какой мере оценки средней и среднего квадратического отклонения, полученные по сгруппированным данным, близки к таким же оценкам, вычисленным без группировки. Вычисления будут точными, если средняя арифметическая всех значений, попавших в соответствующий интервал, совпадает с его центральным значением. Если отклонения средних арифметических интервалов от центральных значений не имеют систематического характера, то будет достигнута достаточная точность. Таким образом, ширина интервала должна быть такой, чтобы центральное значение интервала давало достаточно близкое представление о средней величине всех тех индивидуальных значений, которые входят в соответствующий интервал. Кроме того, весьма важно, чтобы число наблюдений в интервале не было бы слишком малым, а соответственно число групп чрезмерно большим. В общем виде можно сказать, что расчет величины интервала должен основываться на учете степени вариации признака и численности единиц исследуемой совокупности.

[1] *Фишер Р.* Статистические методы для исследователей. М., Статистика, 1958. С. 35.

В условиях машинной обработки информации речь должна идти о тех или иных стандартных процедурах группировки по количественным признакам. Один из вариантов такого стандарта приведен в параграфе 5.1, где для определения величины интервала предложена формула Стэрджесса.

В дополнение к уже рассмотренной процедуре при построении рядов распределения можно воспользоваться свойствами кривой нормального распределения. Учитывая, что в пределах $\bar{x} \pm 3\sigma$ должно находиться 99,7% всех вариантов значений признака, можно предложить следующие интервалы группировки:

1) до $\bar{x} - 3\sigma$; 2) $\bar{x} - 3\sigma \div \bar{x} - 2\sigma$; 3) $\bar{x} - 2\sigma \div \bar{x} - \sigma$; 4) $\bar{x} - \sigma \div \bar{x}$; 5) $\bar{x} \div \bar{x} + \sigma$; 6) $\bar{x} + \sigma \div \bar{x} + 2\sigma$; 7) $\bar{x} + 2\sigma \div \bar{x} + 3\sigma$; 8) $\bar{x} + 3\sigma$ и более.

Такой вариант группировки может использоваться в совокупностях достаточно большого объема.

Для асимметричных распределений может быть предложен вариант группировки с использованием средних величин. Общая средняя величина $\bar{x}_0$, рассчитанная по всей совокупности единиц, позволяет выделить две группы с уровнем признака ниже и выше средней. По этим группам рассчитываются средние $\bar{x}_1$ и $\bar{x}_2$, на основе которых в свою очередь можно аналогичным образом сформировать подгруппы. Таким образом, получим следующие интервалы группировки: 1) менее $\bar{x}_1$; 2) $\bar{x}_1 \div \bar{x}_0$; 3) $\bar{x}_0 \div \bar{x}_2$; 4) $\bar{x}_2$ и более.

В совокупностях, достаточно больших по объему, может быть продолжено деление сформированных групп на подгруппы и получено большее число групп.

Покажем рассмотренный вариант группировки на примере, приведенном в параграфе 5.1. По несгруппированным данным $\bar{x}_0 = 6{,}055$ млрд руб. Тогда в первую группу (ниже средней арифметической) попадут банки, имеющие размер прибыли менее 6,055 млрд руб., а именно 3,7; 4,3; 5,6; 5,1; 4,6; 5,7; 5,9; 5,2; 5,8; 4,9 - всего 10 банков. Остальные десять образуют группу с размером прибыли выше средней арифметической (в млрд руб.): 6,7; 8,1; 6,4; 6,2; 6,3; 7,2; 7,9; 7,6; 7,0; 6,9.

Вычислим средние арифметические по этим группам:

$\bar{x}_1 = 50{,}8/10 = 5{,}08$ млрд руб.; $\bar{x}_2 = 70{,}3/10 = 7{,}03$ млрд руб.

Учитывая, что $\bar{x}_{min} = 3{,}7$ млрд руб., а $\bar{x}_{max} = 8{,}1$ млрд руб., получим такую группировку (табл. 5.13):

Таблица 5.13

Группы банков по размеру прибыли, млрд руб.	Число банков f_i	Плотность распределения $\frac{f_i}{h_i}$	Центральные значения интервала x_i'	$x_i' \cdot \frac{f_i}{h_i}$
1	2	3	4	5
3,7 - 5,1 (+)	5	3,57	4,40	15,71
5,2 - 5,9	5	7,14	5,55	39,63
6,0 - 6,9	5	5,56	6,45	35,86
7,0 - 8,1	5	4,55	7,59	34,35
Итого	20	20,82		125,55

Примечание.

Знак (+), проставленный у верхней границы первого интервала, означает, что значения признака, совпадающие с верхней границей соответствующего интервала, включаются в этот интервал.

Поскольку представленный в табл. 5.13 ряд распределения имеет неравные интервалы, в качестве веса следует воспользоваться показателями плотности распределения (см. графу 3).

$$\bar{x} = \frac{\sum\limits_{i=1}^{k} x_i' \cdot \frac{f_i}{h_i}}{\sum\limits_{i=1}^{k} \frac{f_i}{h_i}} = \frac{125{,}55}{20{,}82} = 6{,}030 \text{ млрд руб.}$$

Расчет средней арифметической по приведенной в этой таблице группировке дает следующий результат: $\bar{x}_0 = 6{,}030$ млрд руб. Средняя арифметическая, рассчитанная по группировке, приведенной в табл. 5.5, равна 6,085 млрд руб.

Таким образом, величина абсолютной ошибки в расчете средней арифметической ($\Delta \bar{x}$) по различным группировкам составила: расчет по табл. 5.5: $\Delta \bar{x} = 6{,}085 - 6{,}055 = 0{,}030$ млрд руб., или (в относительном выражении) 0,5% ($\frac{0{,}030}{6{,}055} 100\%$). Расчет по табл. 5.13: $\Delta \bar{x} = 6{,}030 - 6{,}055 = -0{,}025$ млрд руб., или 0,4%.

Приведенные расчеты позволяют сделать вывод о достаточной точности результатов вычислений средних арифметических значений по обоим вариантам группировки.

В табл. 5.13 мы получили равномерное распределение общего числа наблюдений по группам - в каждую группу попало пять банков. Однако такое распределение единиц по группам совсем не обязательно при использовании приведенного выше алгоритма группировки. Принцип равной наполнённости групп по числу наблюдений может быть обеспечен использованием для определения границ интервалов квартилей (в каждую группу включают 25% общего числа единиц совокупности) или при большом числе наблюдений - децилей (в каждую группу включают 10% от общего числа наблюдений). Использование квартилей для группировки 20 коммерческих банков по размеру прибыли приводит к такому же результату, который уже получен в табл. 5.13.

Контрольные вопросы к главе 5

1. В чем состоят различия в построении рядов распределения с дискретным и непрерывным характером вариации признака?
2. Какие системы показателей используют для характеристики особенностей рядов распределения?
3. В чем состоят особенности расчета средней арифметической, моды и медианы в интервальных рядах распределения?
4. В каких случаях используется плотность распределения при расчете средней арифметической?
5. Что представляет собой вариация признака и в чем состоит значение ее изучения?
6. Какие показатели вариации находят наиболее широкое применение?
7. Что характеризует межгрупповая дисперсия?
8. Напишите формулу для расчета средней из внутригрупповых дисперсий.
9. Напишите соотношение между показателями центра распределения при правосторонней и левосторонней ассиметрии.
10. В чем состоит значение проверки гипотезы о форме распределения?
11. Каковы особенности кривых нормального распределения? Их использование в анализе фактических данных.
12. Какие критерии согласия используются наиболее часто?

Глава 6
ВЫБОРОЧНОЕ НАБЛЮДЕНИЕ

6.1. Понятие о выборочном наблюдении и его теоретические основы

Выборочное наблюдение представляет собой один из наиболее широко применяемых видов несплошного наблюдения. При проведении выборочного наблюдения, как и всякого несплошного наблюдения, обследуются не все единицы изучаемого объекта, или, иными словами, обследуются не все единицы генеральной совокупности, а лишь некоторая, так или иначе отобранная часть этих единиц. Однако наблюдение организовано таким образом, что эта часть отобранных единиц в уменьшенном масштабе репрезентирует (представляет) всю совокупность. Часть единиц генеральной совокупности, подлежащей непосредственному наблюдению, называют **выборочной совокупностью**.

Система правил отбора единиц и способов характеристики изучаемой совокупности исследуемых единиц составляет содержание выборочного метода. 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 метода. 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 г.г. 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 отечественные 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 Всероссийской переписи населения 1999 г. содержит 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%-ного выборочного обследования с целью изучения брачности и рождаемости.

Применение выборочного наблюдения взамен сплошного дает возможность лучше организовать наблюдение, обеспечивает быстроту проведения наблюдения, приводит к экономии средств и затрат труда на получение и обработку информации. Объективную гарантию репрезентативности полученной выборочной совокупности дает применение соответствующих научно обоснованных способов отбора подлежащих обследованию единиц. В процессе формирования выборочной совокупности должен быть обеспечен строго объективный подход к отбору единиц. Нарушение этого принципа, когда наблюдению подвергаются единицы, отобранные на основании субъективного мнения исследователя, приводит к тому, что результаты такого наблюдения относятся не ко всей генеральной совокупности, а только к той ее части, которая была подвергнута наблюдению.

В сравнении с другими видами несплошных наблюдений преимущество выборочного наблюдения заключается в том, что по результатам этого наблюдения можно оценить искомые параметры генеральной совокупности. Между характеристиками выборочной совокупности и искомыми характеристиками (параметрами) генеральной совокупности, как правило, существует некоторое расхождение, которое называют **ошибкой**. Общая величи-

на возможной ошибки выборочной характеристики слагается из ошибок двоякого рода: ошибки регистрации и ошибки репрезентативности.

Ошибки регистрации свойственны любому статистическому наблюдению вообще и появление их может быть вызвано несовершенством измерительных приборов, недостаточной квалификацией наблюдателя, неточностью подсчетов и т.п. Можно полагать, что по сравнению со сплошными наблюдениями опасность возникновения ошибок регистрации при проведении выборочных наблюдений должна быть меньше, так как выборочные наблюдения проводятся с участием более квалифицированных работников и, следовательно, более тщательно. Значительно уменьшается при выборочном наблюдении и опасность преднамеренных искажений данных, так как специально подобранные и обученные наблюдатели в них не заинтересованы.

Ошибки репрезентативности присущи только несплошным наблюдениям и представляют собой расхождение между величиной полученных по выборке показателей и величиной этих показателей, которые были бы получены при проведенном с одинаковой степенью точности сплошном наблюдении.

Ошибки репрезентативности могут быть систематическими и случайными. *Систематические ошибки* могут возникать в связи с особенностями принятой системы отбора и обработки данных наблюдений или в связи с нарушением установленных правил отбора. Возникновение *случайных ошибок* репрезентативности объясняется недостаточно равномерным представлением в выборочной совокупности различных категорий единиц генеральной совокупности, в силу чего распределение отобранной совокупности единиц не вполне точно воспроизводит распределение единиц генеральной совокупности.

Определение *возможной* и *фактически допущенной* ошибки выборки имеет важное значение при применении выборочного метода. Величина ошибки характеризует степень надежности результатов выборки; знание этой величины необходимо при оценке параметров генеральной совокупности. Оценки возможной величины и состава ошибок репрезентативности ложатся в основу планирования проектируемого выборочного наблюдения.

Величина случайной ошибки репрезентативности зависит:

1) от принятого способа формирования выборочной совокупности. Выбор последнего связан с решением вопросов о единице отбора, способе отбора единиц, способе размещения всего объе-

ма отбираемых единиц по различным группам генеральной совокупности;

2) от объема выборки;

3) от степени колеблемости изучаемого признака в генеральной совокупности.

Для каждого конкретного выборочного наблюдения величина ошибки репрезентативности может быть определена по соответствующим формулам.

По способу организации различают следующие виды выборочного наблюдения: собственно случайную или простую, расслоенную (типическую или районированную), серийную, механическую, комбинированную, ступенчатую и многофазную. *По степени охвата* единиц исследуемой совокупности различают большие и малые выборки.

В дальнейшем мы будем применять следующие условные обозначения:

N - объем генеральной совокупности (число входящих в нее единиц);

n - объем выборки (число обследованных единиц);

$\bar{x}$ - генеральная средняя (среднее значение признака в генеральной совокупности);

$\tilde{x}$ - выборочная средняя;

p - генеральная доля (доля единиц, обладающих данным значением признака, в общем числе единиц генеральной совокупности), например, доля числа бракованных единиц в общем количестве единиц в данной партии изделий;

w - выборочная доля;

σ^2 - генеральная дисперсия (дисперсия признака в генеральной совокупности);

S^2 - выборочная дисперсия того же признака;

σ - среднее квадратическое отклонение в генеральной совокупности;

S - среднее квадратическое отклонение в выборке.

6.2. Простая случайная выборка

Развитие современной теории выборочного наблюдения началось с простой случайной выборки. Лежащие в основе простой случайной выборки понятия и категории являются исходными при разработке других форм выборочного наблюдения.

При простой случайной выборке отбор производится из всей массы единиц генеральной совокупности без предварительного расчленения ее на какие-либо группы, и единица отбора совпадает с единицей наблюдения.

В зависимости от способа отбора единиц различают:

1) отбор по схеме возвращенного шара, обычно называемый **повторной выборкой**. При повторном отборе вероятность попадания каждой отдельной единицы в выборку остается постоянной, так как после отбора какой-то единицы, она снова возвращается в совокупность и снова может быть выбранной;

2) отбор по схеме невозвращенного шара, называемый **бесповторной выборкой**. В этом случае каждая отобранная единица не возвращается обратно, и вероятность попадания отдельных единиц в выборку все время изменяется (для оставшихся единиц она возрастает).

Наиболее просто случайный отбор единиц можно организовать для совокупностей, учитываемых по состоянию на данный момент и включающих в себя счетное множество единиц. В таких случаях есть возможность заранее составить пронумерованный список единиц генеральной совокупности. Отбор из списка единиц может быть произведен путем жеребьевки. Для этого на каждую единицу совокупности заготавливают одинаковую карточку (шар) и проставляют на ней соответствующий номер. В соответствии с определенным объемом выборки из тщательно перемешанных карточек (шаров) последовательно отбирают n карточек, записывая каждый раз номер вынутой карточки (шара). Единицы, номера которых были отобраны, подлежат выборочному наблюдению. Для совокупностей большого объема гораздо удобнее использовать для отбора единиц таблицы случайных чисел (эти таблицы публикуются в приложениях к руководству по математической статистике, в данной книге - Приложение III). Допустим, что в выборку должно войти 75 единиц из списка, содержащего 780 единиц. Открыв таблицу случайных чисел, находим там, например, в первой строке такую последовательность чисел: 5489; 5583; 3156; 0835; 1988; 3912; 0938; 7460; 0869 и 4420. В нашу выборку могут войти только единицы, порядковые номера которых равны трехзначным числам меньше 780. Поэтому, используя только три последние цифры каждого числа, отбираем необходимые 75 номеров, которые и будут включены в выборку. Из приводимых чисел это номера: 489, 583, 156, 460 и т.д. Можно было бы использовать и первые три цифры каждого числа, тогда в выборку были бы включены единицы с порядковыми номерами: 548, 558, 315, 83, 198, 391, 93, 746, 86 и т.д.

Поясним сущность процесса случайного отбора и покажем основные свойства простой случайной повторной выборки на условном примере, который имеет чисто методический характер.

Пусть генеральная совокупность состоит из четырех единиц ($N = 4$), например речь идет о рабочих разной квалификации. У этих единиц зарегистрированы следующие значения изучаемого признака:

Порядковый номер рабочего	1	2	3	4
Тарифный разряд, x_i	3	4	4	5

Параметры генеральной совокупности, подлежащие определению, - это прежде всего генеральная средняя $\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$ и генеральная дисперсия $\sigma^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n}$, которые соответственно равны:

$\bar{x} = 4{,}0$ разряд и $\sigma^2 = 0{,}5$.

Поставим задачу определения параметров генеральной совокупности по результатам проведения простой случайной повторной выборки объемом в две единицы, $n = 2$. В нашем примере с одинаковой степенью вероятности могла бы появиться любая из 16 возможных комбинаций единиц, т.е. любая из 16 возможных выборок. Результаты таких выборок приведены в табл. 6.1.

В рассматриваемом примере можно поставить задачу определения доли рабочих с тарифным разрядом, например равным 4. Доля таких единиц в генеральной совокупности равна 0,50 ($p = 2/4 = 0{,}50$). В 16 возможных выборках варианты выборочной доли оказались равными 0; 0,5; 1,0 (гр. 6 табл. 6.1).

Возможные варианты значений выборочных средних и отклонения их от генеральной средней представлены в форме ряда распределения (см. табл. 6.2).

Таблица 6.1

№ варианта выборки	№ единиц, входящих в данную выборку	Значения признака по данным выборки	Выборочная средняя $\tilde{x}_i$	Отклонение выборочной средней от генеральной средней $\tilde{x}_i - \bar{x}$	Выборочная доля w_i
1	2	3	4	5	6
1	1 и 1	3; 3	3,0	-1,0	0,0
2	1 и 2	3; 4	3,5	-0,5	0,5
3	1 и 3	3; 4	3,5	-0,5	0,5
4	1 и 4	3; 5	4,0	0,0	0,0
5	2 и 1	4; 3	3,5	-0,5	0,5
6	2 и 2	4; 4	4,0	0,0	1,0
7	2 и 3	4; 4	4,0	0,0	1,0
8	2 и 4	4; 5	4,5	+0,5	0,5
9	3 и 1	4; 3	3,5	-0,5	0,5
10	3 и 2	4; 4	4,0	0,0	1,0
11	3 и 3	4; 4	4,0	0,0	1,0
12	3 и 4	4; 5	4,5	+0,5	0,5
13	4 и 1	5; 3	4,0	0,0	0,0
14	4 и 2	5; 4	4,5	+0,5	0,5
15	4 и 3	5; 4	4,5	+0,5	0,5
16	4 и 4	5; 5	5,0	+1,0	0,0

Таблица 6.2

Выборочные средние разряды рабочих, $\tilde{x}_i$	Число выборок с данной выборочной средней f_i	Отклонение выборочной средней от генеральной средней $\tilde{x}_i - \bar{x}$	Вероятность появления данного значения выборочной средней (или величины отклонения выборочной средней от генеральной)
1	2	3	4
3,0	1	-1,0	0,0625
3,5	4	-0,5	0,2500
4,0	6	0,0	0,3750
4,5	4	+0,5	0,2500
5,0	1	+1,0	0,0625
Итого	16		1,0000

Нетрудно заметить, что в распределении величин выборочных средних и их отклонений наблюдаются определенные закономерности.

1. Из возможных результатов простой случайной повторной выборки наиболее вероятны такие, при которых величина выборочной средней будет близка к величине генеральной средней, и, следовательно, разность между выборочной и генеральной средней (ошибка выборочной средней) будет близка к нулю (например, значение $\tilde{x}_i$ = 4,0 разряду, совпадающее с $\bar{x}$ = 4,0 разряду, имеет наибольшую вероятность 0,375, а значения $\tilde{x}_i$, равные 3,0 разряду и 5,0 разряду имеют наименьшую вероятность 0,0625). Таким образом, чем больше величина случайной ошибки выборки, тем менее вероятно появление такой ошибки.

2. В отдельных выборках (в 10 из 16) значения выборочных средних не будут точно совпадать с величиной генеральной средней, но если рассчитать среднюю из всех возможных значений выборочной средней (математическое ожидание выборочной средней), то величина этой средней будет совпадать с величиной генеральной средней.

$$\bar{x} = \frac{\sum_{j=1}^{k} \tilde{x}_j f_j}{\sum_{j=1}^{k} f_j} = \frac{64}{16} = 4{,}0 \text{ разряд}.$$

Другими словами, при простой случайной выборке средняя арифметическая является несмещенной оценкой генеральной средней.

3. В нашем примере не встречаются ошибки больше единицы по абсолютной величине, т.е. всегда существует предел расхождений между выборочной и генеральной средней. Чем больше величина случайной ошибки, тем менее вероятно появление такой ошибки.

В математической теории выборочного метода доказывается, что с увеличением объема выборки вероятность появления больших ошибок и пределы максимально возможной ошибки уменьшаются (чем больше обследуется единиц, тем меньше будет величина расхождений выборочных и генеральных характеристик).

Как было отмечено выше, теоретической основой выборочного метода служат теоремы П.Л. Чебышева и А.М.Ляпунова, Я. Бернулли и С. Пуассона. Неравенство П.Л. Чебышева в приложении к выборочному методу может быть сформулировано так: при неограниченном увеличении числа независимых наблюдений

($n \to \infty$) в генеральной совокупности с ограниченной дисперсией с вероятностью, сколь угодно близкой к единице, можно ожидать, что отклонение выборочной средней от генеральной средней будет сколь угодно мало, т.е.

$$P\left(\left|\tilde{x} - \bar{x}\right| \leq \varepsilon\right) \longrightarrow 1 \text{ при } n \longrightarrow \infty,$$

где P - вероятность неравенства, стоящего в скобках;

ε - любое сколь угодно малое положительное число;

$\bar{x}$ - генеральная средняя.

Таким образом, теорема П.Л. Чебышева доказывает принципиальную возможность определения генеральной средней по данным простой случайной повторной выборки. Однако, пользуясь ею, мы не можем указать вероятность появления ошибок определенной величины.

На этот вопрос отвечает центральная предельная теорема А.М. Ляпунова, доказанная в 1901 г. Согласно этой теореме при достаточно большом числе независимых наблюдений в генеральной совокупности с конечной средней и ограниченной дисперсией, вероятность того, что расхождение между выборочной и генеральной средней $\left|\tilde{x} - \bar{x}\right|$ не превзойдет по абсолютной величине некоторую величину $t\mu$, равна интегралу Лапласа.

Можно записать сказанное таким образом:

$$P\left(\left|\tilde{x} - \bar{x}\right| \leq t\mu\right) = \Phi(t),$$

где $\Phi(t)$ представляет собой нормированную функцию Лапласа.

$$\Phi(t) = \frac{1}{\sqrt{2\pi}} \int_{-t}^{+t} e^{-\frac{t^2}{2}} dt = \frac{2}{\sqrt{2\pi}} \int_{0}^{t} e^{-\frac{t^2}{2}} dt.$$

Величина μ есть средняя квадратическая стандартная ошибка выборки. Из этой теоремы непосредственно следует, что при достаточно большом числе независимых наблюдений, распределение выборочных средних (а следовательно, и их отклонений от генеральной средней) приближенно нормально.

Частным случаем теоремы П.Л. Чебышева является теорема Я. Бернулли:

$$p\left[(w - p) < t\sqrt{\frac{p(1-p)}{n}}\right] > 1 - \frac{1}{t^2},$$

где w и p - доля признака соответственно в выборочной и генеральной совокупности.

Вернемся к табл. 6.1, где были представлены все возможные варианты выборочных средних и их отклонения от генеральной средней. Используя данные табл. 6.2, рассчитаем величину стандартной ошибки выборки по формуле средней квадратической.

$$\mu = \sqrt{\frac{\sum_{j=1}^{k}\left(\tilde{x}_j - \bar{x}\right)^2 f_j}{\sum_{j=1}^{k} f_j}} = \sqrt{\frac{4}{16}} = 0{,}5 \text{ разряда.}$$

Однако на практике исследователь оперирует данными какой-то одной конкретной выборки, а потому указанным способом определить стандартную ошибку средней невозможно. При этом важно указать, насколько в среднем возможно отклонение выборочной средней от генеральной средней при данных условиях отбора.

В математической статистике доказывается, что величина *средней квадратической стандартной ошибки простой случайной повторной выборки* может быть определена по формуле:

$$\mu_{\bar{x}} = \sqrt{\frac{\sigma^2}{n}} = \frac{\sigma}{\sqrt{n}}. \qquad (6.1)$$

В нашем примере величина генеральной дисперсии равна 0,5 (см. расчет на с. 162), а объем выборки n равен 2.

Отсюда $\mu_{\bar{x}} = \sqrt{\frac{0{,}5}{2}} = 0{,}5$ разряда.

Из формулы средней квадратической ошибки простой случайной выборки видно, что величина $\mu_{\bar{x}}$ зависит от колеблемости признака в генеральной совокупности (чем больше вариация признака, тем больше ошибка выборки) и от объема выборки (чем больше обследуется единиц, тем меньше будет величина расхождений выборочных и генеральных характеристик).

Величину $t\mu_{\bar{x}}$ называют предельной ошибкой выборки. Обозначив предельную ошибку выборки $\Delta_{\bar{x}}$, получим

$$\Delta_{\bar{x}} = t\mu_{\bar{x}}, \qquad (6.2)$$

т.е. предельная ошибка выборки равна t-кратному числу средних ошибок выборки. Допустим, что $t = 2$.

Тогда[1]

$$P\left(\left|\tilde{x} - \bar{x}\right| \leq 2\mu_{\bar{x}}\right) = \Phi_{(t=2)} = 0{,}9545$$

т.е. с вероятностью, равной 0,9545, можно ожидать, что ошибка выборочной средней не превысит удвоенной средней квадратической ошибки выборки. Таким образом, величина предельной ошибки выборки может быть установлена с определенной вероятностью.

Приведем наиболее часто употребляемые уровни доверительной вероятности и соответствующие значения t для выборок достаточно большого объема ($n \geq 30$):

t	1,00	1,96	2,00	2,58	3,00
$\Phi(t)$	0,683	0,950	0,954	0,990	0,997

Как видно из последней графы, вероятность появления ошибки, равной или большей утроенной средней ошибки выборки, т.е. $|\Delta_{\bar{x}}| \geq 3\mu_{\bar{x}}$ крайне мала и равна 0,003 (1 - 0,997). Такие маловероятные события считаются практически невозможными, а потому величину $|\Delta_{\bar{x}}| = 3\mu_{\bar{x}}$ можно принять за предел возможной ошибки выборки.

Распределение выборочной доли представим в табл. 6.3.

Таблица 6.3

Выборочная доля w_j	Число выборок с данной выборочной долей f_j	Отклонение выборочной доли от генеральной $w_j - p$	$w_j f_j$	$(w_j - p)^2 f_j$
0,0	4	-0,5	0,0	1,0
0,5	8	0,0	4,0	0
1,0	4	+0,5	4,0	1,0
Итого	16		8,0	2,0

В среднем для всех возможных вариантов выборок величина выборочной доли совпадает с долей признака в генеральной совокупности:

[1] Значения интегральной функции Лапласа для различных значений t вычислены и приводятся в таблицах (см. Приложение II).

$$\bar{w} = \frac{\sum_{i=1}^{k} w_i f_i}{\sum_{i=1}^{k} f_i} = \frac{8,0}{16} = 0,5 = p.$$

Средняя квадратическая ошибка доли по выборке:

$$\mu_p = \sqrt{\frac{\sum_{i=1}^{k} (w_i - p)^2 f_i}{\sum_{i=1}^{k} f_i}} = \sqrt{\frac{2,0}{16}} = 0,354 .$$

Ранее было показано, что дисперсия доли σ^2_p равна $p(1-p)$, поэтому величину средней стандартной ошибки выборочной доли можно определить по следующей формуле:

$$\mu_p = \sqrt{\frac{p\,(1-p)}{n}}, \tag{6.3}$$

а предельную ошибку доли по формуле:

$$\Delta_p = t\mu_p \tag{6.3a}$$

В рассматриваемом примере $p = 0,5$; $n = 2$ и

$$\mu_p = \sqrt{\frac{0,5 \cdot 0,5}{2}} = 0,354 .$$

В формулы (6.1) и (6.3) средней ошибки выборки входит дисперсия признака в генеральной совокупности, величина которой, как правило, при проведении выборочного наблюдения неизвестна. Нам приходится использовать выборочную дисперсию в качестве оценки генеральной дисперсии.

Можно доказать, что $\sigma^2 = S^2 \cdot \frac{n}{n-1}$, следовательно, генеральную дисперсию можно оценить по результатам одной выборки по следующей формуле:

$$\sigma^2 = \frac{\sum_{i=1}^{n} \left(x_i - \tilde{x}\right)^2}{n-1},$$

где $(n-1)$ - число степеней свободы для расчета дисперсии по выборочным данным[1].

Применение простой случайной повторной выборки на практике весьма ограничено. Прежде всего практически нецелесообразно, а иногда невозможно повторное наблюдение одних и тех же единиц, а поэтому однажды обследованная единица повторному учету не подвергается. Применение бесповторного отбора взамен повторного диктуется также требованием повышения степени репрезентативности выборки (особенно при недостаточно больших n).

Средняя квадратическая ошибка случайной бесповторной выборки определяется по формулам[2]

$$\mu_x = \sqrt{\frac{\sigma^2}{n}\left(1 - \frac{n}{N}\right)}, \tag{6.4}$$

$$\mu_p = \sqrt{\frac{p\,(1-p)}{n}\left(1 - \frac{n}{N}\right)}. \tag{6.5}$$

Сопоставление формул (6.1) и (6.4); (6.3) и (6.5) свидетельствует о том, что применение бесповторного отбора взамен повторного приводит к уменьшению стандартной ошибки выборки.

В тех случаях, когда численность генеральной совокупности N очень велика по сравнению с числом отобранных единиц n, величина $1 - \frac{n}{N}$ будет близка к единице, а потому ею можно пренебречь. Тогда ошибку случайного бесповторного отбора опреде-

[1] При большом объеме выборки ($n > 30$) разница между дисперсией, вычисленной по формуле $\sigma^2 = \frac{\sum_{i=1}^{n}\left(x_i - \tilde{x}\right)^2}{n}$ и по формуле $\sigma^2 = \frac{\sum_{i=1}^{n}\left(x_i - \tilde{x}\right)^2}{n-1}$, очень невелика, и поправка к выборочной дисперсии необязательна (так уже для $n = 30$ получим $\frac{30}{30-1} = 1,035$, т.е. выборочная дисперсия будет на 3,5% меньше генеральной).

[2] Эти формулы являются упрощением строгой формулы $\mu_{\bar{x}} = \sqrt{\frac{\sigma^2}{n}\left(\frac{N-n}{N-1}\right)}$, так как при достаточно больших N: $\frac{N-n}{N-1} \approx 1 - \frac{n}{N}$.

ляют по формуле простой случайной повторной выборки, что повышает надежность оценок генеральных характеристик по выборочным данным.

Выборочное наблюдение проводится в целях распространения выводов, полученных по данным выборки, на генеральную совокупность. Одной из основных задач является оценка по данным выборки интересующих нас характеристик (параметров) генеральной совокупности. Рассмотрим определение величины средней арифметической генеральной совокупности на основе выборочных данных.

Выборочное наблюдение даст возможность определить среднюю арифметическую выборочной совокупности $\tilde{x}$ и величину предельной ошибки этой средней $\Delta_{\bar{x}}$, которая показывает (с определенной вероятностью), насколько выборочная средняя может отличаться от генеральной средней в большую или меньшую сторону. Тогда величина генеральной средней будет представлена интервальной оценкой, для которой нижняя граница будет равна $\tilde{x} - \Delta_{\bar{x}}$, а верхняя граница - $\tilde{x} + \Delta_{\bar{x}}$. Пределы, в которых с данной степенью вероятности будет заключена неизвестная величина оцениваемого параметра, называют *доверительными*, а вероятность *P - доверительной вероятностью.*

Доверительный интервал для генеральной средней можно записать как:

$$\tilde{x} - t\mu_{\bar{x}} \leqslant \bar{x} \leqslant \tilde{x} + t\mu_{\bar{x}} \qquad (6.6)$$

Чаще всего доверительную вероятность устанавливают равной 0,95 или 0,99 (величины коэффициентов доверия t равны соответственно 1,96 и 2,58).

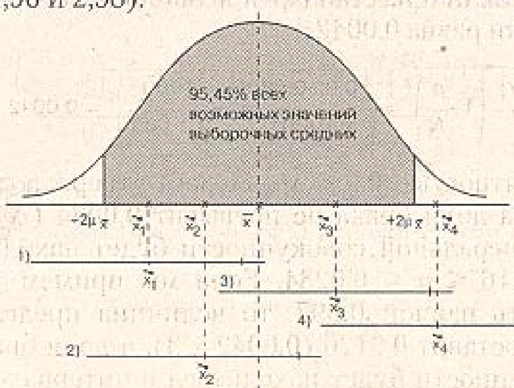

Рис. 6.1. **Доверительные интервалы выборочных средних (P=0,9545)**

На рис. 6.1 показаны доверительные интервалы для четырех различных результатов выборочных наблюдений, построенных при вероятности 0,9545. Для каждой выборки под номерами 1); 2); 3) и 4) доверительный интервал для $\bar{X}$ будет определен так:

$$\tilde{x}_j - 2\mu_{\bar{x}} \leqslant \bar{x} \leqslant \tilde{x}_j + 2\mu_{\bar{x}} \qquad (6.7)$$

где $\tilde{x}_j$ - средняя арифметическая j-той выборки.

Только доверительный интервал, построенный для выборочной средней $\tilde{x}_4$, не включает генеральную среднюю.

Вероятность того, что величина генеральной средней выйдет за доверительные границы, будет равна $1-P$, т.е. при $P = 0{,}9545$ будет равна 0,0455 или 0,01 при $P = 0{,}99$. Событие, обладающие столь малой вероятностью, считается практически невозможным.

Аналогичным образом могут быть записаны доверительные пределы генеральной доли:

$$w - t\mu_p \leqslant p \leqslant w + t\mu_p \qquad (6.7а)$$

Величина доверительного интервала для генеральной средней или генеральной доли зависит от величины предельной ошибки выборки $\Delta_{\bar{x}}$ или Δ_p. Чем больше величина предельной ошибки выборки, тем больше величина доверительного интервала и тем, следовательно, ниже точность оценки.

Поскольку величина предельной ошибки выборки равна $t\mu$, точность оценки параметров генеральной совокупности будет зависеть от принятого уровня доверительной вероятности и от величины стандартной ошибки выборки. Допустим, что доля брака по данным выборки w составила 0,02 (или 2%) при $n = 1000$ и $N = 10\,000$. Величина стандартной ошибки случайной бесповторной выборки равна 0,0042.

$$\mu_p = \sqrt{\frac{w(1-w)}{n}\left(1-\frac{n}{N}\right)} = \sqrt{\frac{0{,}02 \cdot 0{,}98}{1000}\left(1-\frac{1000}{10\ 000}\right)} = 0{,}0042 \ (\text{или } 0{,}42\%).$$

С вероятностью 0,954 мы можем утверждать, что предельная ошибка доли брака не превысит 0,0084 ($\Delta_p = 2\mu_p$) и доля брака в генеральной совокупности будет находиться в интервале: $0{,}0116 \leqslant p \leqslant 0{,}0284$. Если мы примем доверительную вероятность равной 0,997, то величина предельной ошибки выборки составит 0,0126 (0,0042 x 3), и доля брака в генеральной совокупности будет находиться в интервале $0{,}02 \pm 0{,}0126$, т.е. $0{,}74\% \leqslant p \leqslant 3{,}26\%$.

Таким образом, с вероятностью 99,7% можно ожидать, что количество бракованных деталей во всей партии из 10 000 штук будет находиться в интервале от 74 до 327 штук, тогда как с вероятностью 95,4% доверительный интервал составит 116 ÷284 штуки. Изменение точности оценки генеральной доли в связи с изменением доверительной вероятности показано на рис. 6.2.

Незначительно увеличивая достоверность выводов (с 95,4 до 99,7%), мы существенно снижаем точность оценки. В этой связи в экономических расчетах чаще рекомендуется использовать доверительную вероятность $P = 0{,}95$ или $P = 0{,}954$ (соответственно t равно 1,96 и 2,00).

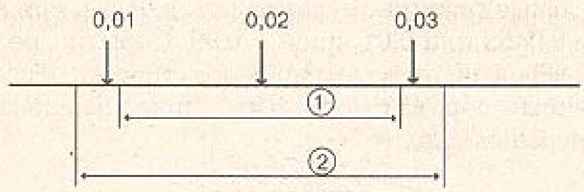

Рис. 6.2. **Доверительные интервалы генеральной доли:**

① - вероятность $P = 0{,}954$; ② - вероятность $P = 0{,}997$.

6.3. Определение необходимой численности выборки

Средняя квадратическая (стандартная) ошибка выборки зависит от объема выборки и степени вариации признака в генеральной совокупности. Уменьшение стандартной ошибки выборки, а следовательно, увеличение точности оценки, всегда связано с увеличением объема выборки. В этой связи уже на стадии организации выборочного наблюдения приходится решать вопрос о том, каков должен быть объем выборочной совокупности, чтобы была обеспечена требуемая точность результатов наблюдений. Рассмотрим формулу предельной ошибки выборки для случая простой случайной повторной выборки

$$\Delta_{\bar{x}} = t\mu_{\bar{x}} = t\sqrt{\frac{\sigma^2}{n}} = \frac{t\sigma}{\sqrt{n}}.$$

Отсюда

$$n = \frac{t^2\sigma^2}{\Delta^2_{\bar{x}}}. \qquad (6.8)$$

При проектировании выборочного наблюдения предполагается заранее заданной величина допустимой ошибки выборки в соответствии с задачами конкретного исследования (следует иметь в виду, что в данной формуле $\Delta_{\bar{x}}$ - абсолютная величина предельной ошибки выборки) и вероятность выводов по результатам наблюдения (величина коэффициента t соответствует принятому уровню доверительной вероятности).

Величина σ^2, характеризующая дисперсию признака в генеральной совокупности, зачастую бывает неизвестна. Поэтому используют приближенные способы оценки генеральной дисперсии.

1. Можно провести «пробное» обследование (обычно небольшого объема), на базе которого определяется величина дисперсии признака, используемая в качестве оценки генеральной дисперсии.

$$\sigma^2 = \frac{\sum_{i=1}^{n}\left(x_i - \tilde{x}_{\text{проб.}}\right)^2}{n_{\text{проб.}} - 1}$$

где $\tilde{x}_{\text{проб.}}$ - средняя арифметическая по результатам пробного обследования;

$n_{\text{проб.}}$ - число единиц, попавших в пробное обследование.

По данным нескольких пробных обследований выбирается наибольшее значение дисперсии.

2. Можно использовать данные прошлых выборочных обследований, проводившихся в аналогичных целях, т.е. дисперсия, полученная по их результатам используется в качестве оценки генеральной дисперсии.

3. Если распределение признака в генеральной совокупности может быть отнесено к нормальному закону распределения, то размах вариации примерно равен 6σ (крайние значения отстоят в ту и другую сторону от средней на расстоянии 3σ), т.е. $R \approx 6\sigma$ откуда $\sigma \approx 1/6R$, где $R = x_{max} - x_{min}$.

При проведении социально-экономических исследований, как правило, можно с достаточной точностью указать максимально и минимально возможные значения признака в исследуемой совокупности. Например, сколько операционистов нужно обследовать в банках региона, чтобы получить характеристику среднего уровня оплаты труда этой категории работников банков в регионе? Положим, нам известно, что разница между наивысшим и наименьшим уровнем оплаты труда операциониста в регионе составляет 300 тыс. руб. Для нормального распределения в промежуток $\bar{x} \pm 3\sigma$ включается 99,7% всех вариантов значений признака, а это означает применительно к рассматриваемой задаче, что 300 тыс. руб. примерно равно шести средним квадратическим отклонениям ($300 \approx 6\sigma$). Поэтому примерная оценка среднего квадратического отклонения заработной платы в генеральной совокупности операционистов региона составит 50 тыс. руб. ($\sigma = \frac{300}{6}$). Для дальнейших расчетов достаточно, чтобы с вероятностью 0,954 предельная ошибка выборки не превышала 10 тыс. руб. Тогда, зная что $\sigma = 50$ тыс. руб. , а $t = 2$, и используя формулу (6.8) для определения необходимого объема выборки, получим:

$$n = \frac{2^2 \cdot 50^2}{10^2} = 100 \text{ человек}$$

Таким образом, при заданных условиях нужно обследовать размер заработной платы у 100 операционистов региона.

В связи с тем, что генеральная дисперсия оценивается приближенно, рекомендуется величину объема выборки, рассчитанную по формуле (6.8), округлять в большую сторону. Эту же формулу рекомендуется использовать для определения объема выборки и в тех случаях, когда проводится случайная бесповторная выборка, так как всегда необходим некоторый «запас» числа обследованных единиц для обеспечения требуемой точности результатов.

Нередко на практике задается не величина абсолютной предельной ошибки $\Delta_{\bar{x}}$, а величина относительной погрешности $\Delta_{\text{относ.}}$, выраженная в процентах к средней:

$$\Delta_{\text{относит.}} = \frac{\Delta_{\bar{x}}}{\bar{x}} \cdot 100\%,$$

откуда

$$\Delta_{\bar{x}} = \frac{\Delta_{\text{относит.}} \cdot \bar{x}}{100\%}$$

Подставив величину $\Delta_{\bar{x}}$, выраженную через относительную погрешность, в формулу (6.8), получим следующее выражение для определения необходимого объема выборки:

$$n = \frac{t^2 \sigma^2}{\Delta^2_{\text{относит.}} \cdot \left(\bar{x}\right)^2} \cdot 100^2$$

Как известно, отношение $\frac{\sigma}{\bar{x}} \cdot 100\%$ представляет собой коэффициент вариации v, откуда

$$n = \frac{t^2 v^2}{\Delta^2_{\text{относит.}}}.$$

Например, по данным пробного обследования коэффициент вариации составляет 40%. Сколько нужно отобрать единиц, чтобы с вероятностью 0,954 предельная относительная ошибка выборки не превышала 5%?

При $v = 40\%$, $\Delta_{\text{относит.}} = 5\%$ и $t = 2$

$$n = \frac{4 \cdot 1600}{25} = 256 \text{ единиц}.$$

При определении по материалам выборки доли признака, а не средней его величины, объем выборочной совокупности определится по следующим формулам:

а) для повторного отбора

$$n = \frac{t^2 w\,(1-w)}{\Delta_p^2}, \tag{6.9}$$

б) для бесповторного отбора

$$n = \frac{t^2 w\,(1-w) N}{N\Delta_p^2 + t^2 w\,(1-w)} \tag{6.10}$$

Для случая, когда частость w даже приблизительно неизвестна, можно произвести примерный расчет численности выборки, вводя в расчет максимальную величину дисперсии доли, равную 0,25:

а) для повторного отбора

$$n = \frac{t^2}{4\Delta_p^2}, \tag{6.9а}$$

б) для бесповторного отбора

$$n = \frac{t^2 N}{4N\Delta_p^2 + t^2}. \tag{6.10а}$$

Очень часто цель выборочного наблюдения сводится к изучению различных признаков, характеризующих данную совокупность. Причем зачастую по степени колеблемости эти признаки могут сильно отличаться друг от друга. Кроме того, для различных признаков может требоваться разная степень точности. В таких случаях при определении необходимого объема выборки следует ориентироваться на тот признак, который при наибольшей колеблемости обладает наименьшей величиной допустимой ошибки.

Например, при проведении выборочного наблюдения требуется определить средний вес изделия и долю изделий первого сорта в совокупности объемом 5000 единиц. Устанавливаем требуемую точность: для среднего веса - 1 г; для доли - 5%. По данным сплошного обследования изделий первой партии, состоящей из 100 единиц, определили дисперсию веса, равную 36, и дисперсию доли, равную 0,05. Требуется определить необходимую численность случайной бесповторной выборки, обеспечивающей заданную точность наблюдения с вероятностью 0,954.

Исходные данные: а) для определения n, обеспечивающего заданную точность оценки среднего веса изделия.

$$N = 5000;\ \sigma^2_x = 36;\ \Delta_{\tilde{x}} = 1;\ P = 0{,}954;\ t = 2.$$

$$n = \frac{t^2\sigma^2 N}{N\Delta^2_{\tilde{x}} + t^2\sigma^2_x} = \frac{4 \cdot 36 \cdot 5000}{5000 \cdot 1 + 4 \cdot 36} = 140 \text{ шт.}$$

б) для определения необходимого объема выборки, обеспечивающего требуемую точность в определении доли изделий первого сорта, мы располагаем такими данными:

$$N = 5000;\ \Delta_p = 0{,}05;\ \sigma^2_p = 0{,}05;\ P = 0{,}954;\ t = 2.$$

$$n = \frac{t^2\sigma^2_p N}{N\Delta^2_p + t^2\sigma^2_p} = \frac{4 \cdot 0{,}05 \cdot 5000}{5000 \cdot 0{,}0025 + 4 \cdot 0{,}05} = 80 \text{ шт.}$$

Очевидно, что число изделий, обеспечивающее требуемую точность результатов выборочного наблюдения для определения доли изделий первого сорта недостаточно, чтобы обеспечить требуемую точность при определении среднего веса изделий. Поэтому следует произвести отбор из генеральной совокупности не менее 140 изделий.

В целом формула предельной ошибки выборочной средней (доли), позволяет решать следующие три группы задач:

1) определять предел возможной ошибки средней (доли), т.е. величину возможных отклонений показателей генеральной совокупности от показателей выборочной совокупности;

2) определять необходимую численность выборки, обеспечивающую требуемую точность, при которой пределы возможной ошибки не превысят некоторой наперед заданной величины;

3) определять вероятность того, что в проведенной выборке ошибка будет иметь заданный предел.

Для решения всех перечисленных задач необходимо знать дисперсию признака в генеральной совокупности σ^2 и в случае бесповторной выборки ее объем N.

Чтобы оценить предельную ошибку выборки, нужно дополнительно задать определенную вероятность заключений, или, что то же, величину коэффициента доверия t и знать объем выборки n.

При расчете объема выборки, необходимого для обеспечения заданной точности результатов наблюдения, нужно знать предельную ошибку выборки и величину коэффициента доверия t. Определяя численность выборки и ее точность, следует учитывать, что чем больше абсолютный объем выборки, тем менее ощутимо влияет на точность результата включение в выборку дополнительных десятков и даже сотен единиц и тем больших затрат требует дальнейшее повышение точности.

В целях упрощения определения объема выборки можно воспользоваться заранее рассчитанными таблицами, в которых указывается необходимое число наблюдений при данной величине доверительной вероятности и допустимой ошибке. Например, при определении доли признака в исследуемой совокупности можно воспользоваться такой таблицей (см. табл. 6.4). Необходимое число наблюдений рассчитано исходя из формулы предельной ошибки повторного отбора при максимальной величине дисперсии доли, равной 0,25.

Таблица 6.4

Доверительная вероятность	Предельная ошибка, %			
	10	5	2	1
0,90	67	270	1690	6763
0,95	96	384	2400	9603
0,99	165	663	4146	16587
0,997	220	880	5504	22018

Из данных таблицы видно, что для того чтобы гарантировать с вероятностью 0,95 величину отклонения от генеральной доли, не превышающую 10%, нужно обследовать 96 единиц. Если предельную ошибку уменьшить до 5%, то нужно подвергнуть обследованию уже 384 единицы, т.е. в четыре раза больше, а для обеспечения с вероятностью 0,95 точности в 1% при определении доли нужно обследовать уже 9603 единицы. Таким образом, таблица позволяет исследователю сделать вывод о целесообразности увеличения объема выборки, а следовательно, и о целесообразности дополнительных затрат, обеспечивающих то или иное повышение точности. Если задана предельная ошибка выборки Δ и объем выборки n, то можно определить величину коэффициента t, зная которую по таблице можно определить вероятность $\Phi(t)$.

Вернемся к примеру, рассмотренному в параграфе 6.2. В выборке объемом 1000 единиц доля бракованных изделий составила 2%. Какова вероятность того, что во всей партии изделий (10 000 штук) доля бракованных изделий будет находиться в пределах от 1,5 до 2,5%?

Доверительная вероятность, которую требуется определить, является функцией t, последняя находится из формулы предельной ошибки выборки $\Delta_p = t\mu_p$, откуда $t = \frac{\Delta_p}{\mu_p}$. Величина предельной ошибки выборки может быть определена как разность между максимально допустимой генеральной долей (по условию она равна 2,5%) и долей бракованных изделий в выборке (по условию 2%).

Таким образом, Δ_p = 0,5% (2,5% - 2,0%). Так как выборка случайная бесповторная, то величина средней ошибки выборки находится по формуле

$$\mu_p = \sqrt{\frac{w(1-w)}{n}\left(1-\frac{n}{N}\right)} = \sqrt{\frac{2 \cdot 98}{1000}\left(1-\frac{1000}{10\ 000}\right)} = 0{,}42\% .$$

Находим величину коэффициента доверия:

$$t = \frac{0{,}5}{0{,}42} = 1{,}19$$

По таблицам интегральной функции Лапласа (см. Приложение II) вероятность, соответствующая данной величине коэффициента t, равна 0,76595.

6.4. Малые выборки

Как уже указывалось, для определенного способа отбора единиц величина стандартной ошибки зависит от объема выборки и степени колеблемости изучаемого признака в генеральной совокупности. Причем, чем меньше объем выборки, тем большую величину стандартной ошибки следует ожидать, а это в свою очередь снижает точность оценки параметров генеральной совокупности.

При выборках небольшого объема величина выборочной дисперсии, используемой в качестве оценки дисперсии генеральной совокупности, может быть в значительной мере подвержена влиянию случайностей выборки. Поэтому при выборках небольшого объема методы оценки результатов выборочного наблюдения видоизменяются в сравнении с применяемыми в теории больших выборок. (К безусловно малым относятся выборки объемом менее 30 единиц.)

Первые работы в области теории малой выборки были выполнены английским статистиком В. Госсетом в 1908 г. (псевдоним - «Стьюдент») и продолжены в исследованиях Р. Фишера.

Для оценки возможных пределов ошибки малой выборки пользуются так называемым отношением Стьюдента:

$$t = \frac{\tilde{x} - \bar{x}}{S : \sqrt{n-1}} . \qquad (6.11)$$

Знаменатель приведенной формулы представляет собой меру случайных колебаний выборочной средней в малой выборке $\mu_{м.в.}$:

$$\mu_{м.в.} = \frac{S}{\sqrt{n-1}} \qquad (6.11а)$$

где S - величина среднего квадратического отклонения для данной выборки: $S = \sqrt{\frac{\sum_{i=1}^{n}\left(x_i - \tilde{x}\right)^2}{n}}$, не рассматриваемая как оценка генеральной дисперсии. Таким образом, теоретическое распределение отношения Стьюдента t имеет дело с величинами, определяемыми непосредственно по данным выборки.

Закон распределения величины t носит название **Закон распределения Стьюдента**, согласно которому плотность распределения ошибок малой выборки $Sk(t)$ равна:

$$Sk(t) = A\left(1+\frac{t^2}{k}\right)^{-\frac{k+1}{2}},$$

где k - число степеней свободы варьирования при определении выборочной дисперсии, равное $n-1$,

A - определяется в зависимости от числа степеней свободы с помощью гамма-функции (Г-функции):

$$A = \frac{\Gamma\left(\frac{k+1}{2}\right)}{\Gamma\left(\frac{k}{2}\right)\cdot\sqrt{\pi k}}.$$

Вероятность того, что ошибка выборки будет не больше заданной величины $t\mu_{м.в.}$ определяется интегральной функцией:

$$Sk(t) = A\int_{-\infty}^{t}\left(1+\frac{t^2}{k}\right)^{-\frac{k+1}{2}}dt.$$

Следует учитывать, что закон Стьюдента приложим только к оценке ошибок выборок, взятых из генеральной совокупности с нормальным распределением признака.

Кривая t - распределения симметрична относительно оси ординат. Как видно из приводимой функции плотности, координаты кривой распределения Стьюдента зависят не только от t, но и от числа степеней свободы k.

Для больших значений k (больших 30) кривая $Sk(t)$ очень близка к кривой нормального распределения. Но при малом числе степеней свободы она значительно отклоняется от кривой нормального распределения, медленнее спускаясь к оси абсцисс, а это значит, что значительные отклонения от средней имеют здесь большую вероятность, чем для нормального распределения. Следовательно, вероятность больших ошибок в малой выборке выше, чем при больших выборках.

Составленные таблицы распределения Стьюдента облегчают применение отношения t на практике (см. Приложение IV). Приведем выдержку из такой таблицы $Sk(t)$ для некоторых значений k и t.

Таблица 6.5

t \ k	4	5	9	10	15	20	25	Нормальное распределение
1,0	0,813	0,818	0,828	0,830	0,833	0,835	0,838	0,841
2,0	0,942	0,949	0,962	0,963	0,968	0,970	0,973	0,977
3,0	0,980	0,985	0,992	0,993	0,995	0,996	0,997	0,999

Приведенная таблица показывает вероятность того, что фактическое отношение Стьюдента $t_ф$ в условиях случайной выборки будет не больше приведенного в таблице, т.е. $P(t_ф < t) = Sk(t)$. Но по этим же таблицам можно определить и вероятность того, что фактическое отношение не превзойдет по абсолютной величине табличное значение, т.е. $P(t_ф < |t|)$. Эта вероятность определяется по формуле:

$$P\left(t_ф < |t|\right) = Sk(t) - \left[1 - Sk(t)\right] = 2Sk(t) - 1.$$

Например, при $k = 20$, вероятность того, что предельная ошибка выборки будет не более 3μ, равна 0,996 (см. табл. 6.5). Вероятность того, что ошибка будет заключена в пределах $\pm 3\mu$ будет равна: $2\cdot 0{,}996 - 1 = 0{,}992$. Таблицу Стьюдента часто излагают в краткой удобной для практического использования форме: для каждого числа степеней свободы k указывают величину отношения Стьюдента t, которая с заданной вероятностью (0,95 или 0,99) не будет превышена по абсолютной величине в силу случайностей отбора (табл. 6.6).

Таблица 6.6

Величины отношения Стьюдента t (краткая выдержка)

k \ P	0,95	0,99	0,997
4	2,776	4,604	6,435
9	2,262	3,250	4,024
14	2,145	2,977	3,583
20	2,086	2,845	3,376
24	2,064	2,797	3,302
28	2,048	2,763	3,250

Приведем пример, иллюстрирующий практическое применение распределения Стьюдента. Оздоровительный центр, рекламируя свои услуги, предлагает клиентам за короткий срок снижение веса до 10 кг. По результатам выборочного обследования 15 женщин, воспользовавшихся услугами центра, были получены следующие данные о снижении их веса:

№ п/п	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Снижение веса, кг	10,2	7,6	6,1	8,4	6,0	5,7	13,7	6,9	5,2	6,1	5,0	3,7	4,7	3,6	3,2

Выборочная средняя составляет 6,41 кг ($\tilde{x} = \frac{96,1}{15} = 6,41$ кг), т.е. среднее снижение веса у обследованных женщин составило 6,41 кг.

Выборочная дисперсия равна 7,061 ($s^2 = \frac{\sum_{i=1}^{15}\left(x_i - \tilde{x}\right)^2}{15} = 7,061$).

Следовательно, средняя квадратическая ошибка выборки составит 0,71 кг.

$$\mu_{м.в.} = \sqrt{\frac{s^2}{n-1}} = \sqrt{\frac{7,061}{14}} = 0,71 \text{ кг}.$$

Оценим с вероятностью 0,99 предел возможных расхождений выборочной средней и генеральной средней. Так как число степеней свободы равно 14 ($k = n-1 = 15-1$), то по таблице 6.6 находим, что значение t, соответствующее вероятности 0,99, равно 2,977.

Тогда с вероятностью 0,99 можно предполагать, что ошибка выборочной средней будет не больше 2,114 кг (2,977 х 0,71), а снижение веса пациентов оздоровительного центра будет находиться в пределах от 4,3 до 8,52 кг (6,41 $\pm$2,11). Следовательно, указанное в рекламе снижение веса на 10 кг имеет столь малую вероятность, что считается событием практически невозможным.

6.5. Статистическая проверка гипотез (Общие понятия. Выбор типа критической области)

Под статистической гипотезой понимаются различного рода предположения относительно характера или параметров распределения случайной переменной, которые можно проверить, опираясь на результаты наблюдений в случайной выборке.

Следует иметь в виду, что статистическая проверка гипотез имеет вероятностный характер, так как принимаемые выводы основываются на изучении свойств распределения случайной переменной по данным выборки, а потому всегда существует риск допустить ошибку. Однако с помощью статистической проверки гипотез можно определить вероятность принятия ложного решения. Если вероятность последнего невелика, то можно считать, что применяемый критерий обеспечивает малый риск ошибки.

При проверке гипотез имеется возможность совершить ошибку двоякого рода:

а) ошибка первого рода - проверяемая гипотеза (ее обычно называют нулевой гипотезой и обозначают H_0) является в действительности верной, но результаты проверки приводят к отказу от нее;

б) ошибка второго рода - проверяемая гипотеза в действительности является ошибочной, но результаты проверки приводят к ее принятию.

Для построения статистического критерия, позволяющего проверить некоторую гипотезу, необходимо следующее:

1) сформулировать проверяемую гипотезу H_0. Наряду с проверяемой гипотезой формулируется также конкурирующая (альтернативная) гипотеза;

2) выбрать уровень значимости α, контролирующий допустимую вероятность ошибки первого рода;

3) определить область допустимых значений и так называемую критическую область;

4) принять то или иное решение на основе сравнения фактического и критического значений критерия.

Уровнем значимости будем называть такое малое значение вероятности попадания критерия в критическую область при условии справедливости гипотезы, что появление этого события может расцениваться как следствие существенного расхождения выдвинутой гипотезы и результатов выборки. Обычно уровень значимости принимают равным 0,05 или 0,01. Исходя из величины уровня значимости, можно определить *критическую область*, под которой понимается такая область значений выборочной характеристики, попадая в которую они будут свидетельствовать о том, что проверяемая гипотеза должна быть отвергнута. К критической области относятся те значения, появление которых при условии верности гипотезы было бы маловероятным.

Допустим, что рассчитанное по эмпирическим данным значение критерия попало в критическую область, тогда при условии верности проверяемой гипотезы H_0 вероятность этого события будет не больше уровня значимости α. Поскольку α выбирается достаточно малым, то такое событие является маловероятным и, следовательно, проверяемая гипотеза H_0 может быть отвергнута.

Если же наблюдаемое значение характеристики не принадлежит к критической области и, следовательно, находится в области допустимых значений, то проверяемая гипотеза H_0 не отвергается. Вероятность попадания критерия в область допустимых значений при справедливости проверяемой гипотезы H_0 равна $1 - \alpha$.

Чем меньше уровень значимости, тем меньше вероятность забраковать проверяемую гипотезу, когда она верна, т.е. меньше вероятность совершить ошибку первого рода. Но при этом расширяется область допустимых значений и, значит, увеличивается вероятность совершения ошибки второго рода.

При данном уровне значимости мы можем по-разному устанавливать критическую область, гарантирующую этот уровень. Пусть мы задаемся уровнем значимости $\alpha = 0{,}05$. В качестве критической области, отвечающей этому уровню значимости, мы можем взять:

1) область больших положительных отклонений так, чтобы вероятность того, что значение x выйдет за пределы $\bar{x} + t_1\mu_{\bar{x}}$ была равна уровню значимости α. При $\alpha = 0{,}05$ сказанное можно записать так:

$$P_1 = P\left(x > \bar{x} + t_1\mu_{\bar{x}}\right) = 0{,}05 .$$

Тогда $$P_1 = \frac{1}{\sqrt{2\pi}}\int_{t_1}^{\infty} e^{-\frac{t^2}{2}}\,dt = \frac{1}{2} - \frac{1}{\sqrt{2\pi}}\int_0^{t} e^{-\frac{t^2}{2}}\,dt,$$ откуда

$$\frac{1}{\sqrt{2\pi}}\int_0^{t} e^{-\frac{t^2}{2}}\,dt = \frac{1}{2} - 0{,}05 = 0{,}45 .$$

По таблице нормированной функции Лапласа находим, что $t_1 = 1{,}64$ (см. Приложение II).

Теперь граница критической области будет найдена из выражения $x > \bar{x} + 1{,}64\,\mu_{\bar{x}}$. В данном случае мы имеет одностороннюю критическую область, графическое представление которой приведено на рис. 6.3.

2) область больших отрицательных отклонений:

$$P_2 = P\left(x < \bar{x} - t_2\mu_{\bar{x}}\right) = 0{,}05, \text{ т.е. } P_2 = P\left(x < \bar{x} - 1{,}64\mu_{\bar{x}}\right).$$

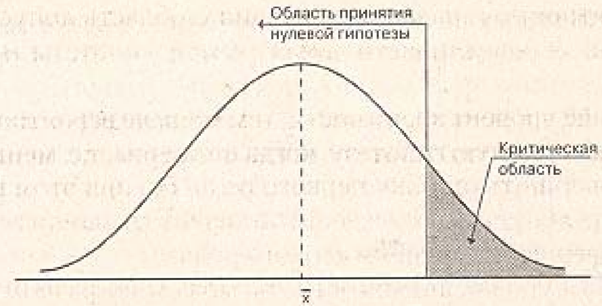

Рис. 6.3. **Правосторонняя критическая область (область больших положительных отклонений)**

3) область больших по абсолютной величине отклонений (см. рис. 6.4).

$$P_3 = P\left(\left|x - \bar{x}\right| > t_3\mu_{\bar{x}}\right) = 0{,}05,$$

т.е. $$0{,}05 = 1 - 2\cdot\frac{1}{\sqrt{2\pi}}\int_0^{t} e^{-\frac{t^2}{2}}\,dt .$$

Тогда $$\frac{1}{\sqrt{2\pi}}\int_0^{t} e^{-\frac{t^2}{2}}\,dt = \frac{1 - 0{,}05}{2} = 0{,}475 .$$

По таблице нормированной функции Лапласа находим, что $t_3 = 1{,}96$. Таким образом, $\left[P\left|x - \bar{x}\right| > 1{,}96\,\mu_{\bar{x}}\right] = 0{,}05$.

Рис. 6.4. **Двусторонняя критическая область при $\alpha = 0{,}05$.**

4) область малых по абсолютной величине отклонений:

$$P_4 = P\left|x - \overline{x}\right| < t_4 \mu_{\overline{x}} = 2\frac{1}{\sqrt{2\pi}}\int_0^t e^{-\frac{t^2}{2}}\,dt = 0{,}05.$$

Тогда $\frac{1}{\sqrt{2\pi}}\int_0^t e^{-\frac{t^2}{2}}\,dt = 0{,}025$.

Величина t, соответствующая вероятности 0,025, равна 0,063 (см. Приложение II).

6.6. Проверка гипотезы о принадлежности «выделяющихся» наблюдений исследуемой генеральной совокупности

В составе собранных данных могут встречаться единичные наблюдения, у которых зарегистрированные значения признака заметно отличаются от общего уровня. Происхождение таких отличий может быть различным. Они могут возникать в результате:

1) ошибок наблюдения; 2) случайного стечения различного рода в отдельности несущественных обстоятельств и 3) нарушения однородности изучаемой совокупности.

Субъективное отбрасывание вариантов ряда, даже резко отличающихся от средней арифметической, не имеет никакого принципиального оправдания, поскольку первичные данные соответствуют конкретным случаям массового явления. Для исключения из дальнейшей обработки выделяющихся наблюдений необходимо применение надлежащим образом обоснованных критериев.

Допустим, что распределение результатов наблюдений, проводимых в обычных условиях, следует нормальному закону с параметрами $\overline{x}$ и σ. При проведении одной из серий таких наблюдений мы получим n значений $x_1 \leqslant x_2 \leqslant x_3 \leqslant \ldots \leqslant x_n$, среди которых максимальное значение x_n (или минимальное x_1) резко отличается по своей величине от остальных $(n-1)$ наблюдений. Возникает вопрос, относятся ли эти значения к данной наблюдаемой в определенных условиях генеральной совокупности или они возникли как результат каких-то экстраординарных обстоятельств. Нулевой гипотезой в этом случае является предположение о том, что

$x_n(x_1)$ принадлежит той же совокупности, что и все остальные $n-1$ наблюдений, т.е. x_n (или x_1) не является результатом ошибки наблюдения или изменения общих условий формирования уровня признаков совокупности.

Проверка этой гипотезы состоит в том, что $x_n(x_1)$ сравнивается по величине с некоторой критической границей x. Если выделяющимся наблюдением является наибольшее, то x_n сравнивается с верхней допустимой границей, выбранной таким образом, чтобы вероятность превзойти ее была равна уровню значимости, т. е. в данном случае мы имеем дело с критической областью вида

$$P\left(x_n > \overline{x} + t\sigma\right) = \alpha.$$

Гипотеза H_0 бракуется, если x_n превосходит по величине указанную границу (см. рис. 6.4).

Если же выделяющимся наблюдением является наименьшее (x_1), то его сравнивают с нижней допустимой границей, которую принимают равной $\overline{x} - t\sigma$, т.е. $P(x_1 < \overline{x} - t\sigma) = \alpha$.

Если же испытанию одновременно подлежат и максимальное и минимальное значения признака, то критическая область будет иметь вид $P(|x_i - \overline{x}| > t\sigma) = \alpha$ (см. рис. 6.4).

Например, известны следующие данные, приведенные в табл. 6.7.

Таблица 6.7

Количество наблюдений	Минимальные значения		Максимальные значения		Разность смежных значений		Среднее значение	Среднее квадратическое отклонение по выборке
	x_1	x_2	x_{n-1}	x_n	x_2-x_1	x_n-x_{n-1}	$\overline{x}$	σ
84	7	13	130	178	6	48	57,3	35,3

Используем критическую область вида $P(x_n > \overline{x} + t\sigma)$, так как минимальное значение x_1 незначительно отличается от следующего за ним значения в ранжированном ряду, т.е. проверим, принадлежит ли выделяющееся наблюдение $x_n = 178$ к рассматриваемой совокупности результатов.

При уровне значимости $\alpha = 0{,}01$ значение нормированной функции Лапласа для рассматриваемой критической области будет равно

$$\frac{1}{\sqrt{2\pi}}\int_0^t e^{-\frac{t^2}{2}}dt = \frac{1}{2} - 0{,}01 = 0{,}49$$

Этому значению в таблице нормированной функции Лапласа соответствует $t = 2{,}33$. Тогда верхняя допустимая граница значений признака, которая не может быть превышена с вероятностью 0,99, будет равна 139,6 $(57{,}3 + 2{,}33 \cdot 35{,}3 = 139{,}6)$. Значение $x_n = 178$ выходит за рассчитанную границу, потому с вероятностью 0,99 можно считать $x_n = 178$, не принадлежащим к изучаемой совокупности и следует исключить его из дальнейших расчетов.

Часто могут иметь место такие случаи, когда параметры генеральной совокупности $\bar{x}$ и σ неизвестны, и для проверки гипотезы о выделяющихся наблюдениях используются полученные по выборке оценки соответствующих параметров. Однако следует учитывать, особенно для малых выборок, что эти оценки являются не вполне надежными. В этой связи для отбрасывания выделяющихся наблюдений по данным малых выборок можно рекомендовать пользоваться критерием Ф. Груббса, предложенным в 1950 г.

Критерий Ф. Груббса основан на отношении двух сумм квадратов отклонений:

1) для испытания наибольшего наблюдения, являющегося выделяющимся в выборке объема n из нормально распределенной совокупности, рассчитывается отношение

$$\frac{S_n^2}{S^2} = \frac{\sum_{i=1}^{n-1}\left(x_i - \bar{x}_n\right)^2}{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2} \leqslant K, \qquad (6.12)$$

где $x_1 \leqslant x_2 \leqslant x_3 \leqslant \ldots \leqslant x_n$; $\bar{x}_n = \frac{\sum_{i=1}^{n-1} x_i}{n-1}$; $\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$;

2) для испытания наименьшего наблюдения, выделяющегося в выборке объема n, рассчитывается отношение

$$\frac{S_1^2}{S^2} = \frac{\sum_{i=2}^{n}\left(x_i - \bar{x}_1\right)^2}{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2} \leqslant K, \qquad (6.12а)$$

где $\bar{x}_1 = \frac{\sum_{i=2}^{n} x_i}{n-1}$.

Рассчитанная величина отношения ($K_{расч.}$) сравнивается с табличной величиной ($K_{табл.}$) при определенном числе наблюдений и заданном уровне значимости. $K_{табл.}$ характеризует ту предельную величину расхождений в суммах квадратов отклонений, которая с вероятностью $(1-\alpha)$ может быть объяснена случайными причинами.

Если $K_{расч.}$ будет равно или меньше табличного, то наименьшее или наибольшее наблюдения не отбрасываются. Если же $K_{расч.} > K_{табл.}$, то вероятность того, что расхождения в суммах квадратов отклонений можно объяснить случайными причинами, равна уровню значимости α и в силу малой вероятности считается событием практически невозможным. В таком случае выделяющееся наблюдение следует отбросить и для дальнейших расчетов использовать оставшиеся $(n-1)$ наблюдений.

Покажем применение критерия Ф. Груббса на следующем примере. Отклонения деталей от номинального размера оказались такими (мм): 0,07; 0,09; 0,10; 0,12; 0,13; 0,15; 0,16; 0,17; 0,25.

Определим, не содержат ли результаты ошибки наблюдения, предполагая, что распределение размеров в генеральной совокупности следует нормальному закону.

Средняя арифметическая по всем отклонениям $\bar{x}$ равна 0,138:

$$\bar{x} = \frac{0{,}06+0{,}09+0{,}10+0{,}12+0{,}13+0{,}15+0{,}16+0{,}17+0{,}25}{9}.$$

Исключая максимальное отклонение 0,25, получим

$$\bar{x}_n = 0{,}124 \left(\bar{x}_n = \frac{0{,}07+0{,}09+0{,}10+0{,}12+0{,}13+0{,}15+0{,}16+0{,}17}{8}\right).$$

Сумма квадратов отклонений от $\bar{x}$ будет равна 0,0231:
$[(0{,}07-0{,}138)^2+(0{,}09-0{,}138)^2+\ldots+(0{,}17-0{,}138)^2+(0{,}25-0{,}138)^2]$.

Сумма квадратов отклонений $(n-1)$ значений от $\bar{x}_n$ равна 0,0088: $[(0{,}07-0{,}124)^2 + (0{,}09-0{,}124)^2 + \ldots + (0{,}16-0{,}124)^2 + (0{,}17-0{,}124)^2]$.

Отношение двух сумм квадратов отклонений $K_{расч.} = \frac{0{,}0088}{0{,}0231} = 0{,}381$.

Таблица 6.8

Таблица Ф. Груббса (выдержка)

Число наблюдений n	Уровень значимости (α)	
	0,01	0,05
3	0,0001	0,0027
4	0,0100	0,0494
5	0,0442	0,1270
6	0,0928	0,2032
7	0,1447	0,2696
8	0,1948	0,3261
9	0,2411	0,3742
10	0,2831	0,4154
⋮		
15	0,4401	0,5559
⋮		
20	0,5393	0,6379
⋮		
25	0,6071	0,6923

При числе наблюдений n = 9 и уровне значимости 0,01 $K_{\text{табл.}}$ равно 0,2411, т.е. $K_{\text{расч.}} > K_{\text{табл.}}$ (см. табл. 6.8). Полученное расчетное значение больше табличного и при уровне значимости 0,05 ($K_{\text{табл.}}$ = 0,3742), а потому отклонение от номинального размера 0,25 мм следует отнести к ошибкам наблюдения.

Рассмотренные выше критерии для отбрасывания выделяющихся наблюдений базировались на предположении нормального распределения признака в генеральной совокупности. На такого же рода допущениях о нормальности распределения признака основана и оценка существенности коэффициента корреляции и т.д.

Поэтому одним из этапов статистической обработки результатов наблюдений является проверка гипотезы о возможности отнесения эмпирического распределения, полученного по данным выборки, к типу кривых нормального распределения (см. гл. 5).

Еще один вариант проверки гипотезы о принадлежности выделяющихся наблюдений к генеральной совокупности исследуемых данных был предложен Дж. Ирвином в 1925 г. Критерий Ирвина λ основан на соотношении разности между x_n и x_{n-1} к среднему квадратическому отклонению в генеральной совокупности.

$$\lambda = \frac{x_n - x_{n-1}}{\sigma}, \qquad (6.13)$$

где $x_1 \leqslant x_2 \leqslant x_3 \leqslant \ldots \leqslant x_{n-1} \leqslant x_n$;

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2}{n-1}}.$$

В нижеприводимой выдержке из таблицы Дж. Ирвина приводится вероятность того, что разность между x_n и x_{n-1} отличается от среднего квадратического отклонения в случайной выборке объемом n единиц более, чем в λ раз.

Таблица 6.9

Значения $P(\lambda)$, где λ критерий Ирвина

Объем выборки	Значения вероятности при λ, равном										
n	1,0	1,1	1,2	1,3	1,4	1,5	1,6	1,7	1,8	1,9	2,0
10	0,152	0,121	0,096	0,075	0,059	0,045	0,038	0,026	0,020	0,015	0,011
20	0,107	0,082	0,062	0,047	0,035	0,026	0,019	0,014	0,010	0,007	0,005
30	0,089	0,068	0,050	0,037	0,027	0,020	0,014	0,010	0,007	0,005	0,004
⋮											
60	0,065	0,048	0,034	0,025	0,017	0,012	0,009	0,006	0,004	0,003	0,002
70	0,061	0,044	0,032	0,022	0,016	0,011	0,008	0,005	0,004	0,002	0,002
80	0,058	0,041	0,030	0,021	0,015	0,010	0,007	0,005	0,003	0,002	0,001

Используя данные примера, приведенного в табл. 6.7, рассчитаем критерий Ирвина λ, который составит 1,36 ($\lambda = \frac{48}{35{,}3} = 1{,}36$). По данным табл. 6.9, при n=80 и λ = 1,4 $P(\lambda)$ равно 0,015. В силу малой вероятности появления таких расхождений между смежными значениями в ранжированном ряду полученных данных можно считать значение x_n = 178, не принадлежащим к изучаемой совокупности, и при дальнейшей обработке информации использовать оставшиеся 83 значения.

6.7. Проверка гипотезы о величине средней арифметической и доли

Известно, например, что средний расход сырья на единицу продукции при существующем методе производства составляет 2,8 условных единиц. После внесения изменений в существующую технологию изготовления продукции по результатам проверки достаточно большой партии изделий средний расход сырья на единицу продукции составил 2,6 условные единицы. Средняя ошибка выборки оказалась равной 0,1. Возникает вопрос, действительно ли применение нового метода обработки приводит к снижению материалоемкости продукции?

Нулевая гипотеза состоит в том, что между новым и существующим методами производства изделий отсутствуют существенные различия с точки зрения влияния их на материалоемкость, т.е. что между генеральными средними при старом и новом методах производства нет существенной разницы, а отклонение выборочной средней от достигнутого уровня при существующем методе обусловлено только случайностями выборки, т.е. H_0 означает, что $\bar{x}_с = \bar{x}_н$, где $\bar{x}_с$ и $\bar{x}_н$ - средний расход сырья на единицу продукции соответственно при существующем и новом методах производства.

Альтернативная гипотеза может быть сформулирована двояко:

1. Применение нового метода обработки приводит к изменению расхода сырья на единицу продукции, т. е. H_α состоит в том, что $\bar{x}_н \neq \bar{x}_с$. Примем уровень значимости α равным 0,05, тогда $P(|\bar{x}_н - \bar{x}_с| \geqslant t\mu_{\bar{x}}) = 0{,}05$ и критическая область соответственно задается неравенством $|\bar{x}_н - \bar{x}_с| > t\mu_{\bar{x}}$. По таблицам интегральной функции Лапласа определяем коэффициент доверия $t = 1{,}96$. Таким образом, величина предельного расхождения двух средних с вероятностью, равной 0,95, не должна превышать $t\mu_{\bar{x}} = 0{,}196(1{,}96 \cdot 0{,}1)$. Следовательно, с вероятностью 0,95 доверительные пределы для генеральной средней при новом методе будут равны $2{,}64 \leqslant \bar{x}_н \leqslant 2{,}996$ ($\bar{x}_с \pm \Delta_{\bar{x}}$).

Средний расход материала при применении новой технологии составляет 2,6, т. е. попадает в критическую область. Следовательно, данные наблюдения не являются совместимыми с выдвинутой гипотезой H_0 о том, что между новым и существующим методами производства изделий отсутствуют существенные различия с точки зрения влияния их на материалоемкость.

2. Применение нового метода обработки приводит к снижению расхода сырья на единицу продукции, т.е. $H\alpha$ состоит в том, что $\bar{x}_н < \bar{x}_с$. В этом случае рассматривается область больших отрицательных отклонений, т.е. при $\lambda=0{,}05$ $P(\bar{x}_н \leq \bar{x}_с - t\mu_{\bar{x}}) = 0{,}05$. В данном варианте критическая область определяется неравенством $\bar{x}_н < \bar{x}_с - t\mu_{\bar{x}}$. Нулевая гипотеза не будет опровергаться, если средний расход материла на единицу продукции будет больше величины 2,636 (2,8 - 0,1 • 1,64). Так как по новой технологии расход сырья составляет 2,6 условных единиц, то с вероятностью 0,995 можно считать, что нулевая гипотеза должна быть отвергнута и что, следовательно, применение новой технологии приводит к снижению расхода сырья на изготовление продукции.

Если гипотеза о величине центра распределения проверяется по результатам малой выборки, то следует учитывать, что отношение разности средних к стандартной ошибке выборки, т. е. величина $t = \frac{\bar{x}_1 - \bar{x}_2}{\mu_{\bar{x}}}$ имеет распределение Стьюдента с n-1 степенями свободы.

Допустим, например, что данные о расходе сырья на единицу продукции были получены по результатам проверки 26 изделий. Используем ранее приведенные данные о среднем расходе сырья при старом способе обработки ($\bar{x}_с$=2,8 усл. ед.) и при новом способе обработки ($\bar{x}_н$=2,6 усл. ед.). По результатам наблюдений значение эмпирического среднего квадратического отклонения составило 0,6. Расчетная величина t - критерия определится по формуле:

$$t_{расч.} = \frac{|\bar{x}_н - \bar{x}_с|}{\frac{S}{\sqrt{n-1}}} = \frac{|2{,}6-2{,}8| \cdot 5}{0{,}6} = 1{,}67.$$

По таблице распределения Стьюдента значение t-критерия для числа степеней свободы ($n-1$), равного 25, и уровня значимости 0,05 составляет 2,06. Так как расчетное значение не превышает табличное, то нулевая гипотеза не опровергается, т.е. у нас нет достаточных оснований считать, что применение новой технологии приводит к снижению материалоемкости.

На следующем примере рассмотрим проверку гипотезы о существенности различия двух выборочных средних (для случая малой выборки).

По результатам оценки размеров дебиторской задолженности, проведенной финансовым директором фирмы, были внесены изменения в ее кредитную политику. По истечении отчетного квартала было решено проанализировать, действительно ли изменения в кредитной политике фирмы оказали влияние на сокращение срока оплаты дебиторской задолженности.

В таблице представлены данные о продолжительности сбора дебиторской задолженности (в днях) при старой и новой кредитной политике фирмы.

Тип кредитной политики фирмы	Продолжительность сбора дебиторской задолженности, дни									
«Старая»	39	42	28	35	38	30	30	29	36	34
«Новая»	28	31	29	30	24	37	27	22	33	31

Положим, что фирма ставит задачу проверить при уровне значимости 5%, привело ли изменение ее кредитной политики к изменению продолжительности сбора дебиторской задолженности.

В данном случае нулевая гипотеза состоит в отсутствии различий в величине выборочных средних, т.е. $H_0 : \tilde{x}_1 = \tilde{x}_2$. Тогда альтернативная гипотеза $H_a : \tilde{x}_1 \neq \tilde{x}_2$. Уровень значимости для проверки выдвинутой гипотезы $\alpha = 0{,}05$. Ответ на поставленный вопрос требует сопоставления разности двух выборочных средних с величиной средней квадратической ошибки этих средних, т.е. должна быть рассчитана по фактическим данным двух выборок величина t:

$$t_{\text{расч.}} = \frac{\tilde{x}_1 - \tilde{x}_2}{\mu_{(\tilde{x}_1 - \tilde{x}_2)}} \qquad (6.14)$$

При выдвинутой выше нулевой гипотезе предполагается, что обе выборки принадлежат к одной и той же генеральной совокупности со средней $\bar{x}$ и дисперсией σ^2.

Поскольку мы располагаем данными двух выборок и величиной двух дисперсий S_1^2 и S_2^2, оценку генеральной дисперсии следует определить как среднюю арифметическую взвешенную из выборочных дисперсий, а в качестве веса использовать число степеней свободы в каждой выборке, т.е. $\sigma^2 = \dfrac{\sum_{j=1}^{2} S_j^2(n_j - 1)}{\sum_{j=1}^{2}(n_j - 1)}$ или

$$\sigma^2 = \frac{S_1^2(n_1 - 1) + S_2^2(n_2 - 1)}{n_1 + n_2 - 2},$$

а поскольку $S_1^2 = \dfrac{\sum_{i=1}^{n_1}(x_{ij} - \tilde{x}_1)^2}{n_1 - 1}$ и $S_2^2 = \dfrac{\sum_{i=1}^{n_2}(x_{ij} - \tilde{x}_2)^2}{n_2 - 1}$, то

$$\sigma^2 = \frac{\sum_{i=1}^{n_1}(x_{i1} - \tilde{x}_1)^2 + \sum_{i=1}^{n_2}(x_{i2} - \tilde{x}_2)^2}{n_1 + n_2 - 2}$$

и формула стандартной ошибки разности двух выборочных средних имеет вид:

$$\mu_{\tilde{x}_1 - \tilde{x}_2} = \sqrt{\frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2}} = \sqrt{\sigma^2 \cdot \frac{n_1 + n_2}{n_1 \cdot n_2}}, \qquad (6.15)$$

где σ^2 - оценка генеральной дисперсии по данным двух выборок;
n_1 и n_2 - количество наблюдений соответственно в первой и во второй выборках.

Результаты расчетов для проверки выдвинутой гипотезы об отсутствии влияния изменений в кредитной политике фирмы на продолжительность сбора дебиторской задолженности приведены в табл. 6.10.

Таблица 6.10

Тип кредитной политики фирмы	Средняя продолжительность сбора дебиторской задолженности, дни	Число обследованных потребителей n_j	Сумма квадратов отклонений $\sum_{i=1}^{n_j}(x_{ij} - \tilde{x}_j)^2$
«Старая»	34,1	10	187,69
«Новая»	29,2	10	165,80
Итого		20	353,49

В предположении отсутствия существенных различий в продолжительности сбора дебиторской задолженности, т.е. верности нулевой гипотезы, величина $t_{расч.}$ не должна превышать табличного значения t при данном уровне значимости и числе степеней свободы (n_1+n_2-2). По таблице распределения Стьюдента (см. Приложение IV) при уровне значимости α=5%, и числе степеней свободы 18 (10+10-2) находим, что $t_{табл.}$ = 2,101. Вероятность получения по результатам расчетов значения $t_{расч.}$, большего по абсолютной величине, чем 2,101, в условиях случайных расхождений выборочных характеристик составляет всего 0,05 и является событием практически невозможным.

Используя данные табл. 6.10, рассчитаем среднюю квадратическую ошибку разности двух выборочных средних:

$$\mu_{\tilde{x}_1-\tilde{x}_2}=\sqrt{\frac{353{,}49}{10+10-2}\cdot\frac{10+10}{10\cdot 10}}=1{,}98 \text{ дня},$$

Тогда

$$t_{расч.}=\frac{\tilde{x}_1-\tilde{x}_2}{\mu_{\tilde{x}_1-\tilde{x}_2}}=\frac{34{,}1-29{,}2}{1{,}98}=2{,}475\ .$$

Поскольку $t_{расч.}$ = 2,475 оказывается больше табличного при уровне значимости 5% ($t_{табл.}$ = 2,101), нулевая гипотеза не подтверждается, т.е. различие между средней продолжительностью сбора дебиторской задолженности при старой и новой кредитной политике фирмы нельзя объяснить случайностями выборки.

Можно получить ответ на выдвинутую нулевую гипотезу иначе. Зная величину средней квадратической ошибки разности двух выборочных средних, можно с заданной вероятностью указать предел возможных расхождений двух выборочных средних $\Delta_{\tilde{x}_1-\tilde{x}_2}=t\mu_{\tilde{x}_1-\tilde{x}_2}$, где t - табличное значение критерия Стьюдента.

В рассматриваемом примере $\Delta_{\tilde{x}_1-\tilde{x}_2}=2{,}101\cdot 1{,}98=4{,}16$ дня. Поскольку фактическая разность двух средних составляет 4,90 дня (34,1-29,2), нулевая гипотеза не подтверждается. Таким образом, можно предполагать, что изменение кредитной политики фирмы дало определенный результат и привело к снижению продолжительности сбора дебиторской задолженности.

Как проверить гипотезу о существенности расхождений двух выборочных долей покажем на следующем примере. По результатам выборочного обследования домохозяйств двух областей Центрально-Черноземного района России в 1992 г., были получены данные о доле доходов от предпринимательской деятельности в источниках доходов домохозяйств.

Таблица 6.11

Номер области	Количество обследованных домохозяйств, тыс. n_i	Удельный вес доходов от предпринимательской деятельности, % w_i
1	14,2	6,9
2	12,5	7,4

Разность двух выборочных долей составляет 0,5% (7,4-6,9). Можно ли считать несущественными различия в доле доходов от предпринимательской деятельности в источниках доходов домохозяйств двух областей? Нулевая гипотеза заключается в том, что отсутствуют существенные различия в роли новых видов доходов, традиционных для рыночной экономики.

Тогда в качестве оценки генеральной доли будет использоваться средняя взвешенная из долей, полученных по результатам выборочных обследований каждой области:

$$P=\frac{\sum_{i=1}^{2} w_i n_i}{\sum_{i=1}^{2} n_i}=\frac{0{,}069\cdot 14{,}2+0{,}074\cdot 12{,}5}{14{,}2+12{,}5}=0{,}0713\ ,$$

т.е. оценка доли дохода от предпринимательской деятельности в генеральной совокупности составляет 7,13%. Ошибка разности двух долей при справедливости нулевой гипотезы будет рассчитываться по формуле:

$$\mu_{w_1-w_2}=\sqrt{\frac{P(1-P)}{n_1}+\frac{P(1-P)}{n_2}}=\sqrt{P(1-P)\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}=$$

$$=\sqrt{0{,}0713\times 0{,}9287\left(\frac{1}{14\,200}+\frac{1}{12\,500}\right)}=0{,}00316\ .$$

Таким образом, средняя квадратическая ошибка разности двух выборочных долей составляет 0,316%.

Поскольку обе выборки достаточно большого объема, можно воспользоваться таблицей нормированной функции Лапласа (см. Приложение II) для определения значения коэффициента доверия t при вероятности 0,95 или 0,99. Соответствующие этим вероятностям значения t равны 1,96 и 2,58.

Расчетное значение t-критерия, равное отношению разности двух выборочных долей, составит:

$$t_{расч.} = \frac{|10,069 - 0,0741|}{0,00316} = 0,631 .$$

Поскольку расчетное значение t меньше табличного как при пятипроцентном уровне значимости, так и при уровне значимости 1%, мы не отвергаем нулевую гипотезу и делаем заключение о несущественности различий доли доходов от предпринимательской деятельности в источниках доходов домохозяйств двух областей.

Можно было определить и максимально возможную величину расхождений двух выборочных долей с заданной вероятностью

$$\Delta_{w_1-w_2} = t\mu_{w_1-w_2} .$$

При вероятности 0,95: $\Delta_{w_1-w_2} = 0,00619$ (1,96•0,00316).

При вероятности 0,99: $\Delta_{w_1-w_2} = 0,00815$ (2,58•0,00316).

Поскольку фактическая разность двух выборочных долей 0,005 меньше найденных предельных ошибок, можно принять нулевую гипотезу.

Таким образом, оба варианта проверки гипотезы о несущественности различий доли доходов от предпринимательской деятельности по результатам выборочных обследований в двух областях дали один и тот же результат.

6.8. Элементы дисперсионного анализа

Вопрос о проверке существенности расхождения двух выборочных характеристик может быть поставлен не только при сравнении двух выборочных средних, но и при сравнении двух выборочных дисперсий. Для сравнения дисперсий применяется критерий, предложенный Рональдом Фишером, который называют дисперсионным отношением, или F-критерием.

Критерий Фишера представляет собой отношение двух дисперсий:

$$F = \frac{S_1^2}{S_2^2}, \tag{6.16}$$

где S_1^2 и S_2^2 рассматриваются в качестве оценок одной и той же генеральной дисперсии.

При вычислении дисперсионного отношения в числителе берется большая из оценок S_1^2 и S_2^2, поэтому величина дисперсионного отношения может быть равна или больше единицы. Если F-критерий равен 1, то это указывает на равенство дисперсий, и вопрос об оценке существенности их расхождения снимается. Если же величина дисперсионного отношения больше единицы, то возникает необходимость оценить случайно ли расхождение между дисперсиями. При этом очевидно, что чем больше величина дисперсионного отношения, тем значительнее расхождение между дисперсиями.

Для определения границ случайных колебаний отношения дисперсий Р.Фишером разработаны специальные таблицы F-распределения (см. Приложение VI). В этих таблицах указываются предельные значения F-критерия для различных комбинаций числа степеней свободы числителя k_1 и знаменателя k_2 [1], которые могут быть превзойдены с вероятностью 0,05 или 0,01. Число степеней свободы k_1, соответствующее большей дисперсии, определяет столбец таблицы, а число степеней свободы k_2, соответствующее дисперсии S_2^2, строку таблицы.

Рассчитанная по фактическим данным величина дисперсионного отношения сопоставляется с соответствующей данному сочетанию числа степеней свободы числителя и знаменателя и принятому уровню значимости табличной величиной дисперсионного отношения.

Гипотеза, которая проверяется с помощью этих таблиц, состоит в том, что сравниваемые дисперсии характеризуют вариацию признака в совокупностях, отобранных из одной и той же нормально распределенной генеральной совокупности, или же отобранных из нормально распределенных генеральных совокупностей с одинаковой дисперсией.

Если фактическое дисперсионное отношение будет больше табличного, то лишь с вероятностью 0,05 или 0,01 можно утверждать, что различие между дисперсиями определяется случайными факторами. Однако события, имеющие столь малую вероят-

[1]Понятие числа степеней свободы связано с тем, что в статистических совокупностях приходится учитывать линейные связи, ограничивающие свободу изменения случайных величин. Например, при исчислении дисперсии в совокупности мы располагаем n-1 степенями свободы, так как любое одно недостающее значение признака мы может определить, зная n-1 значений и среднюю арифметическую.

ность, считаются практически невозможными, а потому в этом случае с вероятностью 1-α можно утверждать существенность различий в величине дисперсий.

Если же фактическое значение дисперсионного отношения будет меньше соответствующего табличного значения, например, при 1%-ном уровне значимости, то с вероятностью 99% можно утверждать, что расхождение между дисперсиями несущественно.

Дисперсионный анализ приобретает самостоятельное значение при оценке существенности расхождения нескольких средних, что позволяет проверить гипотезу о наличии связи между признаком, положенным в основу группировки, и результативным признаком. В зависимости от количества факторов, определяющих вариацию результативного признака, дисперсионный анализ подразделяется на *однофакторный* и *многофакторный*. Методы дисперсионного анализа позволяют также проверить гипотезу относительно формы корреляционной зависимости и оценить целесообразность включения в модель дополнительных факторов (см. подробнее гл. 7).

Рассмотрим применение дисперсионного анализа для случая **однофакторного** комплекса.

Пусть все n наблюдений разбиты на k групп в соответствии с определенным признаком и число наблюдений в j-й группе равна n_j. Систему таких наблюдений в общем виде можно записать таким образом:

Номер группы	Значения результативного признака
1	y_{11} y_{12} y_{13} ... y_{1i} ... y_{1n1}
2	y_{21} y_{22} y_{23} ... y_{2i} ... y_{2n2}
.	
.	
j	y_{j1} y_{j2} y_{j3} ... y_{ji} ... y_{jnj}
.	
.	
k	y_{k1} y_{k2} y_{k3} ... y_{ki} ... y_{knk}

Обозначим общую среднюю для всей совокупности $\bar{y}$, а среднюю в соответствующей группе $\bar{y}_j$. Каждое индивидуальное отклонение y_{ij} от общей средней складывается из двух частей; отклонения от средней в соответствующей группе, т.е. величины $y_{ij}-\bar{y}_j$, и отклонения средней в группе $\bar{y}_j$ от общей средней, т.е. $\bar{y}_j-\bar{y}$, иначе $y_{ij}-\bar{y} = (y_{ij}-\bar{y}_j)+(\bar{y}_j-\bar{y})$. Тогда сумма квадратов отклонений всех наблюдаемых значений от общей средней будет равна

$$\sum_{j=1}^{k}\sum_{i=1}^{n_j}\left(y_{ij}-\bar{y}\right)^2=\sum_{j=1}^{k}\sum_{i=1}^{n_j}\left(y_{ij}-\bar{y}_j\right)^2+\sum_{j=1}^{k}\left(\bar{y}_j-\bar{y}\right)^2 n_j, \qquad (6.17)$$

где первое слагаемое в правой части представляет собой сумму квадратов отклонений наблюдаемых значений от групповых средних и характеризует вариацию внутри групп, а второе слагаемое - сумма квадратов отклонений групповых средних от общей средней – характеризует вариацию между группами.

Колебания изучаемого признака внутри группы вокруг средней возникают под влиянием прочих причин, исключая влияние фактора, положенного в основу группировки. Колеблемость групповых средних вокруг общей средней обусловлена влиянием признака-фактора. Если фактор, положенный в основу группировки, не оказывает влияния на вариацию изучаемого признака, то дисперсия групповых средних будет отражать только влияние тех же самых прочих факторов, которые определяют и вариацию внутри групп, а потому отношение дисперсий будет близко к единице или отличаться от нее в силу наличия случайных колебаний. Предельный размер этих колебаний можно установить по таблицам F-распределения.

При применении дисперсионного анализа для расчета дисперсий учитывается число степеней свободы. Для расчета внутригрупповой дисперсии число степеней свободы равно $n-k$ [1], а для расчета межгрупповой дисперсии равно $k-1$ [2].

Тогда межгрупповая s_1^2 и внутригрупповая дисперсия s_2^2 с учетом числа степеней свободы будут соответственно равны:

$$s_1^2=\frac{\sum_{j=1}^{k}\left(\bar{y}_j-\bar{y}\right)^2 n_j}{k-1}, \qquad (6.18)$$

$$s_2^2=\frac{\sum_{j=1}^{k}\sum_{i=1}^{n_j}\left(y_{ij}-\bar{y}_j\right)^2}{n-k}. \qquad (6.18а)$$

[1] В каждой группе при определении дисперсии мы располагаем n_j-1 степенями свободы, так как любое отклонение можем определить, зная n_j-1 отклонений и $\bar{x}_j$. Поскольку мы использовали k средних, а сумма n_j=n, число степеней свободы для расчета внутригрупповой дисперсии равно n-k.

[2] При расчете межгрупповой дисперсии мы рассматриваем k отклонений групповых средних от общей средней. Зная общую среднюю и k-1 групповых средних, можно определить любое из недостающих отклонений.

Если группировочный признак оказывает влияние на вариацию результативного признака, то вариацию групповых средних нельзя считать обусловленной только случайными воздействиями и это найдет отражение в различии величины межгрупповой и внутригрупповой дисперсии, т.е. $F_{расч.}$ будет больше единицы. Если при этом $F_{расч.} > F_{табл.}$, то с вероятностью 0,95 (0,99) можно утверждать, что между факторным и результативным признаками существует взаимосвязь.

Рассмотрим дисперсионный анализ на следующем примере.

За месяц известны данные о выработке рабочего за время работы в первую и во вторую смены (см. табл. 6.12).

Таблица 6.12

Смена	Выработка рабочего, нормо-час
I	12,1; 11,1; 12,6; 12,9; 11,6; 13,1; 12,6; 12,4; 11,6; 17,3; 12,9; 11,6; 12,4
II	9,9; 11,4; 13,4; 10,4; 12,9; 12,6; 13,9; 13,4; 12,4; 9,9

Можно ли считать, что расхождение между уровнями выработки рабочего в первую и во вторую смены несущественно, т.е. можно ли считать, что генеральные средние в двух подгруппах одинаковы и, следовательно, выработка рабочего может быть охарактеризована общей средней. Для того чтобы ответить на поставленный вопрос, рассчитаем среднюю выработку рабочих в каждой смене (см. графу 2 табл. 6.13). Величина средней выработки в первую и вторую смены различна. Теперь возникает вопрос о том, насколько существенны эти расхождения, т.е. нужно проверить предположение о возможном влиянии сменности на выработку рабочих. Используя данные графы 4 и графы 5, рассчитаем S_1^2 и S_2^2.

Таблица 6.13

Смена	Средняя выработка, нормо-часы $\overline{y}_j$	Число смен в месяце n_j	Сумма квадратов отклонений вариантов от групповой средней $\sum_{i=1}^{n_j}(y_{ij}-\overline{y}_j)^2$	Квадраты отклонений групповых средних от общей средней $(\overline{y}_j-\overline{y})^2$
1	2	3	4	5
I	12,63	13	28,0877	0,0676
II	12,02	10	20,5960	0,1225
Итого	$\overline{y}=12,37$	23	48,6837	-

Число степеней свободы для расчета внутригрупповой дисперсии равно 21(23-2), а для расчета межгрупповой дисперсии 1(2-1).

Следовательно,

$$S_1^2=\frac{0,0676\cdot 13+0,1225\cdot 10}{1}=2,1038;\quad S_2^2=\frac{48,6837}{21}=2,3182$$

Дисперсионное отношение $F_{расч.}=\frac{2,1038}{2,3182}=0,9075$.

В соответствии с числом степеней свободы числителя (1) и знаменателя (21) в таблице F-распределения для 5%-ного уровня значимости находим $F_{табл.}=4,32$, а для 1%-ного уровня значимости $F_{табл.}=8,02$.

Так как $F_{расч.}$ значительно меньше табличных значений, гипотеза о несущественности различия выработки рабочего в первую и во вторую смены не опровергается, т.е. сменность не оказывает влияния на уровень выработки рабочего.

6.9. Различные формы организации выборочного наблюдения

Различные формы организации выборочного наблюдения - типическая (районированная или стратифицированная) выборка, серийная (гнездовая) выборка, механическая выборка, комбинированная выборка - представляют собой дальнейшее развитие и видоизменение простой случайной выборки. Применение этих форм вызывается соображениями удешевления или облегчения процесса наблюдения, особым характером объекта наблюдения, или отсутствием необходимой информации для составления списка единиц наблюдения.

Начнем рассмотрение видов выборки с организации **типической (стратифицированной) выборки**. Как правило, социально-экономические явления характеризуются большим разнообразием и не являются достаточно однородными в отношении изучаемых признаков. При наличии в составе генеральной совокупности различных типов явления с разными уровнями изучаемых признаков желательно так организовать выборку, чтобы обеспечить более равномерное представительство в выборочной совокупности различных частей (типов) явления.

Для этого общий список единиц генеральной совокупности в целом предварительно разбивается на отдельные списки, каждый из которых включает единицы, принадлежащие к одной однородной по определенному признаку группе (типу). В качестве типов могут быть использованы группы, сложившиеся в практике статистики. Например, при обследовании семейных бюджетов рабочих и служащих на первом этапе выделяют группы рабочих и служащих в отдельных отраслях экономики. Затем уже при отборе работников на предприятии – квалифицированных и малоквалифицированных рабочих. Другими словами, типическая выборка опирается на статистическую группировку единиц генеральной совокупности по одному признаку или по комбинации нескольких признаков. Из каждой выделенной группы в случайном порядке отбирается некоторое количество единиц.

Таким образом, при проведении типической выборки необходимо разбить общий объем выборки n между группами и определить число подлежащих наблюдению единиц в каждой группе.

Наиболее часто применяется так называемое пропорциональное размещение единиц, когда количество отбираемых в выборку единиц пропорционально удельному весу данной группы по числу единиц в генеральной совокупности, т.е. число наблюдений по каждой группе определится по формуле

$$n_i = n\frac{N_i}{N}, \tag{6.19}$$

где n_i - число наблюдений из i-й типической группы;
n - общий объем выборки;
N_i - объем i-й типической группы в генеральной совокупности;
N - объем генеральной совокупности.

Возможен и другой вариант (назовем его условно равномерным распределением), когда из каждой группы отбирают одинаковое число единиц, т.е.

$$n_i = \frac{n}{k}, \tag{6.19а}$$

где k - число выделенных типических групп.

Третий вариант решения (назовем его условно оптимальным размещением) учитывает также и степень вариации признака в различных группах генеральной совокупности. Расчет объема выборки из каждой типической группы производится по формуле

$$n_i = n\frac{N_i\sigma_i}{\sum_{i=1}^{k}N_i\sigma_i}, \tag{6.19б}$$

где σ_i - среднее квадратическое отклонение изучаемого признака в i-й группе.

При таком размещении пропорция отбора для групп с большой колеблемостью признака увеличивается, что в свою очередь приводит к соответственному уменьшению возможной случайной ошибки в определении групповой средней. Использование принципа оптимального размещения на практике затрудняется тем, что часто мы не располагаем данными о величинах σ_i в генеральной совокупности.

При типическом отборе в выборку попадают, таким образом, представители всех типических групп, поэтому вероятность получить большую точность выборки здесь выше, чем при простой случайной выборке.

На основе типической выборки можно непосредственно определить средние по группам, т.е. величины $\tilde{x}_i$. Величина средней для всей совокупности $\tilde{x}_0$ есть средняя арифметическая взвешенная из выборочных средних по группам, причем в качестве веса используют объем соответствующей типической группы в генеральной совокупности N_i, т.е.

$$\tilde{x}_0 = \frac{\sum_{i=1}^{k}\tilde{x}_i N_i}{\sum_{i=1}^{k}N_i}. \tag{6.20}$$

При пропорциональном размещении единиц по типическим группам можно применять и формулу

$$\tilde{x}_0 = \frac{\sum_{i=1}^{k}\tilde{x}_i n_i}{\sum_{i=1}^{k}n_i} = \frac{\sum_{i=1}^{k}\tilde{x}_i N_i}{\sum_{i=1}^{k}N_i}. \tag{6.20а}$$

Из приведенной формулы очевидно, что точность типической выборки будет зависеть от степени надежности определения групповых средних, т.е. от того, насколько хорошо отобранные единицы репрезентируют среднюю каждой типической группы, и, следовательно, случайная ошибка типической выборки зависит от величины внутригрупповых дисперсий. Поэтому при определении ошибки выборки для средней величины признака в совокупности берется не общая дисперсия, а средняя из внутригрупповых дисперсий, которую обозначим $\overline{\sigma}^2$. Формулы для определения стандартной ошибки типической выборки для пропорционального и оптимального размещения приведены в табл. 6.14.

Таблица 6.14

Вид размещения единиц по типическим группам	Ошибка средней арифметической		Ошибка доли	
	Способ отбора		Способ отбора	
	повторный	бесповторный	повторный	бесповторный
Пропорциональное	$\sqrt{\frac{\overline{\sigma}^2}{n}}$	$\sqrt{\frac{\overline{\sigma}^2}{n}\left(1-\frac{n}{N}\right)}$	$\sqrt{\frac{w_i\left(1-w_i\right)}{n}}$	$\sqrt{\frac{w_i\left(1-w_i\right)}{n}\left(1-\frac{n}{N}\right)}$
Оптимальное	$\frac{1}{N}\sqrt{\Sigma\frac{\sigma_i^2 N_i^2}{n_i}}$	$\frac{1}{N}\sqrt{\Sigma\frac{\sigma_i^2 N_i^2}{n_i}\left(1-\frac{n_i}{N_i}\right)}$	$\frac{1}{N}\sqrt{\Sigma\frac{w_i\left(1-w_i\right) N_i^2}{n_i}}$	$\frac{1}{N}\sqrt{\Sigma\frac{w_i\left(1-w_i\right) N_i^2}{n_i}\left(1-\frac{n_i}{N_i}\right)}$

Примечание.
σ_i^2 - дисперсия признака в соответствующей типической группе.

Так как общая дисперсия признака в генеральной совокупности согласно теореме сложения дисперсий равна сумме межгрупповой и средней из внутригрупповых дисперсий, величина стандартной ошибки типической выборки всегда меньше, чем при простой случайной выборке. Различия между общей дисперсией и средней из внутригрупповых будут тем больше, чем больше различия между группами (т.е. соответственно чем больше величина межгрупповой дисперсии). Таким образом, типическая выборка в сравнении с простой случайной выборкой имеет преимущества при наличии в составе генеральной совокупности резко различающихся между собой групп, в то время как колеблемость признака внутри групп незначительна.

Для организации типической выборки необходимо располагать даными, позволяющими выделить в составе генеральной совокупности однородные группы, составить списки единиц этих групп и т.д. Такие данные можно получить либо из материалов текущего оперативного учета, либо из данных сплошных переписей или учетов.

При районированной выборке случайный отбор проводится внутри уже сложившихся территориальных (или организационных) единств (районов, кварталов, городов и т.п.); степень одно-

родности выделяемых «районов» не поддается регулировке со стороны исследователя: рассеяние внутри «районов» может оказаться значительным, а порайонные средние, наоборот, могут мало отличаться друг от друга. В этом случае такая выборка утрачивает свои преимущества перед случайной выборкой из общей массы. При организации простой случайной или типической выборки единицы наблюдения совпадают, т.е. отбираются единицы, признаки которых подлежат регистрации при наблюдении.

При **серийной выборке** в случайном порядке отбираются группы (серии, гнезда) единиц, которые подвергаются сплошному обследованию. Серийная выборка перед уже рассмотренными видами выборки имеет преимущества организационного характера и широко используется там, где генеральная совокупность состоит из определенным образом обособленных групп единиц. Например, при статистическом контроле качества готовой продукции серийная выборка находит распространение в случаях применения так называемой «мерной» тары, например, сверла одного типоразмера упаковываются в конические ящики со строго установленным количеством в каждом из них. На практике, таким образом, часто встречается серийный отбор с равными сериями.

Так как внутри серии производится сплошное обследование всех входящих в нее единиц, средние по сериям определяются без случайной ошибки. Таким образом, общая средняя $\tilde{x}_0$, рассматриваемая как оценка генеральной средней, будет зависеть от того, насколько точно средние серий $\tilde{x}_i$ будут репрезентировать генеральную среднюю. Тем самым случайная ошибка средней по всей выборке зависит от величины межсерийной дисперсии δ^2 [1] характеризующей степень колеблемости серийных средних.

При равновеликих сериях стандартная ошибка выборки определится по формулам:

а) при повторном отборе серий

$$\mu_{\bar{x}} = \sqrt{\frac{\delta^2}{s}}; \tag{6.21}$$

---

[1] При равновеликих сериях $\tilde{x}_0 = \frac{\sum_{i=1}^{s}\tilde{x}_i}{s}$, $\delta^2 = \frac{\sum_{i=1}^{s}\left(\tilde{x}_i - \tilde{x}_0\right)^2}{s}$, где s - число отобранных серий.

б) при бесповторном отборе серий

$$\mu_{\bar{x}} = \sqrt{\frac{\delta^2}{s}\left(\frac{S-s}{S-1}\right)} \approx \sqrt{\frac{\delta^2}{s}\left(1-\frac{s}{S}\right)}, \tag{6.21а}$$

где S - общее число серий в генеральной совокупности,

s - число отобранных серий.

Из приведенных формул следует, что случайная ошибка серийной выборки должна быть тем меньше, чем меньше межсерийная дисперсия и чем, следовательно, более однородны внутри себя серии. Однако на практике крайне редко можно встретить такой вариант, когда отобранные в случайном порядке серии не отличаются одна от другой по величине изучаемого признака, тогда как единицы, входящие в отобранную серию, обычно близки по величине изучаемого признака. Это связано с общностью условий их формирования, что обусловливает малую величину внутригрупповой дисперсии.

Поскольку при серийной выборке, как правило, обследуется незначительное число серий, случайная ошибка серийной выборки получается несколько больше, чем при других способах отбора, т.е. серийная выборка менее точна, чем выборка, основанная на индивидуальном отборе. Тем не менее серийный отбор широко применяется в практике выборочных обследований, особенно в тех случаях, когда обследование охватывает обширную территорию и гнездами являются территориальные единицы, что связано со значительной экономией денежных средств на проведение наблюдения.

Средняя ошибка выборочной доли при случайной выборке в случае равновеликих серий определяется по формулам:

а) при повторном отборе

$$\mu_p = \sqrt{\frac{\delta_p^2}{s}} \tag{6.22}$$

где δ_p^2 - межсерийная (межгрупповая) дисперсия доли, определяемая по формуле:

$$\delta_p^2 = \frac{\sum_{i=1}^{s}\left(w_i - \bar{w}\right)^2}{s},$$

где w_i - доли признака i-й серии;

$\bar{w}$ - доля признака по всей выборочной совокупности.

б) при бесповторном отборе

$$\mu_p = \sqrt{\frac{\delta_p^2}{s}\left(1-\frac{s}{S}\right)}. \tag{6.22а}$$

Мы рассмотрели так называемую *одноступенчатую выборку*, когда в случайно отобранных группах генеральной совокупности сплошь обследовали входящие в них единицы. Но можно поступить иначе, сформировав выборочную совокупность в два этапа: сначала в случайном порядке выбираются подлежащие обследованию серии, затем из каждой отобранной серии в случайном порядке отбирается определенное количество единиц, подлежащих непосредственному наблюдению. Ошибка такой выборки будет зависеть от ошибки серийного отбора и от ошибки индивидуального отбора, т.е. *многоступенчатый отбор* дает, как правило, менее точные результаты по сравнению с одноступенчатым, что объясняется возникновением ошибок репрезентативности на каждой ступени выборки.

При проведении бесповторной выборки в равновеликих сериях ошибка двухступенчатого отбора определяется по формуле:

$$\mu_{\bar{x}} = \sqrt{\frac{\delta^2}{s}\left(1-\frac{s}{S}\right)+\frac{\bar{\sigma}^2}{ms}\left(1-\frac{ms}{N}\right)}, \tag{6.23}$$

где m - число отобранных в каждой серии единиц;

$\bar{\sigma}^2$ - средняя из внутрисерийных дисперсий.

Многоступенчатый отбор характеризуется тем, что на всех ступенях, кроме последней, осуществляется только отбор, а наблюдение единиц производится только на последней ступени. Поэтому списки единиц нужны лишь на последней ступени отбора. При многоступенчатой выборке единицы отбора на первых ступенях обычно представляют собой организационные единства единиц наблюдения и на разных ступенях применяются единицы отбора разных порядков. Например, при текущем изучении бюджетов рабочих и служащих единицей наблюдения является семья. Формирование выборочной совокупности производится путем отбора сначала предприятий, а затем лиц, работающих на предприятиях, т.е. членов семей.

Одной из распространенных форм в практике выборочного наблюдения является **механический отбор**. При механическом отборе наблюдению подвергаются единицы, находящиеся на равном расстоянии в определенной последовательности расположения единиц генеральной совокупности. Допустим, что имеется исчерпывающий список единиц генеральной совокупности и что единицы располагаются в порядке, являющемся случайным по отношению к подлежащим изучению признакам (например, список рабочих по алфавиту). В зависимости от объема выборки из списка для обследования выбирается либо каждая четвертая единица, либо каждая десятая единица и т.д. При проведении механической выборки генеральная совокупность в сущности разбивается на равные по численности группы и из каждой группы для обследования отбирается одна единица.

В том случае, когда к механическому отбору прибегают в целях повышения репрезентативности выборочного наблюдения, списки единиц генеральной совокупности составляются в форме ранжированной совокупности, т.е. единицы заносятся в списки в порядке возрастания (убывания) какого-либо признака, тесно связанного с изучаемыми признаками и до известной степени, определяющего их величину. Например, при изучении бюджетов рабочих и служащих используется механический отбор из списков, составленных в порядке убывания величины средней месячной заработной платы работников.

Механический отбор очень удобен и в тех случаях, когда мы не можем заранее составить список единиц генеральной совокупности. Например, выборка берется из постепенно формирующейся во времени совокупности или из практически бесконечной совокупности. Так при обследовании покупок можно наблюдать каждого десятого входящего в магазин покупателя; при контроле качества продукции - проверять, например, каждую пятую сходящую со станка деталь и т.д.

При проведении механической выборки нужно установить шаг отсчета, т.е. расстояние между отбираемыми единицами, и начало отсчета, т.е. номер той единицы, которая должна быть обследована первой. Шаг отсчета устанавливают в зависимости от предполагаемого процента отбора. Например, из генеральной совокупности объемом 1000 единиц обследованию подлежат 100 единиц. Это означает, что из каждых 10 единиц обследованию будет подлежать одна единица, т.е. шаг отсчета равен 10.

Выбор начала отсчета непосредственно связан со способом расположения единиц генеральной совокупности в списках. В случае неупорядоченного расположения единиц, из совокупности единиц первого интервала путем случайного отбора выбирают начальную единицу. Например, на основании жеребьевки номер начальной единицы равен 4. Тогда в выборку попадут единицы, стоящие в списке под номерами 4, 14, 24, 34, ... 984, 994.

Если единицы в списке ранжированы по определенному признаку, за начало отсчета принимают единицу, лежащую в середине первого интервала. В данном примере из первых десяти единиц нужно выбрать либо пятую, либо шестую единицу. Тогда в выборку попадут единицы с порядковыми номерами 5, 15, 25, 35, 45, ... 985, 995, (или 6, 16, 26, 36, 46, ... 986, 996).

Если генеральная совокупность предварительно не ранжируется и предполагается, что расположение единиц в ней является случайным по отношению к изучаемым признакам, то механический отбор можно рассматривать как разновидность простой случайной бесповторной выборки. Таким образом, величина случайной ошибки такого отбора будет определяться по формуле:

$$\mu_{\bar{x}} = \sqrt{\frac{\sigma^2}{n}\left(1 - \frac{n}{N}\right)} \qquad (6.24)$$

В совокупностях бесконечно большого объема отношение n/N очень мало и множитель $(1 - n/N)$ очень мало отличается от единицы. Поэтому для практически бесконечных совокупностей величина случайной ошибки механической выборки может быть определена по формуле:

$$\mu_{\bar{x}} = \sqrt{\frac{\sigma^2}{n}} \qquad (6.24а)$$

При механическом отборе из ранжированной совокупности единицы отбора расположены не в случайном порядке, а в порядке, обеспечивающем более пропорциональное представительство в выборке отдельных групп генеральной совокупности, различающихся по величине признака. Такой отбор можно уподобить типической выборке с пропорциональным размещением. Генеральная совокупность при этом способе отбора как бы разбивается на однородные группы равного объема и из каждой группы выбирается одна единица. Однако такое сопоставление механического отбора из ранжированной совокупности с типической выборкой не вполне правомерно, поскольку границы групп устанавливаются в известной мере произвольно, механически; кроме того, определение ошибки по формулам типической выборки практически невозможно из-за отсутствия данных о вариации по группам. В этой связи величина ошибки механического отбора из ранжированной совокупности приближенно оценивается по формулам простой случайной выборки.

В целях экономии средств данные по некоторым интересующим нас признакам можно анализировать на основании изучения всех единиц выборочной совокупности, а по другим признакам - на основании части единиц выборочной совокупности, которые представляют подвыборку из единиц первоначальной выборки. Этот способ называют **двухфазным отбором**. При наличии нескольких подвыборок можно говорить о **многофазном отборе**.

Многофазный отбор по своей структуре отличается от многоступенчатого отбора, так как при многофазном отборе пользуются на каждой фазе одними и теми же единицами отбора, при многоступенчатом отборе на разных ступенях применяются единицы отбора разных порядков. Многофазным отбором чаще всего пользуются в тех случаях, когда различно число единиц, необходимых для определения отдельных показателей с заданной точностью. Это связано как с различиями в степени колеблемости взаимосвязанных переменных, так и с разной точностью, требуемой для подсчетов. Ошибки при многофазной выборке рассчитывают на каждой фазе отдельно.

Часто бывает целесообразно взять из изучаемой совокупности две или несколько независящих друг от друга выборок, применяя для получения каждой из них один и тот же способ отбора. Такие выборки называют **взаимопроникающими выборками**. Преимущество взаимопроникающих выборок состоит в том, что они позволяют получить отдельные и независимые оценки тех или иных признаков изучаемой совокупности. Если проведение разных выборок поручено разным исследователям, то сравнение результатов обеспечивает взаимную проверку работы самих исследователей.

6.10. Практика применения выборочного метода наблюдения

При правильной организации результаты выборочного обследования по точности не уступают данным сплошного наблюдения. Вот почему для современной организации статистического наблюдения в стране актуальными стали проблемы внедрения выборочного метода. В отечественной статистике уже накоплен определенный опыт выборочных обследований не только в качестве самостоятельного наблюдения, но и в сочетании со сплошным.

Выборка находит применение в расширении программы переписей населения, когда часть вопросов адресована не всему населению, а лишь специально отобранной его части. При выборочной разработке данных появляется возможность более детального описания уменьшенной совокупности, снижаются затраты подготовительного и машинного времени. При проведении переписей населения 1979 и 1989 гг. выборочная совокупность составляла 25% общей численности населения. По окончании переписей производится также выборочный контроль правильности проведенных записей регистратором и инструктором-контролером. В этой области есть возможность более широкого использования выборочного наблюдения, в частности, при выборочной разработке данных, полученных при сплошной и выборочной переписях населения.

Одним из направлений практического применения выборочного метода является **статистический контроль качества продукции**. Статистические методы используют для организации приемочного контроля качества готовой продукции и текущего предупредительного контроля качества в процессе изготовления продукции. Выборочный контроль качества готовой продукции заключается в том, что из всей партии изготовленной продукции отбирается на пробу некоторое число изделий и оценивается качество каждого из них. По доле дефектных изделий в пробе судят о качестве всей партии изделий. Если результат контроля дает основание считать, что доля брака во всей партии не превышает некоторого допустимого предела, то партия принимается без сплошного контроля.

Если же по результатам контроля доля брака в отобранной части изделий превышает допустимый предел, то изготовленная партия подвергается сплошной проверке, т.е. проверяются все оставшиеся $(N-n)$ изделия. Естественно, что сплошной контроль оставшихся изделий невозможен там, где он связан с необходимостью уничтожения или порчи изделия. В таких случаях выборочный контроль качества является единственно возможным источником данных о качестве всей продукции.

Предупредительный статистический контроль качества позволяет систематически контролировать ход производственного процесса в целях предупреждения возможных нарушений, создающих угрозу получению требуемого качества продукции. При проведении предупредительного контроля качества на операциях механической обработки деталей осуществляется контроль положения центра настройки и рассеивания размеров. В качестве показателя центра настройки может быть использована медиана или средняя арифметическая, а рассеивание размеров можно оценить с помощью крайних значений, размаха вариации или среднего квадратического отклонения. В картах статистического контроля содержатся контрольные диаграммы, на которых по оси ординат наносится линия среднего допускаемого размера и границы поля допуска контролируемого размера.

Непременным условием правильной постановки статистического контроля качества является предварительное изучение закономерностей наблюдаемого процесса на основе так называемых больших выборок, которые проводятся непосредственно перед постановкой текущего предупредительного контроля. «Большая» выборка проводится обычно на протяжении нескольких смен. В ходе такого наблюдения производится сплошной контроль всех изготовляемых изделий, что дает возможность регистрировать все последовательные состояния процесса и оценить его устойчивость по контролируемому параметру. По результатам предварительного наблюдения определяются параметры нормального течения процесса и прежде всего - средние уровни и дисперсия. Зная эти параметры, можно установить так называемые контрольные границы для медианы и крайних значений.

«Контрольные границы» будут равны:

а) для медианы:

$$c \pm t\,\mu \qquad \left(\mu = \frac{\sigma}{\sqrt{n}}\right), \tag{6.25}$$

где c - середина поля допуска;

σ - среднее квадратическое отклонение контролируемого размера;

t - принимается равным от 2,9 до 3,9;

n - принятый объем контрольной пробы.

б) для крайних значений:

$$c \pm t\sigma \tag{6.25a}$$

(t принимается равным от 2-х до 3-х).

Сигнал о разладке процесса подается в том случае, если медиана или какое-либо из крайних значений в пробе окажутся вне установленных контрольных границ.

Специальной формой организации выборочного наблюдения является **метод моментных наблюдений**, предложенный в 1938 г. Типпетом и получивший распространение при изучении структуры внутрисменных затрат времени рабочих и в дальнейшем при характеристике использования фонда времени производственного оборудования. Сущность метода моментных наблюдений состоит в периодической фиксации состояний наблюдаемых единиц в заранее установленные или случайно выбранные моменты времени.

Различают *моментное наблюдение со случайным отбором моментов* и периодическое моментное наблюдение. В первом случае момент наблюдения (час и минуту) выбирают с помощью таблиц случайных чисел. *Периодическое моментное наблюдение* организуется по аналогии с механическим отбором: наблюдение состояний процесса осуществляется через определенные промежутки времени.

При подготовке к проведению моментных наблюдений необходимо разработать перечень различных возможных состояний наблюдаемых объектов (например, элементов сменного фонда рабочего времени). Степень детализации элементов затрат рабочего времени или причин простоев и видов работы оборудования зависит от поставленной задачи. Если требуется получить общую информацию, достаточно выделить 3-4 вида простоев и работ.

В тех случаях, когда требуется детальная разработка данных о причинах простоев, необходима более развернутая классификация элементов фонда рабочего времени. Элементы классификации должны быть четко сформулированы, не повторяться и иметь достаточный удельный вес, чтобы представлять интерес для анализа. В практике применения моментных наблюдений следует выделять только те элементы затрат времени, которые имеют удельный вес не менее 5%, поскольку в противном случае резко увеличивается необходимый объем наблюдений.

Число моментов, в которые производится фиксация состояний процесса, определяется на основе формулы предельной ошибки доли

$$n = \frac{t^2 p\,(1-p)}{\Delta_p^2}, \tag{6.26}$$

где p - доля данного вида потерь в сменном фонде времени [1],
Δ_p - абсолютная величина предельной ошибки выборочной доли.

Величину предельной ошибки чаще задают в относительной форме, т.е. задаются величиной:

$$\Delta_{отн.} = \frac{\Delta_p}{p} \cdot 100\% \qquad (6.26а)$$

откуда $\Delta_p = \frac{\Delta_{отн.} \cdot p}{100}$. (6.26б)

Тогда необходимые число наблюдений определится по формуле

$$n = \frac{t^2 p\left(1-p\right)}{\Delta^2_{отн.} \cdot p^2} \cdot 100^2 = \frac{t^2\left(1-p\right)}{\Delta^2_{отн.} \cdot p} \cdot 100^2 \qquad (6.26в)$$

В соответствии с рассчитанным числом человеко-моментов и численностью обследуемых рабочих в данной группе устанавливается необходимое число обходов как частное от деления количества наблюдений на число рабочих, подлежащих обследованию.

Далее встает вопрос о количестве смен, на которые должно быть распределено общее количество обходов. Для ответа на этот вопрос нужно знать продолжительность одного обхода. Так как наблюдатель должен обходить все рабочие места по заранее установленному маршруту, продолжительность одного обхода может быть установлена с достаточной точностью с учетом затрат времени на соответствующие записи и перемещение наблюдателя.

Допустим, что наблюдение проводится на участке цеха, где в смену работает 30 человек. Поставлена задача - выявить важнейшие причины потерь рабочего времени. По данным фотографий рабочего дня известно, что удельный вес потерь рабочего времени по различным причинам составляет не менее 5%, т.е. $p = 0,05$. Приняв $\Delta_{отн.}$ равной 10%, получим, что необходимое число моментов наблюдения равно 7600.

[1] Величина p определяется по данным предыдущих наблюдений или по данным предварительно проведенного пробного обследования. Следует так же отметить, что во всех случаях соблюдается условие $0< p <1$. Поэтому максимум произведения $p(1-p)$ составит 0,25, что соответствует $p = 0,5$. Поэтому приведенную формулу можно применять и тогда, когда информация о генеральной доле полностью отсутствует: $n = \frac{0,25\,t^2}{\Delta^2_p}$, причем величина требуемого для обеспечения заданной предельной ошибки объема выборки окажется завышенной.

Отсюда общее количество обходов равно 254 (7600:30=253,3). Выполнить указанное количество обходов за одну смену практически невозможно, так как это означало бы, что каждая последующая запись должна быть сделана примерно через 2,0 минуты [(8,0•60):254=1,89].

В действительности, затраты времени на один обход составляют по предварительным прикидкам примерно 7 мин, таким образом за смену может быть сделано не более 69 обходов (480:7=68,6). Если принять во внимание, что в некоторых случаях наблюдателю потребуется дополнительное время на выяснение причин простоя, то минимально необходимое время на один обход следует принять равным 10 мин.

Тогда за смену наблюдатель сможет сделать 48 обходов (480:10=48,0), а чтобы выполнить все необходимое число обходов, потребуется не менее 5 смен (254:48=5,3). Таким образом, в нашем примере наблюдение целесообразно проводить ежедневно на протяжении всей рабочей недели (иногда целесообразно провести наблюдение и за более длительный период времени, уменьшив ежедневное число обходов).

Выбор сроков проведения наблюдения обычно должен опираться на дополнительную информацию о закономерностях хода процесса. Наличие известной неравномерности в загрузке оборудования и рабочих всегда имеет место на протяжении рабочих дней недели, а потому выбор пятидневки в качестве периода наблюдения позволит получить более достоверную картину распределения затрат рабочего времени.

Следующим этапом в организации моментного наблюдения является составление графика проведения наблюдений, т.е. речь идет о выборе моментов времени, в которые должны быть сделаны соответствующие записи по наблюдаемым единицам.

С организационной точки зрения часто более удобно использование периодического моментного наблюдения. В этом случае наблюдение состояний процесса осуществляется через определенные промежутки времени. В нашем примере наблюдатель должен подходить к одному и тому же рабочему месту через каждые 10 мин. В случайном порядке отбирается только момент времени, в который должно начинаться наблюдение. Допустим, что смена начинается в 7 час. 30 мин. С помощью жеребьевки установили, что первый обход начинается в 7 час. 34 мин., тогда второй обход в 7 час. 44 мин. третий - в 7 час. 54 мин. и т.д.

Применение периодического моментного наблюдения позволяет получить сведения об использовании рабочего времени на данном участке на протяжении всей смены, тогда как при применении случайного отбора моментов принципиально возможны случаи, когда какой-то из часов смены совсем выпадает из рассмотрения. Для выявления же, например, случаев нарушения трудовой дисциплины (в связи с поздним началом работы и ранним окончанием ее) особенно важно получить сведения и в начале, и в конце смены.

Специальные выборочные наблюдения находят все более широкое применение в социальной статистике. Так важнейшим источником информации об уровне жизни являются данные регулярно проводимых выборочных обследований бюджетов семей рабочих и служащих. Информацию о характере использования бюджета времени трудящихся органы государственной статистики также получают по материалам выборочных обследований. В 1980 г. проводилось обследование бюджета времени в 52 тыс. семей рабочих, служащих и колхозников; в 1985 - в 51 тыс. семей постоянно ведущих такие бюджеты.

Применение выборочного метода при изучении общественного мнения в ряде стран уже имеет свою историю. С 1935 г. в США организацией опросов общественного мнения с применением научных методов занимается Американский институт общественного мнения, который основал Джордж Гэллап. Опросы Гэллапа проводятся два раза в неделю по разнообразным политическим, социальным и экономическим проблемам. Результаты национальных обследований, регулярно проводимых Институтом Гэллапа, основываются, как правило, на персональных интервью с минимумом опрашиваемых (примерно 1500 человек). После разбивки страны по территориальному признаку и по числу проживающих в целях соответствия выборки фактическому распределению населения в случайном порядке отбираются районы в пределах города или округа. В каждом случайно отобранном пункте проводится приблизительно пять интервью в определенных, отобранных на случайной основе домах. В каждом доме интервьюер обращается к самому молодому мужчине или же, если таковой отсутствует, к самой пожилой женщине.

В нашей стране также действуют различные центры по изучению общественного мнения, в частности население широко знакомится с результатами опросов в ходе различных выборных кампаний. Вместе с тем нужно отметить, что практика проведения подобных опросов населения нуждается в совершенствовании и необходимо более четкое их теоретическое обоснование, так как результаты выборов и прогнозы различных центров имели нередко весьма существенные расхождения.

Контрольные вопросы к главе 6

1. В чем преимущества выборочного метода в сравнении с другими видами статистических наблюдений?
2. Что означает ошибка репрезентативности, какие факторы определяют ее величину?
3. Чем отличается распределение ошибок простой случайной выборки при проведении больших и малых выборок?
4. От чего зависит точность оценки параметров генеральной совокупности (генеральный средней и генеральной доли)?
5. Запишите доверительные интервалы генеральной средней с вероятностью 0,95 и 0,99.
6. Чем отличается величина средней квадратической ошибки простой случайной выборки при повторном и бесповторном отборе? Какая из этих ошибок больше?
7. Для решения каких вопросов организации выборочного наблюдения и оценки его результатов может использоваться формула средней квадратической ошибки выборки?
8. Как определяется предельная ошибка при проведении большой и малой выборок?
9. Дайте понятие критической области при статистической проверке гипотез.
10. Какие методы можно использовать для исключения выделяющихся наблюдений?
11. Как оценить существенность расхождений двух выборочных средних (долей)?
12. Какой вид выборочного наблюдения следует использовать, если генеральная совокупность не является однородной?
13. В чем состоят преимущества серийной выборки перед простой случайной выборкой?

14. Какие способы размещения общего объема подлежащих обследованию единиц могут быть использованы при организации типической выборки?

15. В чем преимущества механической выборки и как определяется величина ее стандартной ошибки?

16. Как нужно изменить объем механической выборки, если среднюю квадратическую ошибку следует уменьшить в 1,5 раза?

17. Назовите важнейшие области применения выборочного метода в практике государственной статистики.

Глава 7
КОРРЕЛЯЦИОННАЯ СВЯЗЬ И ЕЕ СТАТИСТИЧЕСКОЕ ИЗУЧЕНИЕ

7.1. Понятие о корреляционной связи и предпосылки ее использования

Статистические распределения характеризуются наличием более или менее значительной вариации в величине признака у отдельных единиц совокупности. Естественно, возникает вопрос о том, какие же причины формируют уровень признака в данной совокупности и каков конкретный вклад каждой из них. *Изучение зависимости вариации признака от окружающих условий и составляет содержание теории корреляции* [1].

Изучение действительности показывает, что вариация каждого изучаемого признака находится в тесной связи и взаимодействии с вариацией других признаков, характеризующих исследуемую совокупность единиц. Вариация уровня производительности труда работников предприятий зависит от степени совершенства применяемого оборудования, технологии, организации производства, труда и управления и других самых различных факторов.

При изучении конкретных зависимостей одни признаки выступают в качестве факторов, обусловливающих изменение других признаков. Признаки этой первой группы в дальнейшем будем называть признаками-факторами (**факторными признаками**); а признаки, которые являются результатом влияния этих факторов, будем называть **результативными**. Например, при изучении зависимости между производительностью труда рабочих и энер-

[1] Основоположниками теории корреляции считаются английские биометрики Ф. Гальтон (1822-1911) и К.Пирсон (1857-1936). Термин «корреляция» был заимствован из естествознания и обозначает соотношение, соответствие. Представление о корреляции как об отношении взаимозависимости между случайными переменными величинами лежит в основе математико-статистической теории корреляции.

говооруженностью их труда уровень производительности труда является результативным признаком, а энерговооруженность труда рабочих - факторным признаком.

Рассматривая зависимости между признаками, необходимо выделить прежде всего две категории зависимости: 1) функциональные и 2) корреляционные.

Функциональные связи характеризуются полным соответствием между изменением факторного признака и изменением результативной величины, и каждому значению признака-фактора соответствуют вполне определенные значения результативного признака.

В **корреляционных связях** между изменением факторного и результативного признака нет полного соответствия, воздействие отдельных факторов проявляется лишь в среднем при массовом наблюдении фактических данных. В простейшем случае применения корреляционной зависимости величина результативного признака рассматривается как следствие изменения только одного фактора (например, энерговооруженность труда рассматривается как причина роста производительности труда). Однако выделенный в данном примере в качестве основного признак-фактор не является единственной причиной изменения результативного признака, а наряду с ним на величину результативного признака влияет множество других причин. Как уже указывалось, на формирование уровня производительности труда на предприятии более или менее существенное влияние оказывают факторы, характеризующие степень совершенства применяемой техники и технологии, уровень механизации и автоматизации труда, специализации производства, состав работающих, текучесть кадров и т.п.

Кроме того, сам признак-фактор в свою очередь может зависеть от изменения ряда обстоятельств. В сложном взаимодействии находится результативный признак - в более общем виде он выступает как фактор изменения других признаков. Отсюда результаты корреляционного анализа имеют значение в данной связи, а интерпретация этих результатов в более общем виде требует построения системы корреляционных связей.

Одновременное воздействие на изучаемый признак большого количества самых разнообразных факторов приводит к тому, что одному и тому же значению признака-фактора соответствует целое распределение значений результативного признака, поскольку в каждом конкретном случае прочие факторные признаки могут изменять силу и направленность своего воздействия.

При сравнении функциональных и корреляционных зависимостей следует иметь в виду, что при наличии функциональной зависимости между признаками можно, зная величину факторного признака, точно определить величину результативного признака. При наличии же корреляционной зависимости устанавливается лишь тенденция изменения результативного признака при изменении величины факторного признака. В отличие от жесткости однозначно функциональной связи корреляционные связи характеризуются множеством причин и следствий и устанавливаются лишь их тенденции.

Необходимо отметить, что экономической теории принадлежит решающее слово в обосновании связей между теми или иными признаками. При этом теоретический анализ должен показать, какие факторы влияют на исследуемый признак или же влияние каких факторов должно быть проверено. Статистическое выражение связи между явлениями может показать, что изменения одного из сопоставляемых признаков сопровождаются изменениями другого. Следовательно, нужно искать объяснение этим изменениям в их содержательном анализе. С помощью статистических методов изучения зависимостей можно установить, как проявляется теоретически возможная связь в данных конкретных условиях. Статистика не только отвечает на вопрос о реальном существовании намеченной теоретическим анализом связи, но и дает количественную характеристику этой зависимости. Зная характер зависимости одного явления от других, можно объяснить причины и размер изменений в явлении, а также планировать необходимые мероприятия для дальнейшего его изменения.

При исследовании корреляционных зависимостей между признаками решению подлежит широкий круг вопросов, к которым следует отнести: 1) предварительный анализ свойств моделируемой совокупности единиц; 2) установление факта наличия связи, определение ее направления и формы; 3) измерение степени тесноты связи между признаками; 4) построение регрессионной модели, т.е. нахождение аналитического выражения связи; 5) оценка адекватности модели, ее экономическая интерпретация и практическое использование.

Для того чтобы результаты корреляционного анализа нашли практическое применение и дали желаемый результат, должны выполняться определенные требования в отношении отбора объекта исследования и признаков-факторов. Одним из важнейших условий правильного применения методов корреляционного анализа является требование однородности тех единиц, которые подвергаются изучению методами корреляционного анализа. Например, при корреляционном анализе зависимостей тех или иных технико-экономических показателей работы предприятий от определенных факторов должны быть отобраны предприятия, выпускающие однотипную продукцию, имеющие одинаковый характер технологического процесса и тип используемого оборудования, для предприятий добывающей промышленности определенную роль играет и географическое размещение предприятий.

При выполнении указанных общих требований далее необходима **количественная оценка** однородности исследуемой совокупности по комплексу признаков. Одним из возможных вариантов такой оценки является расчет *относительных показателей вариации*. Традиционно широкое распространение для этих целей получил *коэффициент вариации*. Несколько реже применяется *отношение размаха вариации к среднеквадратическому отклонению*. Вывод о неоднородности исследуемой совокупности по тому или иному признаку требует проверки гипотезы о принадлежности «выделяющихся» (аномальных) значений признака исследуемой генеральной совокупности (см. гл. 6).

Другим важным требованием, обеспечивающим надежность выводов корреляционного анализа, является требование *достаточного числа наблюдений*. Как уже указывалось, влияние существенных причин может быть затушевано действием случайных факторов, «взаимопогашение» влияния которых на результативный показатель в известной мере происходит при выведении средней результативного показателя для массы случаев.

Определенные требования существуют и в отношении факторов, вводимых в исследование. Все множество факторов, оказывающих влияние на величину результативного показателя, в действительности не может быть введено в рассмотрение, да практически в этом и нет необходимости, так как их роль и значение в формировании величины результативного показателя могут иметь существенные различия. Поэтому при ограничении числа факторов, включаемых в изучение, наряду с качественным анализом целесообразно использовать и определенные количественные оценки, позволяющие конкретно охарактеризовать влияние факторов на результативный показатель (к оценкам можно отнести *парные коэффициенты корреляции, ранговые коэффициенты* при экспертной оценке влияния факторов и др.). Включаемые в исследование факторы должны быть независимыми друг от друга, так как наличие тесной связи между ними свидетельствует о том, что они характеризуют одни и те же стороны изучаемого явления и в значительной мере дублируют друг друга.

Все основные положения теории корреляции разрабатывались применительно к предположению о нормальном характере распределения исследуемых признаков. В этой связи целесообразным является изучение формы распределения, дающее возможность в известной мере обосновать правомерность применения методов корреляционного анализа (см. гл. 5).

Проверку нормальности распределения зависимой переменной можно проводить при каждом фиксированном значении факторного признака или внутри каждого отдельного интервала группирования, на которые разбит диапазон изменения факторного признака, пользуясь различными критериями согласия. Для проверки исходной предпосылки нормальности распределения необходимо в каждой группе иметь достаточно большое количество наблюдений, что в практических исследованиях встречается довольно редко.

Следует отметить, что на практике часто сталкиваются с теми или иными отклонениями от исходных предпосылок. Однако это не означает, что мы должны отказываться от применения методов корреляционно-регрессионного анализа.

И наконец, при построении корреляционных моделей факторы должны иметь количественное выражение, иначе составить модель корреляционной зависимости не представляется возможным.

7.2. Статистические методы выявления наличия корреляционной связи между двумя признаками

Для ответа на вопрос о наличии или отсутствии корреляционной связи используется ряд специфических методов: так называемые элементарные приемы (параллельное сопоставление рядов значений результативного и факторного признаков, графическое изображение фактических данных с помощью поля корреляции, построение групповой и корреляционной таблиц), а также дисперсионный анализ. Простейшим приемом обнаружения связи является **сопоставление двух параллельных рядов** - ряда значений факторного признака и соответствующих ему значений результативного признака. Значения факторного признака располагают в возрастающем порядке и затем прослеживают направление изменения величины результативного признака. Результативный признак (функцию) в дальнейшем будем обозначать через *y*, а факторный признак - через *x*.

Например, по 20 туристическим фирмам были установлены затраты на рекламу (факторный признак) и количество туристов, воспользовавшихся услугами каждой фирмы (результативный признак) [1]. В табл. 7.1 фирмы ранжированы по величине затрат на рекламу.

Таблица 7.1

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 видеть, 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.

[1] Поскольку речь идет об изложении методологии изучения взаимосвязей, мы ограничились совокупностью малого объема. Данные в примере условные.

В тех случаях, когда возрастание величины факторного признака влечет за собой возрастание и величины результативного признака, говорят о возможном наличии прямой корреляционной связи. Если же с увеличением факторного признака, величина результативного признака имеет тенденцию к уменьшению, то можно предполагать обратную связь между признаками.

Однако наличие большого числа различных значений результативного признака, соответствующих одному и тому же значению признака-фактора, 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.

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. 7.2 выписать эти результаты.

Для результативного признака необходимо определить величину интервала. Для этого воспользуемся формулой Стэрджесса:

$$h_y = \frac{y_{max} - y_{min}}{1 + 3{,}322 \lg n} = \frac{1200 - 720}{5} = 96 \text{ человек.}$$

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* (см. табл. 7.2).

Таблица 7.2

Центральное значение интервала, y	768	865	962	1059	1156		
группы по y / группы по x	720-816	817-913	914-1010	1011-1107	1108-1207	f_x	$\overline{y}_i$
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: $\overline{y}_j$ - 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.

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).

Среднее число туристов для первой группы, состоящей из трех фирм, которые тратят на рекламу 8 усл. ден. ед., будет равно 800 человек: $\overline{y}_1 = \dfrac{768 \cdot 2 + 865 \cdot 1}{3} = 800$ человек.

Для следующей группы, состоящей из пяти фирм, у которых затраты на рекламу 9 усл. ден. ед. $\overline{y}_2 = \dfrac{768 \cdot 1 + 865 \cdot 3 + 962 \cdot 1}{5} = 865$ человек и т. д. (рассчитанные таким образом средние представлены в графе 8 табл. 7.2).

Итак, увеличение средних значений результативного признака с увеличением значений факторного признака еще раз свидетельствует о возможном наличии прямой корреляционной зави-

симости 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.

Другим возможным приемом обнаружения связи является построение **групповой таблицы**. Все наблюдения разбиваются на группы в зависимости от величины признака-фактора, и по каждой группе вычисляются средние значения результативного признака (см. табл. 7.3).

Таблица 7.3

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. 7.2) и групповой таблицах (см. графу 3 табл. 7.3) 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 результативного признака от фактора, положенного в основу группировки. Иными словами, предполагается, что все прочие причины, если они носят случайный характер, при определении средней по группам взаимопогашаются, т.е. дают в каждой группе один и тот же результат. Следовательно, различия в величине средних будут связаны только с различиями в величине данного факторного признака. Если бы связи между факторным и результативным признаком не было, то все групповые средние были бы приблизительно одинаковыми по величине. Оценка существенности расхождения групповых средних лежит в основе использования методов дисперсионного анализа для выявления наличия и оценки существенности связи [1] (см. параграф 6.8 гл. 6).

Для предварительного выявления наличия связи и раскрытия ее характера, а в известной мере и для выбора формы связи применяют **графический метод**. Используя данные об индивидуальных значениях признака-фактора и 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. 7.1)

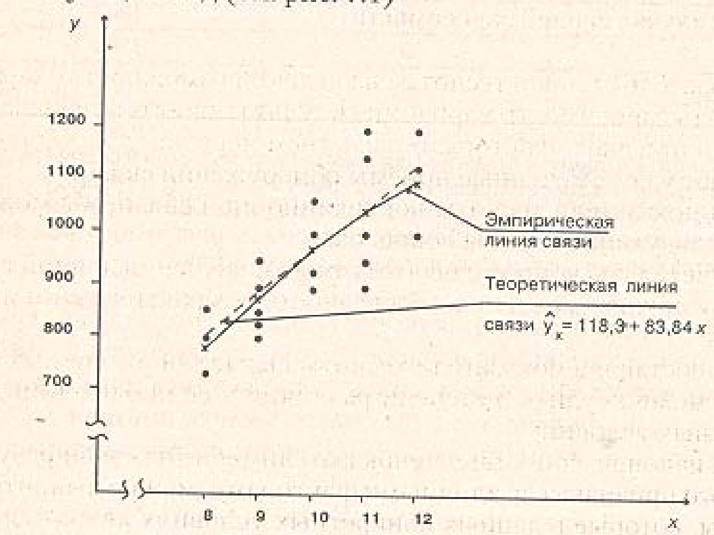

Рис. 7.1. **Зависимость числа клиентов фирмы (y) от ее затрат на рекламу (x)**

[1] Одним из недостатков аналитической группировки является неоднозначность результатов, которые зависят как от числа выделяемых групп, так и от установления границ интервалов.

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. 7.3) и по каждому значению затрат фирм на рекламу определены значения среднего числа туристов в группе (см. гр. 3 табл. 7.3) 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.

7.3. Измерение степени тесноты корреляционной связи в случае парной зависимости

Показатели степени тесноты связи дают возможность охарактеризовать зависимость вариации результативного признака от вариации признака-фактора. В известной мере они дополняют и развивают уже отмеченные приемы обнаружения связи.

Зная показатели тесноты корреляционной связи, мы можем решать следующие группы вопросов:

1) ответить на вопрос о необходимости изучения данной связи между признаками и целесообразности ее практического применения;

2) сопоставляя показатели тесноты связи для различных ситуаций, можно судить о степени различий в ее проявлении для конкретных условий;

3) и, наконец, сопоставляя показатели тесноты связи результативного признака с различными факторами, можно выявить те факторы, которые в данных конкретных условиях являются решающими и главным образом воздействуют на формирование величины результативного признака.

К простейшим показателям степени тесноты связи относят **коэффициент корреляции знаков**, который был предложен немецким ученым Г.Фехнером (1801-1887). Этот показатель основан на оценке степени согласованности направлений отклонений индивидуальных значений факторного и результативного признаков от соответствующих средних. Для его расчета вычисляют средние значения результативного и факторного признаков, а затем проставляют знаки отклонений для всех значений взаимосвязанных пар признаков.

Если ввести обозначения: n_a - число совпадений знаков отклонений индивидуальных величин от средней, $n_в$ - число несовпадений знаков отклонений, то коэффициент Фехнера можно записать таким образом:

$$K_ф = \frac{n_а - n_в}{n_а + n_в} \qquad (7.1)$$

Коэффициент Фехнера может принимать различные значения в пределах от -1 до +1. Если знаки всех отклонений совпадут, то $n_в = 0$ и тогда показатель будет равен 1, что свидетельствует о возможном наличии прямой связи. Если же знаки всех отклонений будут разными, тогда $n_а = 0$ и коэффициент Фехнера будет равен −1, что дает основание предположить наличие обратной связи.

Рассмотрим расчет $K_ф$ на примере, приведенном в табл. 7.4. Средний размер затрат на рекламу по всем 20 фирмам составит 9,95 усл. ден. ед., а среднее число туристов, воспользовавшихся услугами фирмы, - 952 человека. В графах 4 и 5 табл. 7.4 указаны знаки отклонений значений признаков от соответствующей средней.

Подсчитав число совпадений знаков $n_a = 16$ и число несовпадений знаков $n_в = 4$ (см. графу 6 табл. 7.4), рассчитаем коэффициент Фехнера:

$$K_ф = \frac{16-4}{16+4} = 0{,}6\ .$$

Полученная величина коэффициента Фехнера свидетельствует о том, что можно предполагать наличие прямой зависимости между исследуемыми признаками.

Как видно из приведенной формулы для расчета коэффициента Фехнера, величина этого показателя не зависит от величины отклонений факторного и результативного признака от соответствующей средней величины. Поэтому нельзя говорить о степени тесноты корреляционной связи, а тем более об оценке ее существенности на основании только коэффициента Фехнера. При малом объеме исходной информации коэффициент Фехнера практически решает ту же задачу, которая ставится при построении групповых и корреляционных таблиц, т.е. отвечает на вопрос о наличии и направлении корреляционной связи между призна-

Таблица 7.4

Порядковый номер фирмы	Затраты на рекламу, усл. ден. ед. X_i	Количество туристов, воспользовавшихся услугами фирмы, человек Y_i	Знаки отклонений индивидуальных значений признака от средней		Совпадение (а) или несовпадение (в) знаков
			для X_i	для Y_i	
1	2	3	4	5	6
1	8	800	-	-	a
2	8	850	-	-	a
3	8	720	-	-	a
4	9	850	-	-	a
5	9	800	-	-	a
6	9	880	-	-	a
7	9	950	-	-	a
8	9	820	-	-	a
9	10	900	+	-	b
10	10	1000	+	+	a
11	10	920	+	-	b
12	10	1060	+	+	a
13	10	950	+	-	b
14	11	900	+	-	b
15	11	1200	+	+	a
16	11	1150	+	+	a
17	11	1000	+	+	a
18	12	1200	+	+	a
19	12	1100	+	+	a
20	12	1000	+	+	a

ками. В том случае, если построена корреляционная или же групповая таблица, дополнительный расчет коэффициента Фехнера не имеет практической ценности.

Более совершенным показателем степени тесноты связи является **линейный коэффициент корреляции** (r)[1].

При расчете этого показателя учитываются не только знаки отклонений индивидуальных значений признака от средней, но и сама величина таких отклонений, т.е. соответственно для факторного и результативного признаков величины $x_i - \bar{x}$ и $y_i - \bar{y}$. Однако непосредственно сопоставлять между собой полученные абсолютные величины нельзя, так как сами признаки могут быть выражены в разных единицах (как это имеет место в представленном примере), а при наличии одних и тех же единиц измерения средние могут быть различны по величине. В этой связи срав-

[1] Коэффициент корреляции был предложен английским математиком К.Пирсоном.

нению могут подлежать отклонения, выраженные в относительных величинах, т.е. в долях среднего квадратического отклонения (их называют нормированными отклонениями). Так, для факторного признака будем иметь совокупность величин $t_{x_i} = \frac{x_i - \bar{x}}{\sigma_x}$, а для результативного $t_{y_i} = \frac{y_i - \bar{y}}{\sigma_y}$.

Полученные нормированные отклонения можно сравнивать между собой. Для того чтобы на основе сопоставления рассчитанных нормированных отклонений получить обобщающую характеристику степени тесноты связи между признаками для всей совокупности, рассчитывают среднее произведение нормированных отклонений. Полученная таким образом средняя и будет являться линейным коэффициентом корреляции r.

$$r = \frac{\sum_1^n t_{x_i} t_{y_i}}{n} = \frac{\sum_1^n \left(\frac{x_i - \bar{x}}{\sigma_x}\right)\left(\frac{y_i - \bar{y}}{\sigma_y}\right)}{n}, \tag{7.2}$$

или поскольку σ_x и σ_y для данных рядов являются постоянными и могут быть вынесены за скобку, то формула линейного коэффициента корреляции приобретает следующий вид:

$$r = \frac{\sum_1^n (x_i - \bar{x})(y_i - \bar{y})}{n\,\sigma_x\,\sigma_y}. \tag{7.2а}$$

Вычисление коэффициента корреляции по формуле (7.2) является достаточно трудоемкой операцией. Выполнив несложные преобразования, можно получить следующую формулу для расчета линейного коэффициента корреляции:

$$r = \frac{n\sum_1^n x_i y_i - \sum_1^n x_i \sum_1^n y_i}{\sqrt{\left[n\sum_1^n x_i^2 - \left(\sum_1^n x_i\right)^2\right]\left[n\sum_1^n y_i^2 - \left(\sum_1^n y_i\right)^2\right]}}. \tag{7.2б}$$

При пользовании этой формулой отпадает необходимость вычислять отклонения индивидуальных значений признаков от средней величины, что исключает ошибку в расчетах при округлении средних величин.

Линейный коэффициент корреляции может принимать любые значения в пределах от -1 до $+1$. Чем ближе коэффициент корреляции по абсолютной величине к 1, тем теснее связь между признаками. Знак при линейном коэффициенте корреляции указывает на направление связи - прямой зависимости соответствует знак плюс, а обратный зависимости - знак минус.

Если с увеличением значений факторного признака x, результативный признак y имеет тенденцию к увеличению, то величина коэффициента корреляции будет находиться между 0 и 1. Если же с увеличением значений x результативный признак y имеет тенденцию к снижению, коэффициент корреляции может принимать значения в интервале от 0 до -1.

Используем данные табл. 7.1 и рассчитаем линейный коэффициент корреляции.

$$\sum_{i=1}^{20} x_i = 199; \quad \sum_{i=1}^{20} y_i = 19\,050; \quad \sum_{i=1}^{20} x_i y_i = 192\,310$$

$$\left(\sum_i x_i\right)^2 = 39\,601; \quad \left(\sum_i y_i\right)^2 = 362\,902\,500$$

$$\sum_{i=1}^{20} x_i^2 = 2013; \quad \sum_{i=1}^{20} y_i^2 = 18\,497\,700.$$

$$r = \frac{20 \cdot 192\,310 - 199 \cdot 19\,050}{\sqrt{(20 \cdot 2013 - 39\,601)(20 \cdot 18\,497\,700 - 362\,902\,500)}} = 0{,}8105.$$

Полученная величина линейного коэффициента корреляции свидетельствует о возможном наличии достаточно тесной прямой зависимости между рассматриваемыми признаками.

Квадрат коэффициента корреляции (r^2) носит название **коэффициента детерминации**. Для рассматриваемого примера его величина равна 0,6569, а это означает, что 65,69% вариации числа клиентов, воспользовавшихся услугами фирмы, объясняется вариацией затрат фирм на рекламу своих услуг.

В тех случаях, когда исходная информация представлена в виде корреляционной таблицы, нужно учитывать частоты повторений как индивидуальных значений факторного и результативного признаков, так и число повторений данного сочетания значений фактора и результата. При расчете линейного коэффициента корреляции по корреляционной таблице формула (7.2б) имеет такой вид:

$$r=\frac{n\sum_{i=1}^{n}x_i y_i f_{xy}-\sum_{i=1}^{n}x_i f_x\sum y_i f_y}{\sqrt{\left[n\sum_{i=1}^{n}x_i^2 f_x-\left(\sum_{i=1}^{n}x_i f_x\right)^2\right]\left[n\sum_{i=1}^{n}y_i^2 f_y-\left(\sum_{i=1}^{n}y_i f_y\right)^2\right]}}. \qquad (7.3)$$

Здесь еще раз следует напомнить, что сама по себе величина коэффициента корреляции не является доказательством наличия причинно-следственной связи между исследуемыми признаками, а является оценкой степени взаимной согласованности в изменениях признаков. Установлению причинно-следственной зависимости предшествует анализ качественной природы явлений. Но есть и еще одно обстоятельство, объясняющее формулировку выводов о возможном наличии связи по величине коэффициента корреляции.

Связано это с тем, что оценка степени тесноты связи с помощью коэффициента корреляции производится, как правило, на основе более или менее ограниченной информации об изучаемом явлении. Возникает вопрос, насколько правомерно наше заключение по выборочным данным в отношении действительного наличия корреляционной связи в той генеральной совокупности, из которой была произведена выборка?

Принципиально возможны случаи, когда отклонение от нуля полученной величины выборочного коэффициента корреляции оказывается целиком обусловленным неизбежными случайными колебаниями тех выборочных данных, на основании которых он вычислен. Особенно осторожно следует подходить к истолкованию полученных коэффициентов корреляции при незначительных объемах выборочной совокупности.

В этой связи и возникает необходимость оценки существенности линейного коэффициента корреляции, дающая возможность распространить выводы по результатам выборки на генеральную совокупность. В зависимости от объема выборочной совокупности предлагаются различные методы оценки существенности линейного коэффициента корреляции. В отношении приводимых ниже критериев существенности можно сделать общее замечание, касающееся свойств исходной совокупности. Этим свойством является нормальное распределение значений признака в генеральной совокупности.

Рассмотрим следующие критерии, предлагаемые в статистической литературе:

1. При большом объеме выборки, отобранной из исходной нормально распределенной совокупности, можно считать распределение линейного коэффициента корреляции приближенно нормальным со средней, равной r и дисперсией $\sigma_r^2=\frac{(1-r^2)^2}{n-1}$, откуда средняя квадратическая ошибка коэффициента корреляции:

$$\sigma_r=\frac{1-r^2}{\sqrt{n-1}}, \qquad (7.4)$$

где r - линейный коэффициент корреляции, полученный по данным выборки;

n - объем выборки.

Если величина линейного коэффициента корреляции превышает величину средней квадратической ошибки более, чем в t_α раз, то можно говорить о существенности выборочного коэффициента корреляции, где α - уровень значимости 0,01 или 0,05. Если же отношение $\frac{|r|}{\sigma_r}$ окажется меньше t_α, то с вероятностью ($1-\alpha$) следует предполагать отсутствие корреляционной связи в генеральной совокупности.

Доверительный интервал для коэффициента корреляции будет записан так:

$$r-t_\alpha\frac{1-r^2}{\sqrt{n-1}}\leqslant r_{\text{ген.}}\leqslant r+t_\alpha\frac{1-r^2}{\sqrt{n-1}}, \qquad (7.5)$$

где $r_{\text{ген.}}$ - значение коэффициента корреляции в генеральной совокупности.

При малых объемах выборки и линейном коэффициенте корреляции, близком к 1, использование средней квадратической ошибки по формуле (7.4) в качестве критерия существенности r оказывается невозможным в силу того, что распределение выборочного r может значительно отличаться от нормального.

2. Для малого объема выборочной совокупности используется тот факт, что величина $t_{\text{расч.}}=\frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$ при условии $r=0$, распределена по закону Стьюдента с (n-2) степенями свободы.

Полученную величину $t_{расч.}$ сравнивают с табличным значением t-критерия (число степеней свободы равно $n-2$). Если рассчитанная величина $t_{расч.}$ превосходит табличное значение критерия t, то практически невероятно, что найденное значение обусловлено только случайными совпадениями x и y в выборке из генеральной совокупности, для которой действительное значение коэффициента корреляции равно нулю. Если же вычисленная величина $t_{расч.}$ меньше, чем в таблице, то полагают, что коэффициент корреляции в генеральной совокупности в действительности равен нулю и соответственно эмпирический коэффициент корреляции существенно не отличается от нуля.

Применим указанный метод к оценке существенности корреляции между уровнем затрат туристических фирм на рекламу и числом туристов, воспользовавшихся услугами фирм. При объеме выборки, равном 20 и при условии, что величина коэффициента корреляции равна 0,8105,

$$t_{расч.} = \frac{0{,}8105}{\sqrt{1-0{,}8105^2}}\sqrt{20-2} = \frac{0{,}8105 \cdot 4{,}243}{0{,}5857} = 5{,}871$$

В таблице для числа степеней свободы $k = n-2 = 18$ и уровня значимости 1% находим, что $t = 2{,}878$ (см. Приложение IV).

Таким образом лишь с вероятностью меньшей 1% можно утверждать, что величина t=5,871 могла появиться в силу случайностей выборки. Такое событие является маловероятным, а потому можно считать с вероятностью 99%, что в генеральной совокупности действительно существует прямая зависимость между изучаемыми признаками, т.е. отличие выборочного коэффициента корреляции от нуля является существенным.

3. Проверку гипотезы об отсутствии связи можно сделать и без вычислений, пользуясь таблицей, составленной Р. Фишером. В этой таблице показывается величина коэффициента корреляции, которая может считаться существенной при данном количестве наблюдений. При пользовании этой таблицей величину коэффициента корреляции следует искать для числа степеней свободы, равного n-2.

Краткую выдержку из таблицы [1] значений коэффициентов корреляции при различных уровнях критерия значимости приводим в табл. 7.5. В представленном примере коэффициент корреляции для оценки тесноты связи между признаками был рассчитан по 20 данным. По табл. 7.5 находим, что коэффициент корреляции по данным выборки должен быть по крайней мере не ниже 0,5614 для того, чтобы он мог считаться существенным при уровне значимости $\alpha = 0{,}01$.

[1] *Миллс Ф.* Статистические методы. М., Госстатиздат., 1956, табл. IV Приложения. С. 776.

Таблица 7.5

n - 2	$\alpha = 0{,}05$	$\alpha = 0{,}02$	$\alpha = 0{,}01$
4	0,8114	0,8822	0,9172
8	0,6319	0,7155	0,7646
10	0,5760	0,6581	0,7079
13	0,5139	0,5923	0,6411
18	0,4438	0,5155	0,5614
20	0,4227	0,4921	0,5368
25	0,3809	0,4451	0,4869
30	0,3494	0,4093	0,4487
40	0,3044	0,3578	0,3972
50	0,2732	0,3218	0,3541
60	0,2500	0,2948	0,3248
70	0,2319	0,2737	0,3017
80	0,2172	0,2565	0,2830
90	0,2050	0,2422	0,2673
100	0,1946	0,2321	0,2540

При уровне значимости α=0,05 мы могли бы считать существенной действительную связь при коэффициенте корреляции, равном или более 0,4438. По расчету линейный коэффициент корреляции получился равным 0,8105. Сравнение расчетного и табличных значений линейного коэффициента корреляции дает основание предполагать действительное наличие прямой связи между изучаемыми признаками в генеральной совокупности.

4. В тех случаях, когда линейный коэффициент корреляции, полученный по данным относительно малой выборки, близок к единице ($r \geqslant 0{,}8$), для проверки его существенности рекомендуется использовать метод преобразованной корреляции, предложенный Р.Фишером.

Р. Фишер показал, что распределение логарифмической функции линейного коэффициента корреляции (обозначается эта функция Z') приближается к нормальной кривой даже для выборок очень небольшого объема:

$$Z' = \frac{1}{2}\left[\ln(1+r) - \ln(1-r)\right] = \frac{1}{2}\ln\frac{1+r}{1-r}.$$

Средняя квадратическая ошибка Z'-распределения зависит только от объема выборки и определяется по формуле:

$$\sigma_z = \frac{1}{\sqrt{n-3}}. \qquad (7.6)$$

По таблице соотношений между r и Z' (см. Приложение VII), дающей возможность избежать вычислений логарифмов, находим, что коэффициенту корреляции 0,8105 соответствует Z' = 1,13. Для числа наблюдений n = 20, $\sigma_z = \frac{1}{\sqrt{20-3}} = \frac{1}{4,123} = 0,243$. Отношение Z' к средней квадратической ошибке σ_z равно 4,65 (1,13 : 0,243). Поскольку отношение $\frac{z'}{\sigma_z}$ оказалось больше трех, можно полагать действительное наличие связи между признаками в генеральной совокупности.

Коэффициент корреляции достаточно точно оценивает степень тесноты связи лишь в случае наличия линейной зависимости между признаками. При наличии же криволинейной зависимости линейный коэффициент корреляции недооценивает степень тесноты связи и даже может быть равен 0, а потому в таких случаях рекомендуется использовать в качестве показателя степени тесноты связи **эмпирическое корреляционное отношение** η [1].

Расчет корреляционного отношения основан на использовании известной теоремы сложения дисперсий. Общая дисперсия результативного признака σ_0^2 может быть разложена на две составляющие. Первая составляющая - межгрупповая дисперсия δ^2, характеризует ту часть колеблемости результативного признака, которая складывается под влиянием изменения признака-фактора, положенного в основу группировки.

$$\delta^2 = \frac{\sum_{j=1}^{k}\left(\bar{y}_j - \bar{y}_0\right)^2 n_j}{\sum_{j=1}^{k} n_j},$$

где $\bar{y}_j$ - средние значения результативного признака в соответствующих группах;

$\bar{y}_0$ - общая средняя для всей совокупности;

n_j - число наблюдений в соответствующей группе;

k - число выделенных групп.

[1] Показатель был предложен в 1896 г. К.Пирсоном.

Вторая составляющая - средняя из внутригрупповых дисперсий - $\overline{\sigma}^2$ оценивает ту часть вариации результативного признака, которая обусловлена действием других, «случайных» причин.

$$\overline{\sigma}^2 = \frac{\sum_{1}^{k}\sigma_j^2 n_j}{\sum_{1}^{k} n_j},$$

где σ_j^2 - дисперсия результативного признака в соответствующей группе.

Общая дисперсия равна: $\sigma_0^2 = \delta^2 + \overline{\sigma}^2$. Зная общую и межгрупповую дисперсии, можно оценить ту долю, которую составляет вариация под действием фактора x в общей вариации результативного признака y, т.е. найти отношение $\frac{\delta^2}{\sigma_0^2}$.

Извлекая квадратный корень из этого отношения, мы и получим эмпирическое корреляционное отношение

$$\eta = \sqrt{\frac{\delta^2}{\sigma_0^2}}. \qquad (7.7)$$

Величина корреляционного отношения может быть рассчитана и по следующей формуле:

$$\eta = \sqrt{1 - \frac{\overline{\sigma}^2}{\sigma_0^2}}. \qquad (7.7а)$$

Величина корреляционного отношения будет равна нулю, когда нет колеблемости в величине средних по выделенным группам. В тех случаях, когда внутригрупповая дисперсия близка к нулю, т.е. практически вся вариация результативного признака обусловлена действием фактора x, величина корреляционного отношения близка к 1. Направление связи мы легко устанавливаем по данным групповой и корреляционной таблиц.

Воспользуемся данными групповой таблицы 7.3 для расчета величины корреляционного отношения.

Таблица 7.6

Группы туристических фирм по затратам на рекламу, усл.ден.ед.	Число фирм в группе n_j	Среднее количество, воспользовавшихся услугами данной группы фирм, человек $\overline{y}_i$	$(\overline{y}_i - \overline{y}_i)^2 n_i$
1	2	3	4
8	3	790,0	79 218, 75
9	5	860,0	42 781, 25
10	5	966,0	911, 25
11	4	1062,5	48 400, 00
12	3	1100,0	65 268, 75
Итого	20	952,5	236 580, 00

Используем итоговые данные графы 4 таблицы для расчета межгрупповой дисперсии:

$$\delta^2 = \frac{236\ 580}{20} = 11\ 829 .$$

Среднее количество туристов, воспользовавшихся услугами одной фирмы, по всем 20 фирмам составило $\overline{y}_0 = 952{,}5$ человек.

Величина общей дисперсии результативного показателя составит:

$$\sigma_0^2 = \overline{y}_i^2 - \left(\overline{y}_0\right)^2 = \frac{18\ 497\ 700^1}{20} - 907\ 256{,}25 = 17\ 628{,}75 .$$

Следовательно, величина корреляционного отношения, по данным приводимого примера, будет равна 0,819:

$$\eta = \sqrt{\frac{11\ 829}{17\ 628{,}75}} = 0{,}819 .$$

Значимость рассчитанного корреляционного отношения оценивается с помощью дисперсионного отношения[2]:

$$F_{расч.} = \frac{\sum\limits_{j=1}^{k}\left(\overline{y}_i - \overline{y}_0\right)^2 n_j}{k-1} : \frac{\sum\limits_{j=1}^{k}\sum\limits_{i=1}^{n_j}\left(y_{ij} - \overline{y}_i\right)^2}{n-k} .$$

Если фактор, положенный в основу группировки туристических фирм, не оказывает влияния на вариацию изучаемого при-

[1] Указанная сумма квадратов рассчитана по первичным данным, представленным в табл. 7.1.

[2] См. параграф 6.8 гл. 6.

знака, то дисперсия групповых средних будет отражать влияние тех же прочих факторов, которые определяют и вариацию внутри групп, а потому отношение дисперсий будет близко к 1 или отличаться от нее в силу наличия случайных колебаний. В таблицах F-распределения указываются предельные значения F-критерия для различных комбинаций числа степеней свободы $k_1=k-1$ и $k_2=n-k$, которые могут быть превзойдены с вероятностью 0,05 или 0,01 в силу случайных обстоятельств.

В приводимом примере расчетное значения F-критерия будет равно 8,926:

$$F_{расч.} = \frac{236\,580}{5-1} : \frac{99\,395}{20-5} = 8{,}926$$

($\sum\limits_{j=1}^{k}\sum\limits_{i=1}^{n_j}\left(y_{ij} - \overline{y}_i\right)^2 = 99\ 395$ - эта величина рассчитана по данным табл. 7.1 и 7.3; $n = 20$; $k = 5$).

Табличное значение F-критерия при пятипроцентном уровне значимости и числе степеней свободы $k_1=4$ и $k_2=15$ равно 3,06 (см. Приложение VI). Таким образом расчетное значение F-критерия больше табличного, что позволяет с вероятностью 95% утверждать существенность различий в величине дисперсий и соответственно делать вывод о существенности корреляционной связи между анализируемыми показателями.

Следует отметить, что вычисление корреляционного отношения возможно лишь при наличии достаточно большого числа данных, которые представлены либо в форме корреляционной, либо в форме групповой таблицы. Вычисление корреляционного отношения при большом числе групп и малом числе наблюдений в каждой группе лишается смысла.

При недостаточном количестве данных в выделенных группах к рассчитанной величине корреляционного отношения должна быть внесена поправка на группировку

$$\eta^2_{скор.} = 1 - \left(1 - \eta^2\right)\left(\frac{n-1}{n-k}\right), \tag{7.8}$$

где k - число выделенных групп.

Для приводимого нами примера

$$\eta^2_{скор.} = 1 - (1 - 0{,}6708)\left(\frac{20-1}{20-5}\right) = 0{,}5830 \text{, откуда } \eta_{скор.} = 0{,}7635.$$

Определенный интерес представляет сопоставление величины линейного коэффициента корреляции и корреляционного отношения. Сравнив полученную величину корреляционного отношения с абсолютной величиной линейного коэффициента корреляции $|r| = 0{,}8105$, полученной при расчете по несгруппированным данным, можно видеть, что η незначительно больше r. Когда связь между переменными уклоняется от линейной формы, то η и r несколько отличаются по величине, причем η всегда больше r по абсолютной величине [1].

Сопоставление линейного коэффициента корреляции и эмпирического корреляционного отношения имеет смысл только в случае, если эти показатели вычислены для одинаковым образом сгруппированных данных, т.е. при сравнении и коэффициент корреляции и корреляционное отношение должны быть вычислены либо по данным корреляционной таблицы, либо по первичным данным и групповой таблице, что предпочтительнее.

При проверке возможности использования линейной функции в качестве формы уравнения определяют разность квадратов $\eta^2 - r^2$, и если эта разность менее 0,1, то считается возможным применять линейное уравнение корреляционной зависимости. В нашем случае разность квадратов корреляционного отношения и линейного коэффициента корреляции равна 0,060.

$\eta^2 - r^2 = 0{,}819^2 - 0{,}8105^2 = 0{,}0139(0{,}6708 - 0{,}6569)$, что меньше 0,1.

Для проверки гипотезы о линейной зависимости более эффективно использовать величину w^2:

$$w^2 = \frac{\eta^2 - r^2}{k-2} : \frac{1-\eta^2}{n-k}, \tag{7.9}$$

которая подчиняется закону F-распределения с числом степеней свободы числителя $(k\text{-}2)$ и знаменателя $(n\text{-}k)$.

Задаваясь достаточно малым уровнем значимости (например, $\alpha = 0{,}05$), находим по таблицам F-распределения значение $F_{табл.}$ при заданной величине α и соответствующем числе степеней свободы. Для рассматриваемого примера табличное значение F при 5%-ом уровне значимости и числе степеней свободы 3 $(k-2=5-2)$ и 15 $(n-k=20-5)$ равно 3,29.

Если w^2 окажется больше табличного значения F, то гипотезу о линейном виде регрессии можно считать статистически необоснованной.

[1] Доказательство этого положения приведено в книге: Дж. Эдни, М. Дж. Кендэл «Теория статистики» (М., Госстатиздат. 1960. С. 300).

В нашем примере $w^2 = \frac{0{,}6708-0{,}6569}{5-2} : \frac{1-0{,}6708}{20-5} = 0{,}2105$. Сопоставление полученной величины с табличным значением $F_{0{,}05(3;15)} = 3{,}29$ не позволяет отклонить гипотезу о линейности связи между затратами фирмы на рекламу и количеством туристов, воспользовавшихся ее услугами.

Применение линейного коэффициента корреляции для оценки степени тесноты связи между признаками особенно в той части, которая связана с оценкой его существенности, является обоснованным лишь в условиях нормального или близкого к нормальному распределению признаков в изучаемой совокупности. Кроме того, как видно из приводимых выше формул, для определения величины линейного коэффициента корреляции необходимо знать численные значения факторного и результативного признаков. В некоторых же случаях мы можем встретиться с такими качествами, которые не поддаются выражению числом единиц.

Эти обстоятельства заставляют прибегать к использованию так называемых непараметрических методов, позволяющих измерить интенсивность связи как между количественными признаками, форма распределения которых отличается от нормальной, так и между качественными признаками. В основу «непараметрических» методов положен принцип нумерации значений статистического ряда. Каждой единице совокупности присваивается порядковый номер в ряду, который будет упорядочен по уровню признака. Таким образом, ряд значений признака ранжируется, а номер каждой отдельной единицы будет ее рангом.

Можно получить предварительное представление о наличии или отсутствии связи между признаками, если сопоставить последовательность взаимного расположения рангов факторного и результативного признаков. Для этого ранги индивидуальных значений факторного признака располагают в порядке возрастания, и если ранги результативного признака обнаруживают тенденцию к увеличению, можно предполагать наличие прямой связи; если же с увеличением рангов факторного признака ранги результативного признака уменьшаются, то это свидетельствует о возможном наличии между изучаемыми признаками обратной связи.

Коэффициенты корреляции, основанные на использовании рангов, были предложены К. Спирмэном и М.Кендэлом. **Коэф-**

фициент корреляции рангов Спирмэна (был использован им в начале XX в.) основан на рассмотрении разности рангов значений факторного и результативного признаков.

Рассчитаем дисперсию σ_d^2 переменной d, соответствующей разности между рангами переменной x и y, т.е. $d_i = x_i - y_i$, где $i = 1, 2, 3, ..., n$.

$$\sigma_d^2 = \frac{\sum_{i=1}^{n} d_i^2}{n} - \left(\overline{d}\right)^2, \text{ где } \overline{d} = \overline{x} - \overline{y}.$$

Подставим вместо d_i величины $x_i - y_i$ и вместо $\overline{d}$ разность $\overline{x} - \overline{y}$.

Тогда $$\sigma_d^2 = \frac{\sum_{i=1}^{n}(x_i - y_i)^2}{n} - \left(\overline{x} - \overline{y}\right)^2 = \sigma_x^2 + \sigma_y^2 - 2\left(\frac{1}{n}\sum_{i=1}^{n} x_i y_i - \overline{x}\ \overline{y}\right).$$

Как известно, линейный коэффициент корреляции можно определить по формуле

$$r_{xy} = \frac{\frac{1}{n}\sum_{i=1}^{n} x_i y_i - \overline{x}\ \overline{y}}{\sigma_x \sigma_y}.$$

Тогда последнее слагаемое можно записать так:

$$2\left(\frac{1}{n}\sum_{1}^{n} x_i y_i - \overline{x}\ \overline{y}\right) = 2r_{xy} \cdot \sigma_x \sigma_y,$$

откуда $\sigma_d^2 = \sigma_x^2 - 2r_{xy}\sigma_x\sigma_y + \sigma_y^2$ и

$$r_{xy} = \frac{\sigma_x^2 + \sigma_y^2 - \sigma_d^2}{2\sigma_x \sigma_y}. \tag{7.10}$$

Применительно к рангам значений факторного и результативного признаков σ_x^2 и σ_y^2 характеризуют дисперсию натурального ряда чисел от 1 до n, которая составляет $\frac{1}{12}(n^2 - 1)$, а $\overline{x} = \overline{y} = \frac{n+1}{2}$.

Отсюда дисперсия разности рангов

$$\sigma_d^2 = \frac{1}{n}\sum_{1}^{n}(x_i - y_i)^2 - \left(\overline{x} - \overline{y}\right)^2 = \frac{1}{n}\sum_{1}^{n}(x_i - y_i)^2.$$

Подставим величины σ_d^2 и $\sigma_x^2 = \sigma_y^2$ в формулу 7.10 и получим формулу коэффициента корреляции рангов Спирмэна, который обозначают ρ:

$$\rho = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n\left(n^2 - 1\right)}. \tag{7.11}$$

Рассмотрим определение коэффициента Спирмэна на следующем примере.

Эксперты аналитического центра оценивали шансы депутатов на этапе предвыборной компании в городскую Думу следующим образом (см. табл. 7.7 графы 1 и 2).

Таблица 7.7

Порядковый номер кандидата в депутаты	Ранг кандидатов по результатам оценки экспертов	Ранг депутата по числу поданных голосов на выборах	Разница рангов (гр. 2 – гр. 3) $\lvert d_i \rvert$	d_i^2
1	2	3	4	5
1	7	5	2	4
2	4	6	2	4
3	1	2	1	1
4	3	7	4	16
5	10	8	2	4
6	5	3	2	4
7	9	10	1	1
8	2	1	1	1
9	8	9	1	1
10	6	4	2	4
Итого				40

По результатам выборов можно проранжировать депутатов по числу поданных за них голосов избирателей (графа 3 табл. 7.7). Как видим из результатов сопоставлений рангов до выборов и после них, шансы депутатов не всегда оценивались экспертами с достаточной степенью точности, причем в одних случаях более точно, в других точность оценки была существенно ниже (например, у депутата под номером 4). Возникает вопрос, насколько точно результаты экспертной оценки предугадали действительные шансы депутатов быть избранными в городскую Думу?

Для ответа на поставленный вопрос рассчитаем коэффициент корреляции рангов Спирмэна, используя результаты расчетов в графе 5 табл. 7.7.

$$\rho = 1 - \frac{6 \cdot 40}{10\,(100-1)} = 0{,}758\,.$$

Поскольку коэффициенты корреляции рангов могут изменяться в пределах от –1 до +1 (как и линейный коэффициент корреляции), по результатам расчетов коэффициента Спирмэна можно предположить наличие достаточно тесной прямой зависимости между оценками экспертов на стадии предвыборной компании и результатами выборов. Однако нельзя не учесть то обстоятельство, что ранговый коэффициент был рассчитан по небольшому объему исходной информации (n=10). Не является ли отличие рангового коэффициента от нуля лишь результатом случайных совпадений оценок экспертов с результатами выборов по данным малого числа отобранных депутатов, можно ли распространить полученные выводы на генеральную совокупность?

Для совокупностей небольшого объема ($n \leqslant 30$) распределение рангового коэффициента корреляции не является нормальным и нецелесообразно использовать значения t по нормированной функции Лапласа для проверки гипотезы о величине рангового коэффициента корреляции. В Приложении VIII приводится таблица предельных значений коэффициентов корреляции рангов Спирмэна при условии верности нулевой гипотезы об отсутствии корреляционной связи при заданном уровне значимости и определенном объеме выборочных данных.

По таблице Приложения VIII находим, что при объеме выборки в 10 единиц (n=10) и уровне значимости 5% (α=0,05) критическая величина для рангового коэффициента корреляции составляет ±0,6364. Это означает, что вероятность получить величину коэффициента ρ, превышающую критическое значение при условии верности нулевой гипотезы ($H_0 : \rho = 0$), будет менее 5%. В силу малой вероятности такое событие считается практически невозможным, и нулевая гипотеза может быть отвергнута.

Поскольку по результатам расчетов ρ=0,758, что превышает критическую величину рангового коэффициента корреляции, можно принять альтернативную гипотезу о совпадении результатов выборов с оценками экспертов. Однако при уровне значимости α=0,01 критическое значение рангового коэффициента, которое может быть обусловлено случайными совпадениями рангов, составляет 0,7818 (см. Приложение VIII). В этом случае с вероятностью 99% нулевая гипотеза не может быть отвергнута, т.е. величина ρ=0,758 могла быть результатом случайных совпадений рангов обследованных депутатов, тогда как в генеральной совокупности связь между оценками экспертов и результатами выборов может отсутствовать. Поэтому общий вывод по результатам анализа может состоять в необходимости проведения расчетов по большему числу депутатов, а при отсутствии такой возможности - относиться к оценкам экспертов данного аналитического центра с достаточной осторожностью.

М.Кендэл предложил еще одну меру связи между переменными x и y – **коэффициент корреляции рангов Кендэла** – τ:

$$\tau = \frac{2S}{n(n-1)}\,, \quad \text{где } S = P + Q\,. \qquad (7.12)$$

Для вычисления τ надо упорядочить ряд рангов переменной x, приведя его к ряду натуральных чисел. Затем рассматривают последовательность рангов переменной y (см. табл. 7.8).

Таблица 7.8

Ранг депутатов по экспертной оценке (x)	1	2	3	4	5	6	7	8	9	10
Ранг депутатов по результатам выборов (y)	2	1	7	6	3	4	5	9	10	8

Для нахождения суммы S (формула 7.12) находят два слагаемых P и Q. При определении слагаемого P нужно установить, сколько чисел, находящихся справа от каждого из элементов последовательности рангов переменной y, имеют величину ранга, превышающую ранг рассматриваемого элемента. Так, например, первому значению в последовательности рангов переменной y, т.е. числу 2, соответствует 8 чисел (7, 6, 3, 4, 5, 9, 10, 8), которые превышают ранг 2; второму значению 1 соответствует также 8 чисел (7, 6, 3, 4, 5, 9, 10, 8); превышающих 1 и т.д. Суммируя полученные таким образом числа, мы получим слагаемое P, которое можно рассматривать как меру соответствия последовательности рангов переменной y последовательности рангов переменных x. Для нашего примера P = 35 (8+8+3+3+5+4+3+1).

Второе слагаемое Q характеризует степень несоответствия последовательности рангов переменной y последовательности рангов переменной x. Чтобы определить Q подсчитаем, сколько чисел, находящихся справа от каждого из членов последовательности рангов переменной y имеет ранг меньше, чем эта единица. Такие величины берутся со знаком минус.

В рассматриваемом примере

Q = -10 (-1-0-4-3-0-0-0-1-1)

Следовательно, $S = P + Q = 35 - 10 = 25$.

Коэффициент корреляции рангов Кендэла в нашем примере равен:

$$\tau = \frac{2S}{n\,(n-1)} = \frac{2 \cdot 25}{10 \cdot 9} = 0{,}556 .$$

Коэффициент Кендэла также изменяется в пределах от −1 до +1 и равен нулю при отсутствии связи между рядами рангов.

При достаточно большом числе наблюдений между коэффициентами корреляции рангов Спирмэна и коэффициентом корреляции рангов Кендэла существует следующее соотношение:

$$\rho = \frac{3}{2}\,\tau .$$

Существенность коэффициента корреляции рангов Кендэла проверяется при уровне значимости α по формуле

$$\tau > t_\alpha \sqrt{\frac{2\,(2n+5)}{9n\,(n-1)}} \qquad (7.13)$$

где t_α - коэффициент, определяемый по таблице нормального распределения для выбранного уровня значимости α при больших n.

Могут встретиться случаи, когда невозможно установить ранговые различия нескольких смежных значений. В этих случаях принято брать средний ранг (даже если он будет дробным числом) и полученный средний ранг приписывать каждому из таких значений, т.е. говорят, что переходят к матрице переформированных рангов. Например, двум факторам один из экспертов приписывает одинаковый ранг 3. Тогда каждому из факторов присваивается ранг (3+4) : 2 = 3,5, так как они поделили между собой третье и четвертое места, а фактору, имевшему ранг 4, присваивается ранг 5 и т.д.

Если определяется теснота связи между k-м и l-м признаками, в рядах значений которых имеется соответственно q и g групп объединенных рангов, то формула коэффициента корреляции рангов Спирмэна примет вид:

$$\rho = \frac{\dfrac{n^3-n}{6} - \left(T_k + T_l\right) - \sum_{i=1}^{n} d_i^2}{\sqrt{\left(\dfrac{n^3-n}{6} - 2T_k\right)\left(\dfrac{n^3-n}{6} - 2T_l\right)}} , \qquad (7.11a)$$

где

$$T_k = \sum_{i=1}^{q} \frac{t_{k_i}^3 - t_{k_i}}{12} \;;\quad T_l = \sum_{i=1}^{q} \frac{t_{l_i}^3 - t_{l_i}}{12} \;;$$

где t_{k_i} и t_{l_i} определяют количество единиц в i-й группе объединенных рангов соответствующего признака. Скорректированная формула для вычисления коэффициента корреляции рангов Кендэла будет иметь вид:

$$\tau = \frac{S}{\sqrt{\left(\dfrac{n\,(n-1)}{2} - V_k\right)\left(\dfrac{n\,(n-1)}{2} - V_l\right)}} , \qquad (7.12a)$$

где

$$V_k = \frac{\sum_{i=1}^{q} t_{k_i}\left(t_{k_i}-1\right)}{2} \;;\quad V_l = \frac{\sum_{i=1}^{q} t_{l_i}\left(t_{l_i}-1\right)}{2} .$$

Для оценки степени тесноты связи между несколькими признаками при использовании ранговой корреляции **применяется коэффициент конкордации** ω, который вычисляется по формуле:

$$\omega = \frac{12 S_1}{m^2\,(n^3-n)} , \qquad (7.14)$$

где m - число факторов;

n - число ранжируемых единиц;

S - сумма квадратов отклонений рангов.

Если обозначить r_{ij} ранг i-го фактора у j-й единицы, то величина S будет равна:

$$S = \sum_{1}^{n}\left(\sum_{1}^{m} r_{ij}\right)^2 - \frac{\left(\sum_{1}^{n}\sum_{1}^{m} r_{ij}\right)^2}{n} .$$

Рассмотрим вычисление коэффициента конкордации на следующем примере. Восемь предприятий, выпускающих однотипную продукцию, ранжированы экспертами по уровню качества выпускаемой продукции, спросу на продукцию и уровню рентабельности (см. табл. 7.9).

Сумма квадратов отклонений рангов равна 330:

$$S = 1788 - \frac{(108)^2}{8} = 330 .$$

Величина коэффициента конкординации получится равной 0,873:

$$\omega = \frac{12 \cdot 330}{3^2\left(8^3 - 8\right)} = 0{,}873 .$$

Это свидетельствует о возможном наличии достаточно тесной зависимости между изучаемыми признаками.

Таблица 7.9

Порядковый номер предприятия	Ранг по показателям (r_{ij})			$\sum_1^m r_{ij}$	$\left(\sum_1^m r_{ij}\right)^2$
	уровня рентабельности	уровня качества	уровня спроса		
1	4	4	3	11	121
2	1	3	1	5	25
3	3	1	2	6	36
4	7	6	5	18	324
5	5	5	7	17	289
6	6	8	6	20	400
7	2	2	4	8	64
8	8	7	8	23	529
Итого				108	1788

При исследовании степени тесноты связи между качественными признаками, каждый из которых представлен в виде альтернативного признака, используют **коэффициент ассоциации** или **коэффициент контингенции** [1]. Например, нужно оценить влияют ли существующие формы повышения квалификации преподавателей университета на уровень их профессионального мастерства. Располагая данными о результатах аттестации студентами 320 преподавателей, из которых 240 повысили квалификацию, составляем следующую таблицу:

[1] Коэффициент ассоциации предложен английским статистиком Д.Юлом, а коэффициент контингенции К.Пирсоном.

Таблица 7.10

Группы преподавателей	Средний балл по сравнению с предыдущим по результатам аттестации		Всего
	не изменился и возрос	снизился	
Повысившие квалификацию по одной из принятых форм	163 (a)	77 (b)	240
Не прошедшие повышение квалификации по принятым формам	46 (c)	34 (d)	80
Всего	209	111	320

Построенная в такой форме таблица носит название «таблицы четырех полей», частоты которой обозначим соответственно *a, b, c, d*.

Коэффициент ассоциации (K_A) определяется по формуле

$$K_A = \frac{ad - bc}{ad + bc} . \tag{7.15}$$

В приводимом примере его величина будет равна 0,22.

$$\left(\frac{163 \cdot 34 - 77 \cdot 46}{163 \cdot 34 + 77 \cdot 46} = +0{,}22\right).$$

Таким образом, по результатам проведенного в университете обследования вряд ли можно сделать убедительный вывод о повышении профессионального мастерства преподавателей в связи с повышением квалификации по одной из принятых форм (стажировка, факультет повышения квалификации, творческий отпуск и др.), поскольку степень тесноты связи невелика.

В тех случаях, когда хотя бы один из четырех показателей в таблице «четырех полей» отсутствует, величина коэффициента ассоциации будет равна единице, что дает преувеличенную оценку степени тесноты связи между признаками, и предпочтение следует отдать коэффициенту контингенции (K_k):

$$K_k = \frac{ad - bc}{\sqrt{(a+b)(b+d)(a+c)(c+d)}} . \tag{7.16}$$

В качестве критерия наличия связи между качественными показателями с большим числом градаций можно использовать **критерий «хи-квадрат»**. В табл. 7.11 приводятся результаты группировки предприятий отрасли по техническому и организационному уровню развития производства. Чисто визуально трудно ответить на вопрос о взаимосвязи между признаками, поэтому необходимо дать анализ распределения частот в таблице по строкам и графам.

Если признак, положенный в основу группировки по строкам таблицы (технический уровень) не зависит от признака, положенного в основу группировки по столбцам (организационный уровень), то в каждой строке (столбце) распределение частот должно быть пропорционально распределению их в итоговой строке (столбце). Такое распределение можно рассматривать в известной мере в качестве теоретического, частоты которого рассчитаны в предположении отсутствия связи между изучаемыми признаками.

Рассчитаем частоты внутри таблицы пропорционально распределению частот в итоговой строке (эти частоты приведены в табл. 7.11 в скобках):

Таблица 7.11

Распределение предприятий отрасли по техническому и организационному уровню развития

Группы предприятий по техническому уровню	Группы предприятий по организационному уровню			Итого
	ниже среднего уровня	среднего уровня	выше среднего уровня	
Ниже среднего уровня	7 (7,1)	8 (9,4)	10 (8,5)	25
Среднего уровня	4 (3,4)	7 (4,5)	1 (4,1)	12
Выше среднего уровня	4 (4,5)	5 (6,0)	7 (5,5)	16
Итого	15	20	18	53

$$f_{11}=\frac{25\cdot 15}{53}=7{,}1;\quad f_{12}=\frac{25\cdot 20}{53}=9{,}4;\quad f_{13}=\frac{25\cdot 18}{53}=8{,}5\,;$$

$$f_{21}=\frac{12\cdot 15}{53}=3{,}4;\quad f_{22}=\frac{12\cdot 20}{53}=4{,}5;\quad f_{23}=\frac{12\cdot 18}{53}=4{,}1 \text{ и т.д.}$$

Расчетное значение критерия хи-квадрат определится по формуле:

$$\chi^2=\sum_{i=1}^{k_2}\sum_{j=1}^{k_1}\frac{\left(f_{ij}-f_{ij}'\right)^2}{f_{ij}'}\,, \tag{7.17}$$

где f_{ij} и f_{ij}' - соответственно эмпирические и теоретические частоты в i-той строке j-того столбца;

k_1 и k_2 - соответственно число групп в строках и столбцах таблиц.

Расчетное значение критерия по данным табл. 7.11 составит 5,064

$$[\frac{(7-7,1)^2}{7,1}+\frac{(8-9,4)^2}{9,4}+\frac{(10-8,5)^2}{8,5}+\frac{(4-3,4)^2}{3,4}+\frac{(7-4,5)^2}{4,5}+\frac{(1-4,1)^2}{4,1}+$$

$$+\frac{(4-4,5)^2}{4,5}+\frac{(5-6,0)^2}{6,0}+\frac{(7-5,5)^2}{5,5}].$$

По таблицам математической статистики устанавливается либо вероятность появления рассчитанного значения χ^2, соответствующего данному числу степеней свободы в предположении независимости признаков; либо табличное значение критерия хи-квадрат, соответствующего уровню значимости 0,05 или 0,01. В рассматриваемом примере число степеней свободы равно четырем [k = (3-1)(3-1) = 4]. При уровне значимости 0,05 табличное значение χ^2 при $k=4$ составляет 9,5 (см. Приложение V), т. е. $\chi^2_{\text{расч.}} < \chi^2_{\text{табл.}}$ Следовательно гипотеза об отсутствии влияния технического уровня развития на организационный уровень не опровергается при α=0,05.

На основе критерия «хи-квадрат» определяются показатели степени тесноты связи: коэффициенты взаимной сопряженности К.Пирсона и А.А.Чупрова. **Коэффициент сопряженности** К.Пирсона рассчитывается по формуле:

$$P=\sqrt{\frac{\chi^2}{n+\chi^2}}\,, \tag{7.18}$$

где n - общее число наблюдений.

Коэффициент взаимной сопряженности А.А. Чупрова позволяет учесть число групп по каждому признаку и определяется следующим образом:

$$C=\sqrt{\frac{\chi^2}{n\sqrt{(k_1-1)(k_2-1)}}} \tag{7.19}$$

По данным приводимого в табл. 7.11 примера получим следующие значения коэффициентов сопряженности:

$$P=\sqrt{\frac{5{,}064}{53+5{,}064}}=0{,}295,$$

$$C=\sqrt{\frac{5{,}064}{53\sqrt{2\cdot 2}}}=0{,}219.$$

Таким образом, между техническим и организационным уровнем развития отмечается некоторая зависимость, но степень тесноты связи невелика.

7.4. Уравнение регрессии [1]

Изучение корреляционных зависимостей основывается на исследовании таких связей между переменными, при которых значения одной переменной, ее можно принять за зависимую переменную, «в среднем» изменяются в зависимости от того, какие значения принимает другая переменная, рассматриваемая как причина по отношению к зависимой переменной. Как уже было сказано, действие данной причины осуществляется в условиях сложного взаимодействия различных факторов, вследствие чего проявление закономерности затемняется влиянием случайностей. Вычисляя средние значения результативного признака для данной группы значений признака-фактора, мы отчасти элиминируем влияние случайностей. Вычисляя параметры теоретической линии связи, мы производим дальнейшее их элиминирование и получаем однозначное (по форме) изменение y с изменением фактора x.

Теоретической линией регрессии называется та линия, вокруг которой группируются точки корреляционного поля и которая указывает основное направление, основную тенденцию связи. Теоретическая линия регрессии должна отображать изменение средних величин результативного признака y по мере изменения величин факторного признака x при условии полного взаимопогашения всех прочих - случайных по отношению к фактору x - причин. Следовательно, эта линия должна быть проведена так, чтобы сумма отклонений точек поля корреляции от соответствующих точек теоретической линии регрессии равнялась нулю, а сумма квадратов этих отклонений была бы минимальной величиной.

[1] Термин «регрессия» был впервые использован Фрэнсисом Гальтоном в 1877 г.

Важным этапом регрессионного анализа является определение типа функции, с помощью которой характеризуется зависимость между признаками. Главным основанием для выбора вида уравнения должен служить содержательный анализ природы изучаемой зависимости, ее механизма. Вместе с тем теоретически обосновать форму связи каждого из факторов с результативным показателем можно далеко не всегда, поскольку исследуемые социально-экономические явления очень сложны и факторы, формирующие их уровень, тесно переплетаются и взаимодействуют друг с другом. Поэтому на основе теоретического анализа нередко могут быть сделаны самые общие выводы относительно направления связи, возможности его изменения в исследуемой совокупности, правомерности использования линейной зависимости, возможного наличия экстремальных значений и т.п. Необходимым дополнением такого рода предположений должен быть анализ конкретных фактических данных.

Приблизительное представление о линии связи можно получить на основе эмпирической линии регрессии (или линии групповых средних) см. рис. 7.1. Эмпирическая линия обычно является ломаной линией, имеет более или менее значительный излом. Объясняется это тем, что влияние прочих неучтенных факторов, оказывающих воздействие на вариацию результативного признака, в средних погашается не полностью, в силу недостаточно большого количества наблюдений, поэтому эмпирической линией связи для выбора и обоснования типа теоретической кривой можно воспользоваться при условии, что число наблюдений будет достаточно велико.

Можно также использовать опыт предыдущих исследований и там, где выбранные формы уравнений связи давали удовлетворительный результат, можно рекомендовать их использовать и в дальнейшем.

Одним из элементов конкретных исследований является сопоставление различных уравнений зависимости, основанное на использовании критериев качества аппроксимации эмпирических данных конкурирующими вариантами моделей. Наиболее часто для характеристики связей экономических показателей используют следующие типы функций:

линейную $\hat{y} = a + bx$,

гиперболическую $\hat{y} = a + b\frac{1}{x}$,

показательную $\hat{y} = ab^x$,

параболическую $\hat{y} = a + bx + cx^2$,

степенную $\hat{y} = ax^b$,

логарифмическую $\hat{y} = a + b\lg x$,

логистическую $\hat{y} = \frac{d}{1 + e^{a+bx}}$.

В рассматриваемом примере эмпирическая линия регрессии все же больше всего приближается к прямой, и, следовательно, теоретическая линия регрессии может быть представлена уравнением вида:

$$\hat{y} = a + bx.$$

Для нахождения параметров a и b уравнения регрессии используем метод наименьших квадратов. При применении метода наименьших квадратов для нахождения такой функции, которая наилучшим образом соответствует эмпирическим данным, считается, что сумма квадратов отклонений эмпирических точек от теоретической линии регрессии должна быть величиной минимальной.

Критерий метода наименьших квадратов можно записать таким образом:

$$S = \sum_{i=1}^{n}\left(y_i - \hat{y}(x)\right)^2 \longrightarrow \min$$

или, поскольку $\hat{y}_x = a + bx$,

$$S = \sum_{1}^{n}\left[y_i - (a + bx)\right]^2 \longrightarrow \min.$$

Следовательно, применение метода наименьших квадратов для определения параметров a и b прямой, наиболее соответствующей эмпирическим данным, сводится к задаче на экстремум.

Функция двух переменных $S(a,b)$ может достигнуть экстремума в том случае, когда первые частные производные этой функции равняются нулю, т.е. когда

$$\frac{dS}{da} = 0 \quad \text{и} \quad \frac{dS}{db} = 0.$$

Вычисляя эти частные производные, получим:

$$\frac{dS}{da} = -2\sum_{1}^{n}\left(y_i - a - bx_i\right) = 0,$$

$$\frac{dS}{db} = -2\sum_{1}^{n}\left(y_i x_i - ax_i - bx_i^2\right) = 0.$$

После несложных преобразований получим систему нормальных уравнений способа наименьших квадратов для определения величины параметров a и b уравнения прямолинейной корреляционной связи по эмпирическим данным:

$$\begin{cases} \sum_{1}^{n} y_i = an + b\sum_{1}^{n} x_i \\ \sum_{1}^{n} y_i x_i = a\sum_{1}^{n} x_i + b\sum_{1}^{n} x_i^2 \end{cases} \qquad (7.20)$$

Используя результаты расчетов на с. 235, можно записать для нашего примера систему нормальных уравнений:

$$\begin{cases} 20a + 199b = 19\,050 \\ 199a + 2013b = 192\,310 \end{cases}$$

Если данные сгруппированы (например, представлены в виде корреляционной таблицы), то система нормальных уравнений имеет вид:

$$\begin{cases} \sum_{1}^{k} y_i f_y = a\sum_{1}^{k} f_{xy} + b\sum_{1}^{k} x_i f_x \\ \sum_{1}^{k} y_i x_i f_{xy} = a\sum_{1}^{k} x_i f_x + b\sum_{1}^{k} x_i^2 f_x, \end{cases}$$

где f_y - частота повторения данного варианта значения y;

f_x - частота повторения данного варианта значения x;

f_{xy} - частота повторения данного сочетания значений x и y.

Решая систему уравнений (7.20) относительно b, получим следующую формулу для определения этого параметра:

$$b = \frac{\sum_{i=1}^{n} x_i y_i - n\,\bar{x}\,\bar{y}}{\sum_{i=1}^{n} x_i^2 - n\left(\bar{x}\right)^2} \qquad (7.20а)$$

Значение параметра a получим, разделив обе части первого уравнения в формуле (7.20) на n:

$$a = \bar{y} - b\,\bar{x} \qquad (7.20б)$$

В результате решения получим:

$$a = 118{,}3 \quad b = 83{,}84 \quad и \quad \hat{y} = 118{,}3 + 83{,}84x\,.$$

Графическое изображение эмпирической и теоретической линии связи представлено на рис. 7.1.

Параметр b в уравнении называют **коэффициентом регрессии**. При наличии прямой корреляционной зависимости коэффициент регрессии имеет положительное значение, а в случае обратной зависимости коэффициент регрессии – отрицательный.

Коэффициент регрессии показывает, на сколько в среднем изменяется величина результативного признака y при изменении факторного признака x на единицу. Геометрически коэффициент регрессии представляет собой наклон прямой линии, изображающей уравнение корреляционной зависимости, относительно оси x (для уравнения $\hat{y} = a + bx$).

Коэффициент регрессии применяют для определения **коэффициента эластичности**, который показывает, на сколько процентов изменится величина результативного признака y при изменении признака-фактора x на один процент.

Для определения коэффициента эластичности используется формула:

$$Э_x = b \cdot \frac{\bar{x}}{\bar{y}} \qquad (7.21)$$

Для нашего примера коэффициент эластичности будет равен:

$$Э_x = 83{,}84 \cdot \frac{9{,}95}{952{,}5} = 0{,}8758\,.$$

Это означает, что при росте затрат фирмы на рекламу на 1%, количество туристов, воспользовавшихся услугами фирмы, возрастет на 0,8758%.

Зная линейный коэффициент корреляции, оценивающий степень тесноты связи между изменениями факторного и результативного признаков, можно определить коэффициент регрессии в уравнении $\hat{y} = a + bx$ по следующей формуле:

$$b = r\,\frac{\sigma_y}{\sigma_x}, \qquad (7.22)$$

где σ_y и σ_x - средние квадратические отклонения соответственно значений результативного и факторного признаков.

Наличие этого соотношения дает возможность производить вычисление коэффициента корреляции и параметров уравнения линейной регрессии одновременно.

Воспользуемся данными примера, приведенного в табл. 7.1 для расчета параметров уравнения линейной зависимости.

Расчет показателей по несгруппированным данным привел к таким результатам:

$$\sigma_x = 1{,}2836;\ \sigma_y = 132{,}77;\ r = 0{,}8105,\ тогда\ b = 0{,}8105\frac{132{,}77}{1{,}2836} = 83{,}83.$$

Вопрос о целесообразности использования уравнения регрессии при решении задач обоснованного прогноза может быть проиллюстрирован на графике (см. рис. 7.2)

Значению факторного признака x_k соответствует по результатам наблюдения значение результативного признака y_k. Если воспользоваться уравнением регрессии для определения величины результативного признака, получим величину $\hat{y}_k$. Общая величина отклонения наблюдаемого значения y_k от средней $y_k - \bar{y}$ будет тогда представлена на графике двумя отрезками, которые соответствуют: отрезок 1 - отклонению теоретического значения $\hat{y}_k$ от средней величины результативного признака $\hat{y}_k - \bar{y}$ и представляет собой ту часть отклонения y от $\bar{y}$, которая может быть объяснена влиянием фактора x. Вторая часть общего отклонения $\left(y_k - \hat{y}_k\right)$, обозначенная на графике цифрой 2, характеризует остаточную часть общего отклонения и не может быть объяснена влиянием фактора x.

Использование уравнения регрессии приводит к получению лучших оценок для совокупности значений результативного признака по сравнению с ориентацией на общую среднюю, так как уменьшается величина отклонений: $\hat{y}_k - \bar{y} < y_k - \bar{y}$.

Если рассмотреть общую сумму квадратов отклонений результативного признака от средней, т.е. величину $\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2$, и сопоставить с ней меру «остаточной» вариации, т.е. величину

$\sum\left(y_i-\hat{y}_i\right)^2$, мы получим долю вариации, необъясняемой влиянием факторного признака *x*.

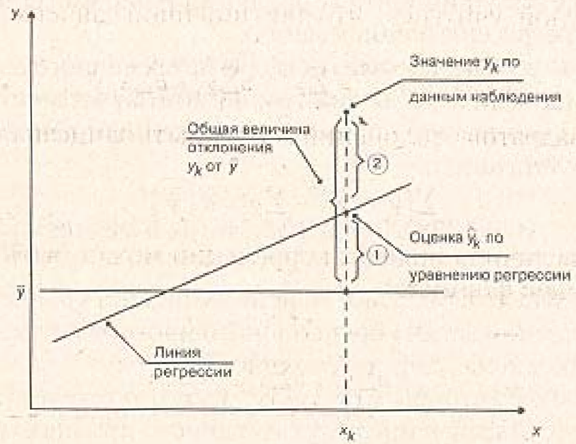

Рис. 7.2. **Соотношение между общей вариацией признака, его составляющей, объясняемой влиянием фактора *x*, и остаточной вариацией.** ① - величина $\left(\hat{y}_k-\bar{y}\right)$; ② - величина $\left(y_k-\hat{y}_k\right)$

Тогда измерителем того, насколько велика степень связи между вариацией факторного и результативного признака, будет величина:

$$1-\frac{\sum_{i=1}^{n}\left(y_i-\hat{y}_i\right)^2}{\sum_{i=1}^{n}\left(y_i-\bar{y}\right)^2}.$$

Извлекая квадратный корень из записанного выражения, получим показатель, называемый **индексом корреляции**:

$$i_{\hat{y}x}=\sqrt{1-\frac{\sum_{i=1}^{n}\left(y_i-\hat{y}_i\right)^2}{\sum_{i=1}^{n}\left(y_i-\bar{y}\right)^2}}. \qquad (7.23)$$

Для расчета индекса корреляции по приведенной формуле нужно предварительно для каждого значения факторного признака рассчитать по уравнению регрессии соответствующее ему значение результативного показателя $\hat{y}_i$. Поскольку такой порядок расчета весьма трудоемок, можно воспользоваться преобразованной формулой, учитывая, что для линейной зависимости:

$$\sum_{i=1}^{n}\left(y_i-\hat{y}_i\right)^2=\sum_{i=1}^{n}y_i^2-a\sum_{i=1}^{n}y_i-b\sum_{i=1}^{n}x_i y_i, \qquad (7.23а)$$

а сумма квадратов отклонений может быть записана так:

$$\sum_{i=1}^{n}\left(y_i-\bar{y}\right)^2=\sum_{i=1}^{n}y_i^2-n\left(\bar{y}\right)^2.$$

Тогда величина индекса корреляции может быть рассчитана по следующей формуле:

$$i_{\hat{y}x}=\sqrt{\frac{a\sum_{i=1}^{n}y_i+b\sum_{i=1}^{n}x_i y_i-n\left(\bar{y}\right)^2}{\sum_{i=1}^{n}y_i^2-n\left(\bar{y}\right)^2}}. \qquad (7.23б)$$

Из записи формул (7.23) и (7.23б) следует, что величина индекса корреляции зависит от того, какая форма уравнения регрессии была выбрана исследователем. А потому индекс корреляции в известной мере следует считать не показателем степени тесноты связи между изучаемыми признаками, а показателем степени близости выбранной теоретической линии регрессии к фактическим данным. Индекс корреляции может принимать значения в пределах от 0 до 1. Если $i_{\hat{y}x}$ равен или близок к нулю, это означает, что между переменными *x* и *y* либо нет связи, или если она и существует, то не может быть охарактеризована выбранной формой уравнения регрессии.

Для рассматриваемой зависимости числа туристов, воспользовавшихся услугами фирмы, и затратами фирмы на рекламу индекс корреляции составляет 0,8312:

$$i_{\hat{y}x}=\sqrt{\frac{118{,}3\cdot 19\,050+83{,}84\cdot 192\,310-20\cdot(952{,}5)^2}{18\,497\,700-20(952{,}5)^2}}=\sqrt{\frac{243\,590}{352\,575}}=0{,}8312.$$

Близость величины индекса корреляции к единице в общем случае означает, что связь между признаками достаточно хорошо описывается избранным уравнением зависимости. Для приводимого примера это означает, что выбор линейной функции в качестве уравнения регрессии достаточно обоснован.

Поскольку не все фактические значения результативного признака лежат на линии регрессии (см. рис. 7.1.), более справедливо для записи уравнения корреляционной зависимости воспользоваться формулой $y = a + bx + e$, где e отражает случайную составляющую вариации результативного признака, представленного на рис. 7.2. отрезком 2. Иногда рассеяние точек корреляционного поля настолько велико, что для принятия решений в управлении нет смысла пользоваться уравнением регрессии, так как погрешность в оценке анализируемого показателя будет чрезвычайно велика. Для всей совокупности наблюдаемых значений рассчитывается средняя квадратическая ошибка уравнения регрессии S_e, которая представляет собой среднее квадратическое отклонение фактических значений y_i, относительно значений, рассчитанных по уравнению регрессии $\hat{y}_i$, т.е.

$$S_e = \sqrt{\frac{\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2}{n-m}}, \qquad (7.24)$$

где S_e - средняя квадратическая ошибка уравнения регрессии;
y_i - фактические значения результативного признака, полученные по данным наблюдения;
$\hat{y}_i$ - значения результативного признака, рассчитанные по уравнению корреляционной связи и полученные подстановкой значений факторного признака x_i в уравнение регрессии $\hat{y} = a + bx$;
m - число параметров в уравнении регрессии.

Заметим, что в данной формуле сумма квадратов отклонений y_i от $\hat{y}_i$ делится на число степеней свободы $(n-m)$, поскольку мы связали себя m степенями свободы в оценке теоретических значений результативного признака по уравнению регрессии с m параметрами. В случае линейного уравнения регрессии m=2.

Используем формулу (7.23а) для упрощения расчета средней квадратической ошибки уравнения. В результате получим следующее выражение для определения величины S_e в случае линейной зависимости между x и y:

$$S_e = \sqrt{\frac{\sum_{i=1}^{n} y_i^2 - a\sum_{i=1}^{n} y_i - b\sum_{i=1}^{n} x_i y_i}{n-2}}, \qquad (7.24а)$$

где a и b являются параметрами уравнения регрессии.

Для рассматриваемого примера величина средней квадратической ошибки уравнения регрессии $\hat{y}$=118,3+83,84x будет равна 81,93 человек.

$$S_e = \sqrt{\frac{18\,497\,700 - 118{,}3 \cdot 19\,050 - 83{,}84 \cdot 192\,310}{20-2}} = 81{,}93 \text{ человек.}$$

Велика или мала полученная величина S_e?

Можно сопоставить ее величину со средним значением результативного признака $\bar{y}$ = 952,5 человек. Тогда $\frac{S_e}{\bar{y}} \cdot 100\%$ будет составлять 8,60%. Можно также сравнить величину S_e со средним квадратическим отклонением результативного признака σ_y. Если $S_e < \sigma_y$, использование уравнения регрессии является целесообразным.

Чем меньше рассеяние эмпирических точек вокруг прямой, тем меньше средняя квадратическая ошибка уравнения. Таким образом, величина S_e служит показателем значимости и полезности прямой, выражающей соотношение между двумя признаками.

Средняя квадратическая ошибка уравнения дает нам возможность в каждом отдельном случае с определенной вероятностью указать, что величина результативного признака окажется в определенном интервале относительно значения, вычисленного по уравнению связи.

Покажем определение доверительных границ для результативного признака, т.е. тех границ, в пределах которых с заданной доверительной вероятностью будет находиться теоретическое значение y. Поскольку параметры уравнения регрессии определяются по выборочным данным, являясь функцией наблюденных значений, оценки параметров a и b содержат некоторую погрешность. Дисперсия значения зависимой переменной, определяемой по уравнению линейной зависимости, будет складываться из дисперсии параметра a и дисперсии параметра b.

Зная дисперсию показателя $\hat{y}$ и задаваясь уровнем доверительной вероятности, определяем доверительные границы результативного признака при значении факторного признака x_0 следующим образом:

$$\hat{y}_{x_0} - t_\alpha \frac{S_e}{\sqrt{n}} \sqrt{1 + \frac{\left(x_0 - \bar{x}\right)^2}{\sigma_x^2}} \leq Y \leq \hat{y}_{x_0} + t_\alpha \frac{S_e}{\sqrt{n}} \sqrt{1 + \frac{\left(x_0 - \bar{x}\right)^2}{\sigma_x^2}}, \qquad (7.25)$$

где t_α - определяется в соответствии с уровнем значимости по t-распределению Стьюдента с $(n - m)$ степенями свободы.

Величина множителя $C_{x_0} = \sqrt{1 + \frac{\left(x_0 - \bar{x}\right)^2}{\sigma_x^2}}$ будет вычисляться

для каждого значения x_0. С удалением значения факторного признака от своего среднего арифметического значения величина C_{x_0} будет возрастать. И если представить доверительные границы графически, то они расположатся выше и ниже линии регрессии в виде ветвей гиперболы, ограничивая доверительную зону (см. рис. 7.3)

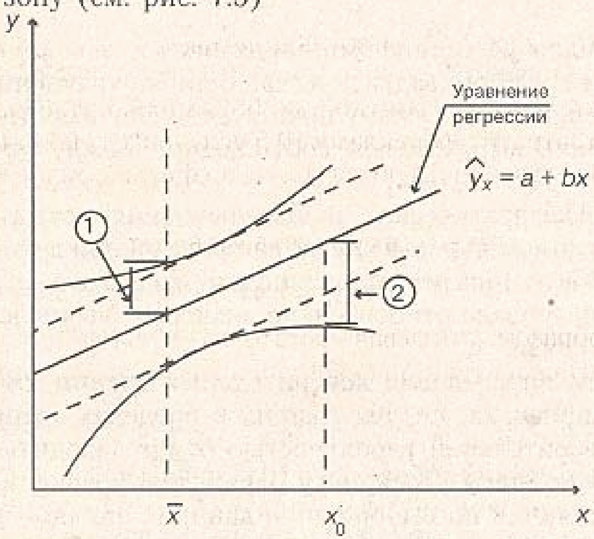

Рис. 7.3. **Доверительные интервалы линейного уравнения регрессии**

$$① - \frac{S_{\hat{y}}}{\sqrt{n}}; \quad ② - t_\alpha \cdot \frac{S_{\hat{y}}}{\sqrt{n}} C_{x_0}, \text{ где } C_{x_0} = \sqrt{1 + \frac{\left(x_0 - \bar{x}\right)^2}{\sigma_x^2}}$$

Покажем расчет доверительных границ числа туристов, которые воспользуются услугами фирмы, если затраты на рекламу будут составлять 10,5 усл. ден. ед.:

1) используя уравнение регрессии $\hat{y} = 118{,}3 + 83{,}84x$, находим число туристов, которые воспользуются услугами фирмы, если ее затраты на рекламу x_0 составят 10,5 усл. ден. единиц:

$$\hat{y}_{x_0} = 118{,}3 + 83{,}84 \cdot 10{,}5 = 998{,}6 \text{ человек.}$$

2) значение коэффициента t_α находим по таблице распределения Стьюдента для уровня значимости $\alpha = 0{,}05$ и числа степеней свободы, равного 18 (см. Приложение IV): $t_\alpha = 2{,}101$.

3) определяем множитель $C_{x_0} = \sqrt{1 + \frac{(10{,}5 - 9{,}95)^2}{1{,}2836^2}} = 1{,}1836.$

4) определяем величину $t_\alpha \frac{S_e}{\sqrt{n}} C_{x_0}$ (см. формулу 7.25).

Рассчитанная на с. 265 величина S_e составила 81,93 человек,

тогда $t_\alpha \frac{S_e}{\sqrt{n}} C_{x_0} = 2{,}101 \frac{81{,}93}{\sqrt{20}} \cdot 1{,}1836 = 45{,}56$ человек.

Следовательно, с вероятностью 95% можно утверждать, что если фирма затратит на рекламу 10,5 усл. ден. ед., количество туристов, которые воспользуются услугами фирмы, будет находиться в интервале:

$$998{,}6 - 45{,}6 \leq y_{x_0} \leq 998{,}6 + 45{,}6, \text{ т.е.}$$

$$953 \leq y_{x_0} \leq 1044.$$

Таким образом, учитывая «остаточную» вариацию результативного признака в данной совокупности фирм, можно полагать с вероятностью 95%, что число туристов, которые воспользуются услугами фирмы при затратах на рекламу 10,5 усл. ден. ед., будет не меньше 953 человек и не более 1044 человек.

Поскольку параметры уравнения регрессии определяются по выборочным данным, являясь функцией наблюденных значений, оценки параметров a и b содержат некоторую погрешность. Поэтому, как и во всех случаях оценки параметров генеральной совокупности по выборочным данным, возникает задача проверки гипотезы о величине коэффициента регрессии.

Средняя квадратическая (стандартная) ошибка коэффициента регрессии b определяется по формуле:

$$S_b = \frac{S_e}{\sqrt{\sum_{i=1}^{n} x_i^2 - n(\bar{x})^2}} \qquad (7.26)$$

или

$$S_b = \frac{S_e}{\sigma_x \sqrt{n}}, \qquad (7.26а)$$

где S_e - средняя квадратическая ошибка уравнения регрессии;
σ_x - среднее квадратическое отклонение x;
S_b - средняя квадратическая ошибка коэффициента регрессии b.

Средняя квадратическая ошибка параметра a уравнения регрессии определится по формуле:

$$S_a = \frac{S_e}{\sqrt{n}}, \qquad (7.27)$$

где S_a - средняя квадратическая ошибка параметра a.

Величину средней квадратической ошибки уравнения можно использовать и при выборе вида той или иной функции в качестве уравнения регрессии. Допустим, что в качестве уравнения регрессии, характеризующего зависимость числа туристов, воспользовавшихся услугами фирмы, от затрат фирмы на рекламу, мы избрали параболу второго порядка.

Тогда $\hat{y}_x = a_0 + a_1 x + a_2 x^2$.

Параметры уравнения a_0, a_1 и a_2 находим путем решения следующей системы нормальных уравнений:

$$\sum_{i=1}^{n} y_i = a_0 n + a_1 \sum_{i=1}^{n} x_i + a_2 \sum_{i=1}^{n} x_i^2,$$

$$\sum_{i=1}^{n} x_i y_i = a_0 \sum_{i=1}^{n} x_i + a_1 \sum_{i=1}^{n} x_i^2 + a_2 \sum_{i=1}^{n} x_i^3,$$

$$\sum_{i=1}^{n} y_i x_i^2 = a_0 \sum_{i=1}^{n} x_i^2 + a_1 \sum_{i=1}^{n} x_i^3 + a_2 \sum_{i=1}^{n} x_i^4.$$

Решив систему уравнений по данным приводимого примера, получим:

$a_0 = -382$; $a_1 = 186$; $a_2 = -5{,}08$.

Уравнение параболы второго порядка примет следующий вид:

$$\hat{y}_x = -382 + 186x - 5{,}08x^2.$$

Величина средней квадратической ошибки параболы S_e определится по формуле:

$$S_e = \sqrt{\frac{\sum_i^n y_i^2 - a_0 \sum_i^n y_i - a_1 \sum_i^n x_i y_i - a_2 \sum_i^n x_i^2 y_i}{n-3}}. \qquad (7.28)$$

В рассматриваемом примере $S_e = 83{,}15$ человек.

Сравнив значения средних квадратических ошибок уравнений прямой и параболы, можно отдать предпочтение уравнению прямой, поскольку величина S_e для этого уравнения регрессии является меньшей.

При практическом использовании уравнений регрессии следует помнить, что экстраполяция допускается только тогда, когда существенно не изменяются условия формирования уровней признаков, которые лежали в основе определения параметров уравнения регрессии. В противном случае использование уравнений для составления прогнозов должно быть отвергнуто. Необходим новый эмпирический материал, который отразит взаимосвязь между признаками в новых условиях с определенными качественными сдвигами.

7.5. Множественная корреляция

В приведенных примерах рассматривалась зависимость между двумя признаками, т.е. речь шла о так называемой парной корреляции. На практике же чаще всего изменение изучаемого признака зависит от действия нескольких причин. В таких случаях изучение корреляционной связи не может ограничиваться парными зависимостями, и в анализ необходимо включить другие признаки-факторы, существенно влияющие на изучаемую зависимую переменную. Одновременное изучение корреляции нескольких переменных проводится на основе использования методов множественной корреляции. Так, рассматривая уровень фондоотдачи на различных предприятиях одной отрасли, мы можем установить, что величина его зависит от размеров предприятия, удельного веса активной части фондов, степени изношенности фондов, их обновления и ряда других факторов.

Если обозначить факторы $x_1, x_2, x_3, \ldots, x_m$, то линейное уравнение множественной зависимости может быть записано так:

$$\hat{y}_{x_1, x_2, \ldots, x_m} = b_0 + b_1 x_1 + b_2 x_2 + b_3 x_3 + \ldots + b_m x_m.$$

Рассчитав параметры уравнения множественной зависимости, определяем значение индекса корреляции по следующей формуле:

$$i = \sqrt{1 - \frac{S^2_{\hat{y}\,x_j}}{\sigma^2_y}}, \qquad (7.23\text{в})$$

где $S^2_{\hat{y}\,x_j}$ – дисперсия эмпирических значений относительно значений, рассчитанных по уравнению регрессии, которая определяется делением остаточной суммы квадратов отклонений результативного признака на $(n-m-1)$;

σ_y^2 – дисперсия эмпирических значений результативного признака.

По параметрам полученного уравнения можем оценить долю каждого из факторов в изменении уровня результативного показателя y. Это может быть сделано путем прямой оценки по величине коэффициентов регрессии при каждом из факторов, а также по коэффициентам эластичности $Э_{x_j}$, стандартизированным частным коэффициентам регрессии β-коэффициентам и Δ-коэффициентам.

Коэффициенты уравнения множественной регрессии показывают абсолютный размер влияния факторов на уровень результативного показателя и характеризуют степень влияния каждого фактора на анализируемый показатель при фиксированном (среднем) уровне других факторов, входящих в модель.

Для сравнения оценок роли различных факторов в формировании моделируемого показателя следует дополнить абсолютные величины относительными. Так, частные коэффициенты эластичности показывают, на сколько процентов в среднем изменяется y с изменением признака-фактора x_j на один процент при фиксированном положении других факторов, и рассчитываются по формуле:

$$Э_j = b_j \frac{\bar{x}_j}{\bar{y}}, \qquad (7.29)$$

где b_j - коэффициент регрессии при j-м факторе.

β-коэффициенты показывают, на какую часть среднего квадратического отклонения σ_y изменится зависимая переменная y с изменением соответствующего фактора x_j на величину своего среднеквадратического отклонения (σ_j). Этот коэффициент позволяет сравнивать влияние колеблемости различных факторов на вариацию исследуемого показателя, на основе чего выявляются факторы, в развитии которых заложены наибольшие резервы изменения результативного показателя:

$$\beta_j = b_j \frac{\sigma_j}{\sigma_y}. \qquad (7.30)$$

Коэффициенты эластичности и β-коэффициенты взаимосвязаны следующим образом:

$$\beta_j = Э_j \cdot \frac{v_j}{v_y}, \qquad (7.31)$$

где v_j - коэффициент вариации j-того факторного признака;

v_y - коэффициент вариации результативного признака.

Чтобы оценить долю влияния каждого фактора в суммарном влиянии факторов, включенных в уравнение регрессии, рассчитывают Δ-коэффициенты:

$$\Delta_j = \frac{r_{jy} \cdot \beta_j}{\sum_j r_{jy} \cdot \beta_j} = \frac{r_{jy} \cdot \beta_j}{R^2}. \qquad (7.32)$$

Содержательный анализ моделей в целях уточнения приоритетности факторов опирается на сравнение перечисленных коэффициентов. В этих целях, особенно при достаточно большом числе факторов, включаемых в уравнение регрессии, производится ранжирование факторов по величине коэффициентов эластичности, бета и дельта - коэффициентов.

Рассмотрим принципы анализа степени влияния факторов на следующем примере, когда в уравнение регрессии были включены пять факторов.

Таблица 7.12

Факторы	Значения коэффициентов			Ранг факторов по величине коэффициентов			Средний ранг
	$Э_j$	β_j	Δ_j	$Э_j$	β_j	Δ_j	
1	2	3	4	5	6	7	8
x_1	0,173	0,204	0,162	2	3	2	2
x_2	0,133	0,114	0,076	4	5	5	5
x_3	0,108	0,144	0,104	5	4	4	4
x_4	0,158	0,253	0,133	3	2	3	3
x_5	1,795	0,732	0,525	1	1	1	1

Если сопоставить значения коэффициентов эластичности (гр. 2 и 5 табл. 7.12), то можно видеть, что главным фактором изменения результативного показателя является фактор x_5: при его изменении на 1% y возрастает на 1,795%. Вторым по силе влияния на результативный показатель является фактор x_1 и т.д. (см. гр. 5 табл. 7.12).

Сравнение значений β_j позволяет сделать вывод, что с учетом уровня колеблемости факторов наибольшие резервы в изменении результативного показателя заложены в увеличении факторов x_5, x_4 и x_1 (см. гр. 6 табл. 7.12).

Сопоставление значений коэффициентов Δ_j позволяет сделать вывод, что наибольшая доля влияния падает на фактор x_5: роль этого фактора в вариации результативного показателя составляет 52,5% общего влияния пяти факторов на результативный показатель. Доля влияния других факторов значительно уступает фактору x_5. Следовательно, наибольшие возможности в изменении результативного показателя связаны с изменением фактора x_5, затем фактора x_1 и далее x_4.

При построении многофакторных корреляционных моделей одной из предпосылок обоснованности конечных результатов является требование возможно меньшей коррелированности включенных в модель признаков-факторов (отсутствие мультиколлинеарности). В случае линейной зависимости между факторами система нормальных уравнений не будет иметь однозначного решения, в результате чего коэффициенты регрессии и другие оценки окажутся неустойчивыми. Кроме того, наличие взаимосвязи факторов затрудняет экономическую интерпретацию урав-

нения связи, так как изменение одного из факторов влечет, как правило, изменение факторов, с ним связанных. Для исключения мультиколлинеарности было предложено несколько методов. На практике чаще всего используют чисто эмпирический подход к решению этой проблемы [1].

Рассмотрим матрицу парных коэффициентов корреляции (индекс «0» присвоен результативному признаку y), по которой можно оценить степень взаимной коррелированности признаков-факторов.

Таблица 7.12а

Факторы	y	x_1	x_2	...	x_j	...	x_m
y	1	r_{10}	r_{20}	...	r_{j0}	...	r_{m0}
x_1	r_{01}	1	r_{21}	...	r_{j1}	...	r_{m1}
x_2	r_{02}	r_{12}	1	...	r_{j2}	...	r_{m2}
...	...	...	...	...	...	...	...
x_j	r_{0j}	r_{1j}	r_{2j}	...	1	...	r_{mj}
...	...	...	...	...	...	...	...
x_m	r_{0m}	r_{1m}	r_{2m}	...	r_{jm}	...	1

В качестве критерия мультиколлинеарности может быть принято соблюдение следующих неравенств:

$$r_{x_j y} > r_{x_j x_k}$$

$$r_{x_k y} > r_{x_j x_k}$$

Если приведенные неравенства (или хотя бы одно из них) не выполняются, то исключается тот фактор x_j или x_k, связь которого с результативным показателем y будет менее тесной. Вместе с тем окончательный вывод о наличии или отсутствии мультиколлинеарности должен быть сделан в соответствии с экономическим содержанием и логикой взаимосвязи конкретных факторов.

Величина **совокупного коэффициента корреляции** по значениям парных коэффициентов может быть определена следующим образом:

[1] Более эффективным средством решения сформулированной выше задачи исключения коллинеарных факторов является использование факторного анализа.

$$R^2 = 1 - \frac{\begin{vmatrix} 1 & r_{10} & r_{20} & \cdot & r_{m0} \\ r_{10} & 1 & r_{21} & \cdot & r_{m1} \\ r_{20} & r_{12} & 1 & \cdot & r_{m2} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ r_{0m} & r_{1m} & r_{2m} & \cdot & 1 \end{vmatrix}}{\begin{vmatrix} 1 & r_{21} & \cdot & r_{m1} \\ r_{12} & 1 & \cdot & r_{m2} \\ \cdot & \cdot & \cdot & \cdot \\ r_{1m} & r_{2m} & \cdot & 1 \end{vmatrix}} \tag{7.33}$$

Величина R^2, называемая еще **коэффициентом детерминации**, показывает, в какой мере вариация результативного признака обусловлена влиянием признаков-факторов, включенных в рассматриваемое уравнение корреляционной зависимости.

Величина совокупного коэффициента корреляции изменяется в пределах от 0 до 1 и численно не может быть меньше, чем любой из образующих его парных коэффициентов корреляции. Чем ближе совокупный коэффициент корреляции к единице, тем меньше роль неучтенных в модели факторов и тем больше оснований считать, что параметры регрессионной модели отражают степень эффективности включенных в нее факторов.

Для случая зависимости результативного признака от двух факторных признаков формула совокупного коэффициента корреляции имеет вид:

$$R_{y_{x_1 x_2}} = \sqrt{\frac{r_{10}^2 + r_{20}^2 - 2 \cdot r_{10} r_{20} r_{12}}{1 - r_{12}^2}} \tag{7.33а}$$

Расчет параметров уравнения зависимости для такого варианта можно вести по формулам, в которых используются уже вычисленные значения парных коэффициентов корреляции и средних квадратических отклонений.

Так, для уравнения вида $\hat{y}_{x_1 x_2} = b_0 + b_1 x_1 + b_2 x_2$ параметры b_0, b_1, b_2 могут быть найдены по следующим формулам:

$$b_1 = \frac{\sigma_y}{\sigma_1} \cdot \frac{r_{10} - r_{20} \cdot r_{12}}{1 - r_{12}^2},$$

$$b_2 = \frac{\sigma_y}{\sigma_2} \cdot \frac{r_{20} - r_{10} \cdot r_{12}}{1 - r_{12}^2}, \tag{7.34}$$

$$b_0 = \overline{y} - b_1 \overline{x}_1 - b_2 \overline{x}_2,$$

где σ_1, σ_2, σ_y - средние квадратические отклонения соответственно признаков x_1, x_2 и y;

r_{10}, r_{20}, r_{12} - парные коэффициенты корреляции.

Для более глубокого исследования связей между явлениями целесообразно установить степень тесноты связи между результативным признаком y и каждым из факторных признаков при исключении влияния других факторных признаков.

Для решения поставленной задачи определяют так называемые **коэффициенты частной корреляции**, выявляющие степень «чистого» влияния факторного признака на результативный признак. Для расчета частных коэффициентов корреляции могут быть использованы парные коэффициенты корреляции.

Для случая зависимости y от двух признаков можно будет вычислить два коэффициента частной корреляции:

1) частный коэффициент корреляции $r_{01.2}$ между результативным признаком y и фактором x_1 при элиминировании фактора x_2 показывает, какую часть колеблемость y, вызванная фактором x_1, составляет в колеблемости y под действием всех факторов, кроме фактора x_2.

$$r_{01.2} = \frac{r_{10} - r_{20} \cdot r_{12}}{\sqrt{\left(1 - r_{12}^2\right)\left(1 - r_{20}^2\right)}}; \tag{7.35а}$$

2) второй частный коэффициент корреляции $r_{02.1}$ характеризует зависимость результативного признака от фактора x_2 при исключении влияния фактора x_1.

$$r_{02.1} = \frac{r_{20} - r_{10} \cdot r_{12}}{\sqrt{\left(1 - r_{10}^2\right)\left(1 - r_{12}^2\right)}} \tag{7.35б}$$

Для общего случая частные коэффициенты корреляции можно определить таким образом:

$$r_{0m.1,2,\ldots,m-1} = \sqrt{\frac{R_m^2 - R_{m-1}^2}{1 - R_m^2}}, \tag{7.35}$$

где R_m^2 - коэффициент детерминации результативного признака y с комплексом факторных признаков $x_1, x_2, \ldots, x_{m-1}, x_m$;

R_{m-1}^2 - коэффициент детерминации результативного признака y с комплексом признаков $x_1, x_2, \ldots, x_{m-1}$;

$r_{0m.1,2,\ldots,m-1}$ - частный коэффициент корреляции y с факторным признаком x_m при исключении влияния факторных признаков $x_1, x_2, \ldots, x_{m-1}$.

При малых значениях $r_{0m.1,2,\ldots,m-1}$ нет смысла вводить в уравнение m-й фактор, так как эффективность нового уравнения регрессии, характеризующего зависимость от m-факторов, возрастает незначительно.

Построение многофакторных регрессионных моделей позволяет дать количественное описание основных закономерностей изучаемых явлений, выделить существенные факторы, обусловливающие изменение экономических показателей, и оценить их влияние.

Полученные модели в основном используются в двух направлениях - для сравнительного анализа и в прогнозировании. Например, для выявления внутриотраслевых резервов повышения эффективности производства рассчитывается уравнение множественной зависимости, рассматриваемое в качестве экономико-статистической модели анализируемого показателя эффективности и характеризующее основные закономерности в формировании этого показателя для совокупности предприятий отрасли. На основе такого уравнения можно проанализировать и сравнить влияние каждого фактора на повышение эффективности в среднем по отрасли.

Разбив все предприятия на группы, различающиеся по условиям производства, определяемых организационно-техническим уровнем развития, можно с помощью модели соответствующего показателя эффективности сравнить результаты деятельности этих групп предприятий между собой и со средним отраслевым уровнем.

Процедура сравнительного анализа с использованием уравнений регрессии сводится к следующему. Для каждой группы предприятий, сформированных по организационно-техническому уровню развития, находятся не только средние значения показателей, характеризующих эффективность деятельности, но и показателей, определяющих уровень результативных (под ними понимаются признаки-факторы, включенные в соответствующую регрессионную модель). Затем проводится сравнение средних значений факторных признаков сопоставляемых групп предприятий и определяются величины $\Delta\bar{x}_{ij}$ где j – номер признака-фактора;

i-номер группы. Умножая разности $\Delta\bar{x}_{ij}$, найденные по сравниваемым группам, на величину коэффициентов регрессии b_j в многофакторной регрессионной модели, получаем эффект влияния различий в уровнях факторов по группам (см. табл. 7.13). Величина полученного эффекта $(\Delta\bar{x}_{ij} \cdot b_j)$ позволяет судить о вкладе каждого фактора в изменение результативного показателя в данной группе по сравнению с другой группой. Рассмотрим в табл. 7.13 процедуру сравнительного анализа, когда вся совокупность предприятий (единиц наблюдения) разбита на две группы.

Таблица 7.13

Общая схема сравнительного анализа с применением уравнений регрессии

Факторы	Средние значения факторов по группе: I группа	Средние значения факторов по группе: II группа	Разность средних значений факторов (гр.2–гр.3)	Коэффициенты регрессии b_j	Эффект влияния на результативный показатель различий в средних уровнях факторов (гр.4·гр.5)
1	2	3	4	5	6
x_1	$\bar{x}_{11}$	$\bar{x}_{12}$	$\Delta\bar{x}_1$	b_1	$b_1\cdot\Delta\bar{x}_1$
x_2	$\bar{x}_{21}$	$\bar{x}_{22}$	$\Delta\bar{x}_2$	b_2	$b_2\cdot\Delta\bar{x}_2$
x_3	$\bar{x}_{31}$	$\bar{x}_{32}$	$\Delta\bar{x}_3$	b_3	$b_3\cdot\Delta\bar{x}_3$
.	.	.	.	.	.
x_j	$\bar{x}_{j1}$	$\bar{x}_{j2}$	$\Delta\bar{x}_j$	b_j	$b_j\cdot\Delta\bar{x}_j$
.	.	.	.	.	.
x_m	$\bar{x}_{m1}$	$\bar{x}_{m2}$	$\Delta\bar{x}_m$	b_m	$b_m\cdot\Delta\bar{x}_m$
y	$\bar{y}_1$	$\bar{y}_2$	$\Delta\bar{y}$		$\sum_{j=1}^{m} b_j\,\Delta\bar{x}_j$

Общая величина различий значений результативного показателя $\Delta\bar{y} = \bar{y}_1 - \bar{y}_2$ в определенной части обусловлена изменением признаков-факторов, включенных в уравнение регрессии. Размер этого влияния определяется величиной $\sum_{j=1}^{m} b_j\,\Delta\bar{x}_j$. Разность $(\Delta\bar{y} - \sum_{j=1}^{m} b_j\,\Delta\bar{x}_j)$ может быть объяснена влиянием прочих неучтенных факторов. Чем выше доля влияния включенных в уравнение регрессии факторов, тем большую значимость приобретает использование уравнения регрессии в управлении деятельностью исследуемой совокупности предприятий. По результатам такого сравнительного анализа можно дать объективную оценку возможностей повышения эффективности производства и выявить внутрихозяйственные резервы в целом по отрасли.

Применение регрессионных моделей в прогнозировании требует осторожности в тех случаях, когда мы сталкиваемся с выходом фактических значений факторных признаков за границы значений признаков, на основе которых было рассчитано уравнение регрессии. Необходима проверка возможности экстраполяции, связанная с оценкой неизменности общих условий формирования уровней признаков в совокупности.

Возможность широкого применения методов корреляционно-регресионного анализа еще в недалеком прошлом сдерживалась высокой трудоемкостью необходимых расчетных процедур. Сегодня широкое распространение получили **пакеты прикладных программ** (ППП) по статистике [1], ликвидировавшие эти ограничения. Однако роль исследователя остается чрезвычайно важной как на этапе предварительной подготовки массива исходной информации, так и на этапе содержательной интерпретации полученных уравнений регрессии и их практическом применении. Прежде всего исследователь обосновывает наличие причинной зависимости между признаками-факторами и результативным признаком. При предварительном анализе массива фактических данных важно оценить однородность совокупности исследуемых единиц, проверить возможность наличия выделяющихся наблюдений, обосновать принципы группировки единиц по значениям факторного признака. По результатам такого анализа формируется информационный массив для расчета показателей степени тесности связи, а затем и параметров уравнения регрессии.

Целесообразны предварительная проверка гипотезы о линейности связи между факторным и результативным признаком до выбора математической формы уравнения регрессии и анализ матрицы парных коэффициентов корреляции. Все сказанное свидетельствует о необходимости выделения промежуточных этапов в последовательности обработки данных с помощью пакетов прикладных программ для внесения необходимых корректировок в процессе обработки данных.

[1] См. с. 65.

Контрольные вопросы к главе 7

1. В чем состоит отличие между корреляционной и функциональной связью?
2. Какие основные проблемы решает исследователь при изучении корреляционных зависимостей?
3. Какова роль групповых и корреляционных таблиц при анализе взаимосвязей?
4. Какие методы целесообразно использовать для выявления возможного наличия связи между факторным и результативным признаком при небольшом объеме фактических данных?
5. Какие показатели являются мерой тесноты связи между двумя признаками?
6. Как оценить существенность линейного коэффициента корреляции?
7. Какие показатели используют для измерения степени тесноты связи между качественными признаками?
8. В чем состоит значение уравнения регрессии?
9. Что характеризует коэффициент регрессии?
10. Какая существует связь между линейным коэффициентом корреляции и коэффициентом регрессии?
11. Какое значение имеет расчет индекса корреляции?
12. Способы расчета средней квадратической ошибки уравнения, ее роль в оценке надежности уравнения регрессии.
13. Как осуществить прогноз значений результативного признака, опираясь на использование для этой цели уравнения регрессии?
14. Как измерить долю общей вариации результативного признака, которая объясняется влиянием вариации признака-фактора x?
15. Если в случае линейной зависимости между признаками 60% вариации результативного признака объясняется влиянием факторного признака, чему будет равна величина коэффициента корреляции?
16. Как подходить к отбору факторов для включения их в уравнение множественной регрессии?

17. Каким образом можно выделить факторы, в изменении которых заложены наибольшие возможности в управлении изменением результативного признака?

18. Что означает величина коэффициента эластичности 0,58?

19. Для чего рассчитываются частные коэффициенты корреляции?

20. В каких пределах заключена величина совокупного коэффициента корреляции и как она соотносится с величиной парных коэффициентов корреляции?

Глава 8
РЯДЫ ДИНАМИКИ

8.1. Виды рядов динамики

Важной задачей статистики является изучение изменений анализируемых показателей во времени. Эти изменения можно изучать, если иметь данные по определенному кругу показателей на ряд моментов времени или за ряд промежутков времени, следующих друг за другом.

Ряд расположенных в хронологической последовательности значений статистических показателей, представляет собой временной (динамический) ряд. Каждый временной ряд состоит из двух элементов: во-первых, указываются моменты или периоды времени, к которым относятся приводимые статистические данные; во-вторых, приводятся те статистические показатели, которые характеризуют изучаемый объект на определенный момент или за указанный период времени.

Статистические показатели, характеризующие изучаемый объект, называют **уровнями ряда**. Вид ряда динамики зависит не только от характера показателей, оценивающих изучаемый объект, но и от того, дается ли показатель за какой-либо период или по состоянию на определенный момент времени. Статистические показатели, приводимые в динамическом ряду, могут быть абсолютными, относительными или средними величинами. В качестве примера ряда динамики абсолютных величин рассмотрим показатели ввода в действие жилых домов в России.

Каждый уровень приведенного динамического ряда характеризует ввод в действие жилых домов за определенный промежуток времени - год. Такие ряды динамики называются *интервальными*.

Таблица 8.1

Ввод в действие жилых домов (млн. кв. м)

Годы	Всего построено	В том числе		
		государственными предприятиями и организациями	жилищно-строительными кооперативами	населением за свой счет и с помощью кредита
1	2	3	4	5
1990	61,7	48,4	2,9	6,0
1991	49,4	37,6	2,4	5,4
1992	41,5	29,2	2,1	4,9
1993	41,8	15,0	1,9	5,6
1994	38,5	10,7	1,8	7,0

Источник. Россия в цифрах. 1995. Краткий статистический сборник. Госкомстат России. М., 1995. С. 258.

Приведем данные о численности населения России (табл. 8.2) Уровни этого динамического ряда характеризуют численность населения на определенную дату. Такой ряд динамики называется *моментным динамическим рядом.*

Таблица 8.2

Численность населения России (на начало года)

	1990 г.	1991 г.	1992 г.	1993 г.	1994 г.
Численность населения, млн человек	147,7	148,2	148,3	148,3	148,0

Отличительной особенностью интервальных рядов динамики абсолютных величин является возможность суммирования их уровней. Так же как и каждый уровень приведенного интервального ряда получился от суммирования уровней за более короткие промежутки времени (месяцы, кварталы), так и сумма уровней за годы характеризует ввод жилых домов за больший промежуток времени. В результате суммирования уровней интервального динамического ряда получаем так называемые *накопленные итоги,* которые имеют реальное содержание.

Сумма же уровней моментного ряда динамики не имеет никакого реального содержания, и «накопленные итоги» для этих рядов не рассчитываются. Определенный смысл имеет расчет разностей уровней моментного динамического ряда, который будет

характеризовать изменение уровня за определенный период времени. Например, за 1990 г. численность населения России выросла на 0,5 млн человек, за 1991 г. - на 0,1 млн, а за 1993 г. численность населения уменьшилась на 0,3 млн человек.

Ряды динамики относительных и средних величин состоят из производных статистических показателей, полученных в результате сопоставления между собой суммарных абсолютных данных.

Рядом динамики относительных величин называется ряд цифровых данных, характеризующих изменение относительных размеров изучаемых явлений во времени. Темпы роста потребительских цен в России за 8 месяцев 1995 г. характеризуются такими данными.

Таблица 8.3

Индексы потребительских цен (за 8 месяцев 1995 г.)

Месяцы	Сводный индекс потребительских цен на товары и услуги	
	в % к предыдущему месяцу	в % к декабрю предыдущего года
январь	118,0	118,0
февраль	111,0	131,0
март	108,9	142,0
апрель	108,5	154,0
май	107,9	167,0
июнь	106,7	178,0
июль	105,4	187,0
август	104,6	196,0

Источник. Статистическое обозрение. 1995. № 9. С. 44.

И, наконец, в табл. 8.4 приведен ряд динамики, характеризующий изменение **среднего размера признака**.

Важнейшей проблемой построения динамических рядов является проблема сопоставимости уровней этих рядов, относящихся к различным периодам. Соблюдение требования сопоставимости уровней ряда означает, что полезно лишь такое сравнение, которое учитывает существо изучаемого явления и цель, к которой оно приводит. Показатели динамического ряда, подлежащие сопоставлению, должны быть однородны по экономическому содержанию. Однако вследствие многих обстоятельств однородность величин, составляющих динамический ряд, может нарушиться, и таким образом нарушается сопоставимость уровней динамического ряда.

Таблица 8.4

Среднемесячная заработная плата за 1995 г. в России

Месяцы	Среднемесячная заработная плата	
	тыс. руб.	долл. США (по курсу ММВБ)
январь	302,6	78,4
февраль	321,0	75,4
март	361,5	76,1
апрель	386,2	76,6
май	429,9	85,1
июнь	480,6	101,4
июль	499,5	110,6

Источник. Статистическое обозрение. 1995. № 9. С. 60.

Прежде всего должна быть обеспечена одинаковая **полнота охвата** различных частей явления. Требование одинаковой полноты охвата разных частей изучаемого объекта означает, что уровни динамического ряда за отдельные периоды времени должны характеризовать размер того или иного явления по одному и тому же кругу, входящих в его состав частей. Например, при характеристике динамики численности студентов высших учебных заведений по годам нельзя в одни годы учитывать только численность студентов дневного обучения, а в другие - численность студентов всех видов обучения.

Вопрос о том, является ли непременным условием сопоставимости уровней динамического ряда **одинаковость границ территории**, может быть решен различно. Так, при характеристике экономической мощи страны надо использовать данные в изменяющихся границах территории, а при изучении темпов экономического развития целесообразно использовать данные по территории в одних и тех же границах.

При определении сравниваемых уровней динамического ряда необходимо использовать единую методологию их расчета. Особенно часто эта проблема возникает при международных сопоставлениях.

Несопоставимость показателей, возникающая в силу неодинаковости применяемых единиц измерения, сама по себе очевидна. С различием применяемых единиц измерения приходится

встречаться при учете продукции в натуральном выражении. Поэтому приведение к сопоставимому виду разнообразной продукции основывается на ее выражении в ценностных или трудовых измерителях. При анализе показателей объема продукции, измеренной в **ценностных единицах**, следует учитывать, что во-первых, с течением времени происходит непрерывное изменение цен, а, во-вторых, существует несколько видов цен (цены производителей и цены потребителей). В этой связи при характеристике стоимостных показателей объема продукции во времени должно быть устранено влияние изменения цен. На практике для решения этой задачи количество продукции, произведенное в разные периоды, оценивают *в ценах одного периода*, которые называют фиксированными или сопоставимыми (подробнее см.гл.9).

Одним из условий сопоставимости уровней интервального динамического ряда является **равенство периодов**, за которые приводятся данные. Например, для характеристики степени ритмичности работы предприятия данные об удельном весе продукции по определенным декадам использовать нецелесообразно, так как число рабочих дней отдельных декад может оказаться различным, что приводит к различиям в объеме выпуска продукции.

Все выше названные обстоятельства следует учитывать при подготовке информации для анализа изменений явлений во времени.

Вернемся к данным о динамике средней месячной заработной платы, приведенным в табл. 8.4. По этим данным можно сделать вывод о тенденции роста показателя в течение всего рассматриваемого периода. Учитывая, что показатели средней заработной платы входят в систему показателей уровня жизни населения, можно ли сделать вывод о его повышении в 1995 г.? Вот здесь-то и возникает проблема сопоставимости доходов в денежном выражении, поскольку в условиях высокой инфляции (индексы потребительских цен приведены в табл. 8.3) покупательная способность рубля будет снижаться. Поэтому наряду с номинальными денежными доходами при изучении уровня жизни следует исчислять реальные денежные доходы населения, на основе сопоставления которых можно анализировать тенденции в динамике уровня жизни.

Наглядно представить процесс развития явлений во времени позволяет графическое изображение изменения уровней временного ряда. Способы графического представления динамики весьма разнообразны, однако их можно объединить в две большие группы: **диаграммы** и **картограммы**.

Наибольшее распространение имеют линейные, столбиковые, ленточные, секторные и фигурные диаграммы.

Линейные диаграммы в прямоугольной системе координат - наиболее распространенный вид графических изображений динамических рядов. Обычно при построении линейной диаграммы на оси абсцисс располагают отрезки, представляющие собой даты или периоды времени, а на оси ординат - уровни ряда динамики. Каждая точка линейной диаграммы соответствует уровню динамического ряда, относящемуся к определенному моменту или периоду времени. Часто на одном и том же графике приводятся несколько кривых, которые позволяют дать сравнительную характеристику динамики различных показателей или одного и того же показателя в разных регионах.

При графическом изображении сильно отличающихся по величине уровней применяется полулогарифмическая сетка, в которой по оси абсцисс откладываются в натуральном масштабе периоды времени, а по оси ординат - логарифмы уровней динамического ряда. Существенно усиливает аналитическое содержание графика применение двух и трех шкал. Так, например, на основной шкале можно привести данные о производстве электроэнергии в натуральном выражении, а на сопряженной шкале - относительное изменение в процентах к базисному году.

Кроме линейной диаграммы для графического изображения развития явления во времени применяются *радиальные диаграммы*. Они удобны в том случае, когда динамика явлений носит периодический характер (например, при наличии сезонности). Пример радиальной диаграммы дан на рис. 8.1. В радиальных диаграммах на внешнем круге обозначается время, а на диаметрах отсчитываются уровни, начиная от центра круга. Получаемые концентрические окружности изображают весь ряд динамики с повторением колебаний по годам.

Для наглядного сравнения объемов изучаемого явления во времени или пространстве используют *столбиковые диаграммы*. Они широко применяются для популяризации развития во времени явлений за короткие промежутки времени. Этот вид диаграмм может быть применен и для изображения динамики структуры явлений. Столбики располагаются вплотную или раздельно. Они должны иметь одинаковое основание, а их высота - пропорциональна числовым значениям изображаемых показателей, так как о соотношении между уровнями судят по высоте столбиков. Возможно и горизонтальное расположение прямоугольников. В этом случае диаграмма называется *ленточной*. Размеры показателей представлены в виде ленты (полосы), ширина которой одинакова, а длина представляет величину этих показателей в соответствии с выбранным масштабом.

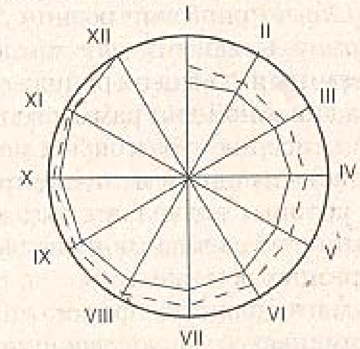

Рис. 8.1. Производство электроэнергии ТЭЦ по месяцам года:

——— - 1993 г.

— — - 1994 г. (1 см - 20 млн кВт-ч).

В ряде случаев применяются *фигурные диаграммы*, в которых величина изображаемого показателя пропорциональна площади фигуры, изображающей данное явление. Например, представление о производстве автомашин (или другой продукции) в разных странах или в одной стране за ряд лет можно представить с помощью изображения автомашин разного размера, площадь которых пропорциональна изображаемым числовым показателям. Для таких диаграмм необходимо давать сопроводительные числовые подписи, так как зрительное сопоставление площадей таких фигур довольно затруднительно.

В публикациях ООН достаточно часто для графических изображений экономических показателей применяются картограммы и картодиаграммы. (подробнее см.гл. 3).

8.2. Показатели ряда динамики и методы их исчисления

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 основных закономерностей динамики исследуемого явления на отдельных этапах и в целом за рассматриваемый период; 4) выявление факторов, обусловливающих изменение изучаемого объекта во времени; 5) прогноз развития явления на будущее.

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 сравнению 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 (см. рис. 8.2).

Абсолютный прирост (Δ_i) определяется как разность между двумя уровнями динамического ряда и показывает на сколько данный уровень ряда превышает уровень, принятый за базу сравнения:

$$\Delta_i = y_i - y_0, \qquad (8.1)$$

где Δ_i - абсолютный прирост,
y_i - уровень сравниваемого периода,
y_0 - уровень базисного периода.

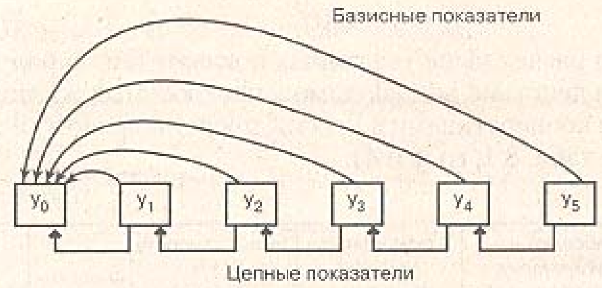

Рис. 8.2. **Принципы построения цепных и базисных показателей динамики**

При сравнении с переменной базой абсолютный прирост будет равен:

$$\Delta_i = y_i - y_{i-1}, \qquad (8.1а)$$

где y_{i-1} - уровень непосредственно предшествующего периода.

Абсолютный прирост с переменной базой иначе называют *скоростью роста.*

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

При сравнении с постоянной базой $K_i = \dfrac{y_i}{y_0}$. (8.2)

При сравнении с переменной базой $K_i = \dfrac{y_i}{y_{i-1}}$. (8.2а)

Если коэффициенты роста выражают в процентах, то их называют *темпами роста:* $T_p = K \cdot 100\%$. (8.3)

Темп прироста показывает на сколько процентов уровень данного периода больше (или меньше) базисного уровня. Этот показатель может быть рассчитан двояко:

1) как отношение абсолютного прироста к уровню, принятому за базу сравнения

$$T_п = \frac{y_i - y_0}{y_0} \cdot 100\% \qquad (8.3а)$$

или

$$T_п = \frac{y_i - y_{i-1}}{y_{i-1}}, \qquad (8.3б)$$

2) как разность между темпом роста (в процентах) и 100%

$$T_п = T_р - 100\% . \qquad (8.3в)$$

Рассмотрим расчет вышеуказанных показателей по ряду динамики ввода в действие жилых домов, построенных жилищно-строительными кооперативами в России, представленному в табл. 8.5 (см. данные табл. 8.1; гр. 1 и 4).

Таблица 8.5

Годы	Всего построено (млн кв. м) жилищно-строительными кооперативами	Абсолютное изменение млн кв м		Коэффициенты роста по сравнению		Темпы прироста (%) по сравнению		Абсолютное значение 1% прироста, тыс. кв. м	Пункты роста, %
		по сравнению с уровнем 1990 г.	по сравнению с предшеств. годом	с 1990 г.	с предшествующим годом	с 1990 г.	с предшествующим годом		
1	2	3	4	5	6	7	8	9	10
1990	2,9	-	-	-	-	-	-	-	
1991	2,4	-0,5	-0,5	0,8276	0,8276	-17,24	-17,24	290	-17,24
1992	2,1	-0,8	-0,3	0,7241	0,8750	-27,59	-12,50	240	-10,35
1993	1,9	-1,0	-0,2	0,6552	0,9048	-34,48	-9,52	210	-6,89
1994	1,8	-1,1	-0,1	0,6207	0,9474	-37,93	-5,26	190	-3,45

Между показателями динамики, вычисленными с постоянной и переменной базой, существует определенная связь. Так, сумма абсолютных приростов с переменной базой дает общий прирост за исследуемый период.

$$\sum_{i=1}^{n-1} \Delta_i = y_n - y_1 = \Delta_{n/1},$$

где n - число уровней динамического ряда [1] (-0,5 - 0,3 - 0,2 - 0,1 = = -1,1) (см. гр. 3 и 4 табл. 8.5).

Равным образом может быть осуществлен переход от коэффициентов роста с постоянной базой к коэффициентам роста, вычисленным с переменной базой и наоборот. Если известны коэффициенты роста с постоянной базой, то частные от последовательного деления этих коэффициентов равны соответствующим коэффициентам роста с переменной базой.

$$K_{i/i-1} = \frac{K_{i/1}}{K_{i-1/1}}.$$

[1] В подстрочных обозначениях стоят индексы уровней: в числителе - сравниваемого, а в знаменателе - базы сравнения.

При наличии коэффициентов роста с переменной базой соответствующий базисный коэффициент роста находится перемножением цепных коэффициентов роста.

$$K_{i/1} = K_{2/1} \cdot K_{3/2} \cdot \ldots \cdot K_{i-1/i-2} \cdot K_{i/i-1}$$

(0,8276 • 0,8750 • 0,9048 • 0,9474 = 0,6207) (см. гр. 5 и 6 табл. 8.5).

При анализе относительных показателей динамики (темпов роста и темпов прироста) не следует рассматривать их изолированно от абсолютных показателей (уровней ряда и абсолютных приростов). Сравнение абсолютного прироста и темпа прироста за одни и те же периоды времени показывает, что замедление темпов прироста не всегда сопровождается уменьшением абсолютных приростов. Поэтому, чтобы правильно оценить значение полученного темпа прироста, его рассматривают в сопоставлении с показателем абсолютного прироста. Результат выражают показателем, который называют *абсолютным значением одного процента прироста* A_i.

$$A_i = \frac{y_i - y_{i-1}}{T_{п\, i/i-1}}, \qquad (8.4)$$

т.е. этот показатель рассчитывают как отношение абсолютного прироста к темпу прироста (в%) за тот же период времени.

Если преобразовать эту формулу, то получим следующее выражение:

$$A_i = \frac{y_i - y_{i-1}}{\dfrac{y_i - y_{i-1}}{y_{i-1}} \cdot 100\%} = \frac{y_{i-1}}{100} = 0{,}01\, y_{i-1} \qquad (8.4а)$$

Абсолютные значения одного процента прироста представлены в гр. 9 табл. 8.5.

В тех случаях, когда сравнение производится с достаточно отдаленным периодом времени, принятым за базу сравнения, рассчитывают так называемые пункты роста, которые представляют собой разность базисных темпов роста (прироста), в% двух смежных периодов (см. гр. 10 табл. 8.5). Так, для 1994 г. пункт роста составит - 3,45 = - 37,93 - (-34,48) (см. гр. 7 табл. 8.5).

В отличие от темпов прироста, которые нельзя ни суммировать, ни перемножать, пункты роста можно складывать, в результате чего получаем темп прироста соответствующего периода по сравнению с базисным периодом, т.е. по данным нашего примера сумма пунктов роста равна - 37,93, что соответствует темпу прироста уровня 1994 г. по сравнению с 1990 г. (-17,24 - 10,35 - 6,89 - 3,45 = - 37,93) (см. гр. 7 табл. 8.5).

При сопоставлении динамики развития двух явлений можно использовать показатели, представляющие собой отношения темпов роста или темпов прироста за одинаковые отрезки времени по двум динамическим рядам. Эти показатели называют **коэффициентами опережения**.

$$K_{0п} = \frac{T_р'}{T_р''} \quad \text{или} \quad K_{0п} = \frac{T_п'}{T_п''}, \tag{8.5}$$

где $T_р'$, $T_р''$ и $T_п'$, $T_п''$ - соответственно темпы роста и темпы прироста сравниваемых динамических рядов.

С помощью этих коэффициентов могут сравниваться динамические ряды одинакового содержания, но относящиеся к разным территориям (странам, регионам, районам и т.п.) или к различным организациям (министерствам, предприятиям) или ряды разного содержания, характеризующие один и тот же объект. Например, по данным таблицы 8.3 и 8.4 рассчитаем коэффициенты опережения роста потребительских цен над ростом средней месячной заработной платы, в апреле - мае 1995 г. (средняя месячная заработная плата в декабре 1994 г. составила 354,2 тыс. руб.).

Таблица 8.6

Месяцы 1995 г.	Темп роста потребительских цен (в % к декабрю 1994 г.)	Темп роста средней месячной заработной платы (в % к декабрю 1994 г.)	Коэффициенты опережения, рассчитанные	
			по темпам роста	по темпам прироста
1	2	3	4	5
апрель	154	109	1,41	6,0
май	167	121	1,38	3,19
июнь	178	136	1,31	2,17

Коэффициенты опережения, рассчитанные разными способами, различны по величине (см. гр. 4 и 5 табл. 8.6), однако тенденция, которая отображена в их изменении, - одинакова: за апрель-июнь 1995 г. рост потребительских цен опережал рост средней месячной заработной платы, хотя размеры этого опережения имели тенденцию к снижению.

8.3. Средние характеристики ряда динамики

Для обобщающей характеристики динамики исследуемого явления за ряд периодов определяют различного рода средние показатели. Рассмотрим две категории этих показателей:

1) средние уровни ряда,

2) средние показатели изменения уровней ряда.

Метод расчета **среднего уровня ряда** динамики зависит от вида временного ряда.

Для *интервального* ряда динамики абсолютных показателей средний уровень за период определяется по формуле простой средней арифметической,

$$\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n}, \tag{8.6}$$

где n - число уровней ряда.

Средний уровень *моментного* динамического ряда определяется несколько иначе. Рассмотрим вычисление среднего уровня для такого моментного динамического ряда, когда промежутки между датами одинаковы. Например, определим средний месячный остаток материалов на складе предприятия в течение первого квартала текущего года, если известно, что остаток материалов на складе на 1 января составил 242 млн руб., на 1 февраля - 251 млн , на 1 апреля - 186 млн руб. При вычислении среднего уровня моментного ряда условно предполагается непрерывное, равномерное изменение уровня в промежутках между двумя датами. Основываясь на этом предположении, определяем средние остатки материалов на складе за каждый месяц как полусумму остатков на начало и конец месяца. Средние остатки за месяц соответственно будут равны:

за январь - 246,5 млн руб. ($\frac{242+251}{2}$), за февраль - 232,0 млн руб. ($\frac{251+213}{2}$) и за март - 192,5 млн руб. ($\frac{213+186}{2}$).

Средний остаток за квартал определяется как простая средняя арифметическая $\bar{y} = \frac{246{,}5+232{,}0+192{,}5}{3} = 229$ млн руб.

Приведенный расчет среднего уровня можно представить формулой средней хронологической:

$$\bar{y}_{xp} = \frac{\frac{y_1}{2} + y_2 + y_3 + ... + y_{n-1} + \frac{y_n}{2}}{n-1} \quad (8.7)$$

где n - число дат, $y_1, y_2, ..., y_n$ - уровни ряда в последовательные моменты времени.

В нашем примере средние остатки на складе будут равны:

$$\bar{y}_{xp} = \frac{\frac{242}{2} + 251 + 213 + \frac{186}{2}}{4-1} = 229 \text{ млн руб.}$$

Для определения среднего уровня моментного ряда с неравными промежутками между временными датами вычисляется средняя арифметическая взвешенная; в качестве весов принимается продолжительность промежутков времени между моментами, в которые происходят изменения в уровнях динамического ряда:

$$\bar{y} = \frac{\sum_{i=1}^{n} y_i t_i}{\sum_{i=1}^{n} t_i}, \quad (8.8)$$

где t_i - количество дней (месяцев) между смежными датами.

Например на 1 января отчетного года стоимость основных средств предприятий составляла 75 млрд руб. В марте были приобретены основные средства на сумму 2 млрд руб., в мае выбыло основных средств на 7 млрд руб., а в сентябре было приобретено еще основных средств на 8 млрд руб. Определим среднюю годовую стоимость основных средств предприятия. Для удобства расчетов данные представим в табл. 8.7.

Таблица 8.7

Даты времени	Стоимость основных средств, млрд руб.	Число месяцев, в течение которых стоимость не изменялась	$y_i t_i$
1/1	75	3	225
I/IV	77	2	154
I/VI	70	4	280
I/X	78	3	234
Итого		12	893

Среднегодовая стоимость основных средств предприятия по данным приводимого примера составит 74,417 млрд руб.

$$\bar{y} = \frac{893}{12} = 74,417 \text{ млрд руб.}$$

Если же ориентироваться на стоимость основных средств предприятия на начало и конец отчетного года, т.е. например использовать показатели балансового отчета предприятия, то получим иной показатель среднегодовой стоимости основных средств предприятия:

$$\bar{y} = \frac{75+78}{2} = 76,5 \text{ млрд руб.},$$

где 78 млрд руб. стоимость основных средств на конец года (78=75+2-7+8).

Вот почему для анализа деятельности предприятия следует пользоваться данными внутренней отчетности, что позволяет получить более достоверную оценку результатов.

При определении средних уровней временного ряда нужно иметь в виду, что средняя будет достаточно надежной характеристикой ряда динамики, если она характеризует период с более или менее стабильными условиями развития. Если же за исследуемый период можно выделить этапы, в течение которых условия развития существенно менялись, то пользоваться общей средней не всегда целесообразно, а предпочтение нужно отдать средним, рассчитанным по отдельным этапам.

Средний абсолютный прирост (или средняя скорость роста) рассчитывается как средняя арифметическая из показателей скорости роста за отдельные промежутки времени.

$$\bar{\Delta} = \frac{\sum_{i=1}^{n-1} \Delta_i}{n-1}, \quad (8.9)$$

где n - число уровней ряда;

Δ_i - абсолютные изменения по сравнению с предшествующим уровнем.

По данным приводимого в табл. 8.5 примера,

$$\bar{\Delta} = \frac{-1,1}{4} = -0,275 \text{ млн кв. м}$$

Это означает, что в течение 1990–1994 гг. в среднем строительство жилья, выполняемого жилищно-строительными кооперативами в России, снижалось на 275 тыс. кв. м.

Формулу 8.9 можно преобразовать. Как уже было отмечено на с. 290, $\sum_{i=1}^{n-1} \Delta_i = y_n - y_1$, тогда

$$\overline{\Delta} = \frac{y_n - y_1}{n-1}, \qquad (8.9а)$$

где y_n и y_1 - соответственно конечный и начальный уровни динамического ряда.

Средний коэффициент роста вычисляется по формуле средней геометрической из показателей коэффициентов роста за отдельные периоды:

$$\overline{K} = \sqrt[n-1]{K_1 \cdot K_2 \cdot ... \cdot K_{n-1}}, \qquad (8.10)$$

где $K_1, K_2, ..., K_{n-1}$ - коэффициенты роста по сравнению с уровнем предшествующего периода;

n - число уровней ряда.

Эту формулу можно записать иначе, если использовать взаимосвязь между коэффициентами роста, вычисленными с переменной базой, и коэффициентами роста, рассчитанными с постоянной базой, т.е. учитывая, что $K_1 \cdot K_2 \cdot K_3 \cdot ... \cdot K_{n-1} = \frac{y_n}{y_1}$, средний коэффициент роста можно определить по формуле:

$$\overline{K} = \sqrt[n-1]{\frac{y_n}{y_1}}. \qquad (8.10а)$$

По данным примера таблицы 8.5, средний коэффициент роста будет равен:

$$\overline{K} = \sqrt[4]{\frac{1,8}{2,9}} = \sqrt[4]{0,6207} = 0,8876,$$

т.е. формула (8.10а) позволяет определить средний коэффициент роста на основе данных о начальном и конечном уровнях ряда динамики. Средний коэффициент роста показывает, во сколько раз в среднем за отдельные составляющие рассматриваемого периода изменялись уровни динамического ряда.

Средний темп роста представляет собой средний коэффициент роста, выраженный в процентах.

$$\overline{T} = \overline{K} \cdot 100\%,$$

где $\overline{K}$ - средний годовой коэффициент роста. Для приводимого примера $\overline{T}$ = 88,76%, а это означает, что в среднем ежегодно строительство жилья составляло 88,76% к уровню предыдущего года. Более доходчива интерпретация относительных изменений с помощью среднего темпа прироста. $\overline{T}_п = \overline{T}$ - 100%, т.е., продолжая пример, можно сказать, что в среднем ежегодно строительство жилья жилищно-строительными кооперативами в России за 1990-1994 гг. снижалось на 11,24%.

Для практического применения средний темп роста, рассчитанный по данным о конечном и начальном уровнях временного ряда, можно использовать только в случае более или менее равномерного изменения уровней. В случае сильной колеблемости уровней ряда использование средней геометрической может дать искаженное выражение средней интенсивности изменения их уровней. В таких случаях предлагается средний коэффициент роста рассчитывать как параболический средний темп роста, средний геометрический темп роста выравненного ряда показательной функции и т.д.

Особую осторожность при применении средних абсолютных приростов или средних темпов роста (прироста) следует соблюдать в тех случаях, когда появляется перелом в имевшей место тенденции изменения уровней динамического ряда. Например, известны данные о производстве приборов по месяцам полугодия.

Таблица 8.8

	Январь	Февраль	Март	Апрель	Май	Июнь
Произведено приборов, (тыс. шт.)	2,0	2,1	2,3	2,6	2,8	1,9
Темпы роста (в % к предыдущему месяцу)	100,0	105,0	109,5	113,0	107,7	67,9

Средний месячный коэффициент роста объема выпуска за первое полугодие составит 92,54% ($\sqrt[5]{0,679} \cdot 100\%$), т.е. по результатам расчетов в среднем за полугодие ежемесячно выпуск продукции снижался на 7,46%. Этот результат противоречит приведенным темпам роста объема выпуска за январь-май отчетного года. В действительности за январь-май ежемесячно объем продукции увеличивался в среднем на 8,8% $(\sqrt[4]{\frac{2,8}{2,0}} \cdot 100\% - 100\%)$, что наглядно видно на рис. 8.3.

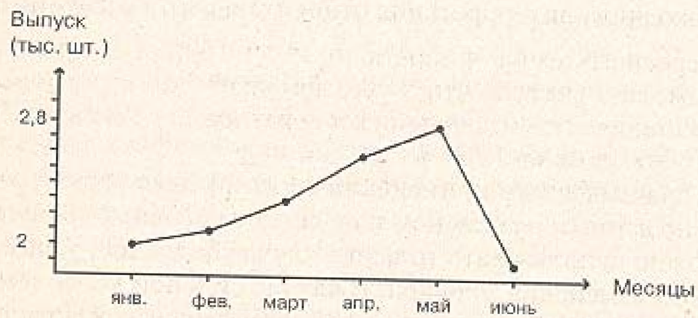

Рис. 8.3. **Производство приборов за первое полугодие**

Резкий спад объема производства приборов в июне может означать либо резкое изменение общей тенденции, либо влияние каких-то случайных факторов. Поэтому в данном случае можно ориентироваться на средний темп роста за период с января по май, когда средний месячный темп роста составит $(\sqrt[4]{\frac{2,8}{2,0}} \cdot 100\%)$ 108,8%. Для последующих периодов следует выполнить самостоятельный анализ.

8.4. Выявление и характеристика основной тенденции развития

Одной из задач, возникающих при анализе рядов динамики, является установление закономерности изменения уровней изучаемого показателя во времени.

В некоторых случаях эта закономерность, общая тенденция развития объекта вполне ясно отображается уровнями динамического ряда. Так, в приведенном в табл. 8.5 примере уровням динамического ряда свойственна тенденция к снижению, не нару-

шаемая на протяжении всего рассматриваемого периода. В других рядах динамики наблюдается систематическое увеличение уровней ряда (например, при изучении динамики средней заработной платы по данным табл. 8.4). Однако часто приходится встречаться с такими рядами динамики, когда уровни ряда претерпевают самые различные изменения (то возрастают, то убывают) и можно говорить лишь об общей тенденции развития явления, либо о тенденции к росту, либо к снижению. В этих случаях для определения основной тенденции развития явления, достаточно устойчивой на протяжении данного периода, используют особые приемы обработки рядов динамики.

Уровни ряда динамики формируются под совокупным влиянием множества длительно и кратковременно действующих факторов и в том числе различного рода случайных обстоятельств. Выявление основной закономерности изменения уровней ряда предполагает ее количественное выражение, в некоторой мере свободное от случайных воздействий. *Выявление основной тенденции развития (тренда) называется в статистике также выравниванием временного ряда, а методы выявления основной тенденции - методами выравнивания.* Выравнивание позволяет характеризовать особенность изменения во времени данного динамического ряда в наиболее общем виде как функцию времени, предполагая, что через время можно выразить влияние всех основных факторов.

Один из наиболее простых приемов обнаружения общей тенденции развития явления - **укрупнение интервала динамического ряда**. Смысл приема заключается в том, что первоначальный ряд динамики преобразуется и заменяется другим, показатели которого относятся к бо́льшим по продолжительности периодам времени. Например, ряд, содержащий данные о месячном выпуске продукции, может быть преобразован в ряд квартальных данных. Вновь образованный ряд может содержать либо абсолютные величины за укрупненные по продолжительности промежутки времени (эти величины получают путем простого суммирования уровней первоначального ряда абсолютных величин), либо средние величины. При суммировании уровней или при выведении средних по укрупненным интервалам отклонения в уровнях, обусловленные случайными причинами, взаимопогашаются, сглаживаются и более четко обнаруживается действие основных факторов изменения уровней (общая тенденция).

Выявление основной тенденции может быть осуществлено также **методом скользящей средней**. Для определения скользящей средней формируем укрупненные интервалы, состоящие из одинакового числа уровней. Каждый последующий интервал получаем, постепенно сдвигаясь от начального уровня динамического ряда на один уровень. Тогда первый интервал будет включать уровни $y_1, y_2, \ldots y_m$; второй - уровни $y_2, y_3, \ldots y_{m+1}$ и т.д. Таким образом, интервал сглаживания как бы скользит по динамическому ряду с шагом, равным единице. По сформированным укрупненным интервалам определяем сумму значений уровней, на основе которых рассчитываем скользящие средние. Полученная средняя относится к середине укрупненного интервала. Поэтому при сглаживании скользящей средней технически удобнее укрупненный интервал составлять из нечетного числа уровней ряда. Нахождение скользящей средней по четному числу уровней создает неудобство, вызываемое тем, что средняя может быть отнесена только к середине между двумя датами. В этом случае необходима дополнительная процедура центрирования средних (подробнее см. пример на с. 329). Покажем порядок расчета скользящих средних, используя данные о дневном выпуске продукции предприятия за месяц (см. табл. 8.9).

Возьмем в качестве укрупненного интервала период в 3 дня, тогда первая скользящая сумма будет равна объему выпуска за первый, второй и третий рабочие дни, вторая скользящая сумма - за второй, третий и четвертый рабочие дни и т.д. В табл. 8.9 в гр. 3 и 4 приведены скользящие суммы за трехдневный и пятидневный промежутки. Скользящая средняя, рассчитанная по трехдневным скользящим суммам будет отнесена ко второму дню каждой трехдневки (см. гр. 5). Скользящая же средняя, рассчитанная по пятидневным суммам (см. гр. 6), относится к третьему дню соответствующей пятидневки. Как следует подходить к выбору интервала сглаживания? Нередко выбор интервала сглаживания осуществляется произвольно, однако при этом нужно учитывать количество уровней в анализируемом ряду динамики, так как при использовании приема скользящей средней сглаженный ряд сокращается по сравнению с исходным рядом на число уровней, равное (m - 1). Вместе с тем, чем продолжительнее интервал сглаживания, тем сильнее усреднение, а потому выявляемая тенденция развития получается более плавной. Чаще всего интервал сглаживания может состоять из трех, пяти или семи уровней. Первоначальные и выравненные динамические ряды с помощью скользящих средних изображены на рис. 8.4.

Таблица 8.9

Динамика выпуска продукции предприятия по дням отчетного месяца

Рабочие дни месяца	Выпуск продукции, млн руб.	Скользящие суммы, млн руб.		Скользящие средние, млн руб.	
		3-х дневные	5-ти дневные	3-х дневные	5-ти дневные
1	2	3	4	5	6
1	37	-	-	-	-
2	42	112	-	37,3	-
3	33	120	215	40,0	43,0
4	45	136	233	45,3	46,6
5	58	158	247	52,7	49,4
6	55	169	284	56,3	56,8
7	56	181	308	60,3	61,6
8	70	195	324	65,0	64,8
9	69	213	324	71,0	64,0
10	74	214	340	71,3	68,0
11	71	231	370	77,0	74,0
12	86	227	370	75,7	74,0
13	70	248	393	82,7	78,6
14	92	230	387	76,7	77,4
15	68	253	404	84,3	80,8
16	93	242	423	80,7	84,6
17	81	263	425	87,7	85,0
18	89	264	460	88,0	92,0
19	94	286	476	95,3	95,2
20	103	306	494	102,0	98,8
21	109	311	516	103,7	103,3
22	99	319	-	106,3	-
23	111	-	-	-	-

Изучение основной тенденции развития методом скользящей средней является лишь эмпирическим приемом предварительного анализа. Рассмотренные приемы сглаживания динамических рядов (укрупнение интервала и метод скользящей средней) могут рассматриваться как важное вспомогательное средство, облегчающее применение других методов и, в частности, более строгих методов выявления тенденции. Для того чтобы представить количественную модель, выражающую общую тенденцию изменений уровней динамического ряда во времени, используется **аналитическое выравнивание ряда динамики**. В этом случае факти-

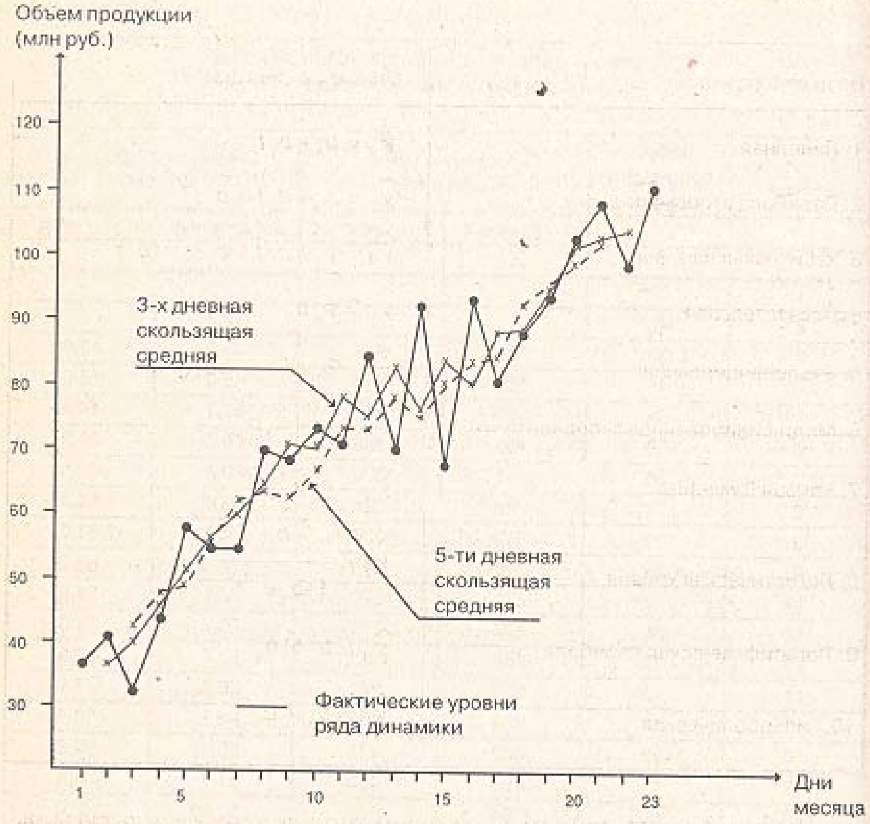

Рис. 8.4. **Сглаженный ряд динамики объема выпуска продукции**

ческие уровни заменяются уровнями, вычисленными на основе определенной кривой. Предполагается, что она отражает общую тенденцию изменения во времени изучаемого показателя.

При аналитическом выравнивании ряда динамики закономерно изменяющийся уровень изучаемого показателя оценивается как функция времени $\hat{y}_t = f(t)$, где $\hat{y}_t$ - уровни динамического ряда, вычисленные по соответствующему аналитическому уравнению на момент времени t. В табл. 8.10 приводятся различные виды трендовых моделей, наиболее часто используемые для аналитического выравнивания.

Таблица 8.10

Название функции	Описание функции
1. Линейная	$\hat{y}_t = b_0 + b_1 t$
2. Парабола второго порядка	$\hat{y}_t = b_0 + b_1 t + b_2 t^2$
3. Кубическая парабола	$\hat{y}_t = b_0 + b_1 t + b_2 t^2 + b_3 t^3$
4. Показательная	$\hat{y}_t = b_0 b_1^t$
5. Экспоненциальная	$\hat{y}_t = b_0 e^{b_1^t}$
6. Модифицированная экспонента	$\hat{y}_t = b_0 + b_1 b_2^t$
7. Кривая Гомперца	$\hat{y}_t = b_0 b_1^{b_2^t}$
8. Логистическая кривая	$\hat{y}_t = \dfrac{b_0}{1+b_1 e^{-b_2 t}}$
9. Логарифмическая парабола	$\hat{y}_t = b_0 b_1^t b_2^{t^2}$
10. Гиперболическая	$\hat{y}_t = b_0 + b_1 \dfrac{1}{t}$

Выбор формы кривой во многом определяет результаты экстраполяции тренда. Основанием для выбора вида кривой может использоваться содержательный анализ сущности развития данного явления. Можно опираться также на результаты предыдущих исследований в данной области.

На практике для этих целей прибегают к анализу графического изображения уровней динамического ряда (линейной диаграммы). Однако из графического представления эмпирических данных не всегда удается произвести однозначный выбор формы уравнения. Поэтому целесообразно воспользоваться графическим изображением сглаженных уровней, в которых случайные и волнообразные колебания в некоторой степени оказываются погашенными (см. рис. 8.4).

При выборе вида кривой для выравнивания динамического ряда возможно также использование **метода конечных разностей**, который основан на свойствах различных кривых, применяемых при выравнивании [1].

Свойства конечных разностей заключаются в следующем:

1. Если общая тенденция выражается линейным уравнением

$$\hat{y}_t = b_0 + b_1 t, \tag{8.11}$$

тогда получаем

постоянными первые разности: $\Delta_i' = y_i - y_{i-1}$,

нулевыми вторые разности $\Delta_i'' = \Delta_i' - \Delta_{i-1}'$.

Равенство Δ' постоянной величине и Δ'' нулю при вычислении их по фактическим данным рассматривается не для каждого отдельного случая, а лишь в целом по всей совокупности уровней. Так, при

$$\begin{array}{lllll} t=0, & y_0 = b_0, & & & \\ & & \Delta_1' = y_1 - y_0 = b_1 & & \\ t=1, & y_1 = b_0 + b_1, & & \Delta_1'' = \Delta_2' - \Delta_1' = 0, & \\ & & \Delta_2' = y_2 - y_1 = b_1 & & \\ t=2, & y_2 = b_0 + 2b_1, & & \Delta_2'' = \Delta_3' - \Delta_2' = 0, & \\ & & \Delta_3' = y_3 - y_2 = b_1 & & \\ t=3, & y_3 = b_0 + 3b_1. & & & \end{array} \tag{8.11a}$$

2. Если тенденция выражается параболой второго порядка $\hat{y}_t = b_0 + b_1 t + b_2 t^2$ (8.12), то получим постоянными - вторые разности, нулевыми - третьи разности.

Если рассмотрим уравнение параболы при разных t, то получим при:

$$\begin{array}{lllll} t=0, & y_0 = b_0, & & & \\ & & \Delta_1' = b_1 + b_2; & & \\ t=1, & y_1 = b_0 + b_1 + b_2, & & \Delta_1'' = 2b_2; & \\ & & \Delta_2' = b_1 + 3b_2, & & \Delta_1''' = 0, \\ t=2, & y_2 = b_0 + 2b_1 + 4b_2, & & \Delta_2'' = 2b_2; & \\ & & \Delta_3' = b_1 + 5b_2, & & \Delta_2''' = 0, \\ t=3, & y_3 = b_0 + 3b_1 + 9b_2, & & \Delta_3'' = 2b_2, & \\ & & \Delta_4' = b_1 + 7b_2, & & \\ t=4, & y_4 = b_0 + 4b_1 + 16b_2 & & & \end{array} \tag{8.12a}$$

и т.д.

[1] Обязательным условием является равенство интервалов между уровнями динамического ряда.

Основываясь на указанных свойствах конечных разностей для различных видов кривых, Б.С. Ястремский [1] сделал вывод о применимости для выравнивания линейной функции, если любые три равностоящие уровня имеют нулевую вторую разность. Порядок разностей, остающихся примерно равными друг другу, принимается за степень выравнивающего многочлена, т.е. если примерно одну и ту же величину имеют вторые разности, то для выравнивания используется парабола второго порядка.

Можно указать и ряд других признаков, которые могут помочь при выборе формы кривой: если примерно постоянными оказываются темпы роста, то для выравнивания применяется показательная функция; если первые разности имеют тенденцию уменьшаться с постоянным темпом, то следует остановиться на модифицированной экспоненте; если средние уровни, нанесенные на полулогарифмическую сетку, близки к прямой линии, то предпочтительнее простая экспонента; если первые разности обратных значений средних уровней изменяются на один и тот же процент, то следует остановиться на логистической кривой.

При выборе формы уравнения следует исходить и из объема имеющейся информации. Чем больше параметров содержит уравнение тренда, тем больше должно быть наблюдений при одной и той же степени надежности оценивания.

Выбор формы кривой может осуществляться и на основе принятого критерия, в качестве которого может служить сумма квадратов отклонений фактических значений от значений, рассчитанных по уравнению тренда. Из совокупности кривых выбирается та, которой соответствует минимальное значение критерия.

Рассмотрим аналитическое выравнивание ряда динамики по прямой, т.е. аналитическое уравнение вида:

$$y_t = b_0 + b_1 t,$$

где t - порядковый номер периодов или моментов времени.

Параметры b_0 и b_1 прямой рассчитываются по *методу наименьших квадратов* (МНК). Система нормальных уравнений в данном случае имеет вид:

[1] Некоторые вопросы математической статистики. 3-е изд. М., Госстатиздат. 1961.

$$\begin{cases} \sum\limits_1^n y_i = b_0 n + b_1 \sum\limits_1^n t_i \, ; \\ \sum\limits_1^n y_i t_i = b_0 \sum\limits_1^n t_i + b_1 \sum\limits_1^n t_i^2 \, . \end{cases}$$

Поиск параметров уравнения можно упростить, если отсчет времени производить так, чтобы сумма показателей времени изучаемого ряда динамики была равна нулю ($\sum\limits_1^n t_i = 0$). При нечетном числе уровней ряда динамики для получения $\sum\limits_{i=1}^n t_i = 0$ уровень, находящийся в середине ряда, принимается за условное начало отсчета времени (этому периоду или моменту времени придается нулевое значение). Даты времени, стоящие выше этого уровня, обозначаются натуральными числами со знаком минус (–1, –2, –3 и т.д.), а ниже - натуральными числами со знаком плюс (+1, +2, +3 и т.д. (см. гр. 3 табл. 8.11).

Если число уровней динамического ряда четное, периоды времени верхней половины ряда (до середины) нумеруются –1, –3, –5 и т.д., а нижней – +1, +3, +5, и т.д. При этом условии $\sum\limits_1^n t_i$ будет равна нулю и система нормальных уравнений преобразуется следующим образом:

$$\begin{cases} \sum\limits_1^n y_i = b_0 n \, ; \\ \sum\limits_1^n y_i t_i = b_1 \sum\limits_1^n t_i^2 , \end{cases} \quad \text{откуда} \quad b_0 = \frac{\sum\limits_1^n y_i}{n} = \bar{y} \, ; \; b_1 = \frac{\sum\limits_1^n y_i t_i}{\sum\limits_1^n t_i^2} \, .$$

Рассмотрим аналитическое выравнивание по прямой ряда динамики строительства жилья жилищно-строительными кооперативами в России (см. данные табл. 8.5). Расчет параметров уравнения прямой представлен в табл. 8.11.

Таблица 8.11

Годы	Построено жилищно-строительными кооперативами, млн кв.м y_i	Условные обозначения периодов, t_i	$y_i t_i$	t_i^2	Выравненные уровни ряда динамики, млн кв. м $\hat{y}_t$	$y_i - \hat{y}_t$	$(y_i - \hat{y}_t)^2$
1	2	3	4	5	6	7	8
1990	2,9	-2	-5,8	4	2,76	0,14	0,0196
1991	2,4	-1	-2,4	1	2,49	-0,09	0,0081
1992	2,1	0	0	0	2,22	-0,12	0,0144
1993	1,9	+1	1,9	1	1,95	-0,05	0,0025
1994	1,8	+2	3,6	4	1,68	0,12	0,0144
Итого	11,1		-2,7	10	11,10	0,00	0,0590

Используя итоги граф 2, 4 и 5,определим параметры уравнения прямой:

$$b_0 = \frac{11{,}1}{5} = 2{,}22 \, ; \quad b_1 = \frac{-2{,}7}{10} = -0{,}27 \, .$$

По рассчитанным параметрам записываем уравнение прямой ряда динамики, характеризующего строительство жилья жилищно-строительными кооперативами:

$$\hat{y}_t = 2{,}22 - 0{,}27t \, .$$

Используя приведенное уравнение, рассчитаем для каждого года теоретические значения:

для 1990 г. $\hat{y}_{t=-2} = 2{,}22 - 0{,}27 \cdot (-2) = 2{,}76$ млн. кв. м,

для 1991 г. $\hat{y}_{t=-1} = 2{,}22 - 0{,}27 \cdot (-1) = 2{,}49$ млн кв. м и т.д. (см. гр. 6 табл. 8.11).

Правильность расчета уровней выравниваемого ряда динамики может быть проверена следующим образом: сумма значений эмпирического ряда должна совпадать с суммой вычисленных уровней выравненного ряда, т.е. $\sum\limits_{i=1}^n y_i = \sum\limits_{i=1}^n \hat{y}_t$ (см. итоги гр. 2 и 6).

Продление в будущее тенденции, наблюдавшейся в прошлом, носит название *экстраполяции*. Экстраполируя при $t = 3$, находим уровень 1995 г., равный 1,41 млн кв. м (2,22 - 0,27 • 3).

Возможность экстраполяции обеспечивается двумя обстоятельствами:

1) общие условия, определяющие тенденцию развития в прошлом, не претерпевают существенных изменений в будущем;

2) тенденция развития явления характеризуется тем или иным аналитическим уравнением.

Общая тенденция развития может быть охарактеризована с помощью содержательного экономического анализа. Вместе с тем расчет таких показателей, как скорость роста, темпы роста, пункты роста, позволяет ориентироваться в наличии или отсутствии устойчивой тенденции развития и обосновать форму уравнения тренда. Если условия формирования уровней ряда изменяются, то расчет параметров уравнения не следует вести по данным за весь рассматриваемый период времени. В этом случае целесообразно разбить ряд динамики на ряд этапов, ориентируясь на устойчивость абсолютных приростов или пунктов роста. Значение $\hat{y}_t$, полученное в результате экстраполяции, используют для определения прогнозного значения на будущее.

При составлении прогнозов оперируют не точечной, а интервальной оценкой, определяя так называемые *доверительные интервалы прогноза*. Величина доверительного интервала определяется в общем виде так:

$$\hat{y}_t \pm t_\alpha \frac{s_{\hat{y}}}{\sqrt{n}}, \qquad (8.13)$$

где $s_{\hat{y}}$ - среднее квадратическое отклонение от тренда;

t_α - табличное значение t-критерия Стьюдента при уровне значимости α.

Величина $s_{\hat{y}}$ определяется по формуле:

$$s_{\hat{y}} = \sqrt{\frac{\sum_{1}^{n}\left(y_i - \hat{y}_t\right)^2}{n-m}}, \qquad (8.14)$$

где y_i и $\hat{y}_t$ - соответственно фактические и расчетные значения уровней динамического ряда;

n - число уровней ряда;

m - количество параметров в уравнении тренда (для уравнения прямой $m = 2$).

Используя данные гр. 8 табл. 8.11, рассчитаем среднюю квадратическую ошибку линейного уравнения тренда:

$$s_{\hat{y}} = \sqrt{\frac{0,0590}{5-2}} = 0,14 \text{ млн кв. м.}$$

Отсюда величина относительной ошибки составляет 6,32% ($\frac{s_{\hat{y}}}{\bar{y}} \cdot 100\% = \frac{0,14}{2,22} \cdot 100\%$).

Величина средней квадратической ошибки может быть рассчитана иным способом, позволяющим избежать ошибки при округлениях величины ($y_i - \hat{y}_t$). Для уравнения линейного тренда $\hat{y}_t = b_0 + b_1 t$ величина $s_{\hat{y}}$ определяется по формуле:

$$s_{\hat{y}} = \sqrt{\frac{\sum_{i=1}^{n} y_i^2 - b_0 \sum_{i=1}^{n} y_i - b_1 \sum_{i=1}^{n} y_i t}{n-2}}. \qquad (8.14а)$$

Для примера, приведенного в табл. 8.11, получим $s_{\hat{y}} = 0,14$ млн кв. м.

$$\left[s_{\hat{y}} = \sqrt{\frac{25,43 - 2,22 \cdot 11,1 - 0,27 \cdot 2,7}{5-2}} = 0,14 \text{ млн кв. м} \right]$$

Если воспользоваться методом конечных разностей для выбора формы уравнения тренда, то из данных гр. 4 табл. 8.5 можно видеть, что абсолютные приросты не являются постоянными, а начиная с 1992 г. постоянными будут вторые разности (разности абсолютных приростов). В этом случае, как было показано выше, для выравнивания может использоваться парабола второго порядка:

$$\hat{y}_t = b_0 + b_1 t + b_2 t^2.$$

Система нормальных уравнений для нахождения параметров уравнения параболы (при соблюдении принципа отсчета от условного начала) будет иметь вид:

$$\begin{cases} \sum_{1}^{n} y_i = b_0 n + b_2 \sum_{1}^{n} t_i^2; \\ \sum_{1}^{n} y_i t_i = b_1 \sum_{1}^{n} t_i^2; \\ \sum_{1}^{n} y_i t_i^2 = b_0 \sum_{1}^{n} t_i^2 + b_2 \sum_{1}^{n} t_i^4. \end{cases}$$

Расчет параметров этого уравнения тренда представлен в табл. 8.12.

Таблица 8.12

Годы	y_i	t_i	t_i^2	$y_i t_i$	$y_i t_i^2$	t_i^4	$\hat{y}_t$	$y_i - \hat{y}_t$
1	2	3	4	5	6	7	8	9
1990	2,9	-2	4	-5,8	11,6	16	2,889	+0,011
1991	2,4	-1	1	-2,4	2,4	1	2,426	-0,026
1992	2,1	0	0	0	0	0	2,091	+0,009
1993	1,9	+1	1	1,9	1,9	1	1,886	+0,014
1994	1,8	+2	4	3,6	7,2	16	1,809	-0,009
Итого	11,1	0	10	-2,7	23,1	34	11,10	0,00

Подставляем итоги гр. 2, 4, 5, 6 и 7 табл. 8.12 и получаем следующую систему уравнений для данного временного ряда:

$$\begin{cases} 5b_0 + 10b_2 = 11{,}1 \\ 10b_1 = -2{,}7 \\ 10b_0 + 34b_2 = 23{,}1 \end{cases}$$

Решая систему уравнений, определим значения параметров уравнения параболы второго порядка:

$$b_0 = 2{,}0914;\ b_1 = -0{,}27;\ b_2 = 0{,}0643.$$

Отсюда уравнение параболы второго порядка, характеризующего тенденцию строительства жилья жилищно-строительными кооперативами, будет записано так:

$$\hat{y} = 2{,}0914 - 0{,}27t + 0{,}0643t^2.$$

Величина среднего квадратического отклонения фактических уровней динамического ряда от выравненных для уравнения параболы второго порядка определится по формуле:

$$s_{\hat{y}} = \sqrt{\frac{\sum_{i=1}^{n} y_i^2 - b_0 \sum_{i=1}^{n} y_i - b_1 \sum_{i=1}^{n} y_i t_i - b_2 \sum_{i=1}^{n} y_i t^2}{n-3}}, \qquad (8.14б)$$

т.е.

$$s_{\hat{y}} = \underset{\text{млн кв. м}}{0{,}02377} \left[\sqrt{\frac{25{,}43 - 2{,}0914 \cdot 11{,}1 - 0{,}27 \cdot 2{,}7 - 0{,}0643 \cdot 23{,}1}{5-3}} = \sqrt{0{,}00057} = 0{,}02377 \right],$$

относительная ошибка уравнения составит 1,07%. Сравнив полученные значения $s_{\hat{y}}$ для уравнения прямой и параболы второго порядка, можно сделать вывод о том, что парабола более точно описывает основную тенденцию ряда динамики, характеризующего строительство жилья жилищно-строительными кооперативами.

Прогноз на 1995 г. по уравнению параболы второго порядка будет составлен следующим образом:

1) экстраполируя для t=3, получаем $\hat{y}_t$ = 1,860 млн кв. м;

2) значение критерия Стьюдента при уровне значимости 5% и числе степеней свободы $n-m$ = 2 равно 4,303 (см. Приложение IV);

3) величина $t_\alpha \cdot \frac{s_{\hat{y}}}{\sqrt{n}} = 4{,}303 \cdot \frac{0{,}02377}{\sqrt{5}} = 0{,}04574$ млн кв. м;

4) $1{,}860 - 0{,}046 \leqslant y_{\text{прогноз.}} \leqslant 1{,}860 + 0{,}046$

$$1{,}814 \leqslant y_{\text{прогноз.}} \leqslant 1{,}906 \text{ млн кв. м.}$$

Таким образом, с вероятностью 95% можно ожидать, что в 1995 г. жилищно-строительные кооперативы должны построить жилья не меньше, чем 1,814 млн кв. м, но не больше 1,906 млн кв. м.

Приведенные расчеты следует рассматривать не как завершающую стадию прогнозирования, а лишь как предварительный этап в разработке прогноза. Для составления прогноза должна быть привлечена дополнительная информация, не содержащаяся в самом динамическом ряду.

Если в изменениях уровней обнаруживается тенденция к постоянству темпов роста, то выравнивание ряда следует проводить по показательной кривой:

$$\hat{y}_t = b_0 b_1^t, \qquad (8.15)$$

где b_1 - коэффициент роста.

Техника выравнивания по показательной кривой аналогична технике выравнивания по прямой, за исключением того, что выравниваются здесь не уровни ряда, а их логарифмы:

$$\begin{cases} \sum_{1}^{n} \lg y = n \lg b_0 + \lg b_1 \sum_{1}^{n} t; \\ \sum_{1}^{n} \lg y \cdot t = \lg b_0 \cdot \sum_{1}^{n} t + \lg b_1 \sum_{1}^{n} t^2. \end{cases}$$

Приравнивая $\Sigma t = 0$, получаем:

$$\begin{cases} \sum \lg y = n \lg b_0; \\ \sum \lg y \cdot t = \lg b_1 \sum t^2. \end{cases} \tag{8.15a}$$

По вычисленным значениям логарифмов определяем величины параметров уравнения показательной кривой b_0 и b_1.

Современные компьютерные программы по анализу временных рядов позволяют автоматически определять тип модели, адекватной исходным данным, на основе соответствующего критерия.

Аналитическое уравнение представляет собой математическую модель развития явления и дает выражение статистической закономерности, проявляющейся в рядах динамики. Следует помнить, что прием аналитического выравнивания содержит в себе ряд условностей, связанных прежде всего с тем, что уровни, характеризующие тот или иной динамический ряд, рассматриваются как функция времени. В действительности же развитие явлений обусловлено не тем, сколько времени прошло с отправного момента, а тем, какие факторы влияли на развитие, в каком направлении и с какой интенсивностью. Развитие явлений во времени выступает как внешнее выражение этих факторов, как их суммарное действие, оказывающее влияние на изменение уровня в отдельно взятые промежутки или моменты времени. Выявить основную тенденцию развития явления методом наименьших квадратов можно лишь тогда, когда выяснено, что изменяющиеся во времени процессы протекают на всем рассматриваемом промежутке времени одинаково, что их количественное и качественное изменение происходит под действием одного и того же комплекса основных факторов, определяющих движение данного ряда динамики.

Модели, учитывающие общие закономерности изменения экономического явления в изучаемый интервал времени и изменения во времени влияния комплекса факторов, называют *многофакторными динамическими моделями.*

Допустим, что величина исследуемого показателя y зависит от изменения нескольких факторов $x_1, x_2, x_3, \ldots, x_m$. Располагая дан-

ными по некоторой совокупности объектов за ряд лет, можно построить корреляционную модель, характеризующую зависимость y от указанных факторов для каждого периода времени.

Предположим, что зависимость может быть представлена линейной функцией. Тогда модель будет иметь вид:

для периода $t = 1$: $\hat{y}_1 = a_{01} + a_{11}x_1 + a_{21}x_2 + \ldots + a_{m1}x_m$;

для периода $t = 2$: $\hat{y}_2 = a_{02} + a_{12}x_1 + a_{22}x_2 + \ldots + a_{m2}x_m$;

для периода $t = 3$: $\hat{y}_3 = a_{03} + a_{13}x_1 + a_{23}x_2 + \ldots + a_{m3}x_m$

и т.д.

Для всех периодов получим систему из n уравнений, и для каждого из факторов будет n коэффициентов регрессии, т.е. будем иметь временные ряды для каждого из коэффициентов регрессии:

$$\begin{array}{l} a_{01}, a_{02}, a_{03}, \ldots, a_{0n}; \\ a_{11}, a_{12}, a_{13}, \ldots, a_{1n}; \\ a_{21}, a_{22}, a_{23}, \ldots, a_{2n}; \\ \ldots\ldots\ldots\ldots\ldots \\ a_{m1}, a_{m2}, a_{m3}, \ldots, a_{mn}. \end{array}$$

Рассматривая каждый из этих временных рядов, можно представить a_m как функцию времени и, используя аналитическое выравнивание, построить прогнозы коэффициентов регрессии на период времени (t), т.е. определить значение величин $a_{0t}, a_{1t}, a_{2t}, \ldots, a_{mt}$.

Тогда величина признака y на период t может быть представлена так:

$$\hat{y}_t = a_{0t} + a_{1t}x_1 + a_{2t}x_2 + \ldots + a_{mt}x_m.$$

Значения факторов $x_1, x_2, \ldots, x_m$ необходимо определить также на момент времени t. Обычно для этого используют либо контрольные цифры, либо экстраполяцию по линии тренда.

Еще раз следует подчеркнуть, что прогноз уровня, характеризующего определенный объект с опорой на метод аналитического выравнивания, основан на предположении, что те же самые условия, в которых формировались уровни ряда в прошлом, будут существовать и в будущем. Использование экстраполяции в изменившихся условиях, появление новых факторов, оказывающих влияние на формирование уровней временного ряда, будет сопряжено с более или менее значительными ошибками. Причем экстраполяция на отдаленные даты подвержена более значительным ошибкам, чем краткосрочная экстраполяция. Это в известной степени может объясняться определенной инерционностью развития управляемых объектов: чем выше уровень управления, тем более инерционен объект управления. Однако с удалением периода прогноза от фактических уровней временного ряда влияние инерционности развития снижается и возрастает влияние новых условий и факторов. Поэтому целесообразно постоянно обновлять временные ряды по мере получения новых данных, что обусловливает и корректировку уравнений тренда.

Выделим особенности моделей аналитического выравнивания уровней динамического ряда, накладывающие ограничения на их использование. Во-первых, динамические ряды, к которым применяется аппроксимация, должны быть достаточно длинными. Во-вторых, применение аппроксимации наиболее целесообразно в случае медленно и плавно меняющегося уровня. В-третьих, аппроксимация как метод моделирования практически не адаптируется к изменяющимся условиям формирования уровней ряда; при появлении новых данных построение модели должно быть проведено заново. В-четвертых, при использовании для расчета параметров уравнения метода наименьших квадратов (МНК) считается, что значимость информации в пределах отрезка аппроксимации одинакова независимо от давности полученных данных, в то время как более поздние данные имеют большую ценность.

Помимо этого, динамические ряды экономических показателей часто имеют небольшую длину и подвержены значительным колебаниям, которые аппроксимация предвидеть не может. Поэтому в практике статистического анализа экономических процессов большое распространение получили **методы адаптивного моделирования и прогнозирования** [1].

В основе адаптивных методов лежит модель экспоненциального сглаживания, возможность использования которой для прогнозирования была доказана Р. Брауном. Сущность этого метода заключается в том, что временной ряд сглаживается с помощью взвешенной скользящей средней, в которой веса распределяются по экспоненциальному закону. Такая взвешенная скользящая средняя характеризует значения динамического ряда в конце интервала сглаживания, т.е. является характеристикой последних уровней ряда.

[1] Этот раздел параграфа написан доц. Бычковой С.Г.

Экспоненциальная средняя первого порядка для исходного ряда записывается следующим образом:

$$S_t^{(1)}(y) = \alpha \sum_{i=0}^{n} (1-\alpha)^i y_{t-i},$$

где $S^{(1)}{}_t(y)$ - экспоненциальная средняя первого порядка;

α - коэффициент сглаживания.

Экспоненциальная средняя *k*-го порядка, соответственно, определяется следующим образом:

$$S_t^{(k)}(y) = \alpha S_t^{(k-1)}(y) + (1-\alpha) S_{t-1}^{(k)}(y). \quad (8.16)$$

Коэффициенты полиномов, используемых для прогнозирования, могут быть получены через сглаженные значения ряда, и для линейной модели их формулы имеют следующий вид:

$$\begin{cases} a_0 = 2S_t^{(1)}(y) - S_t^{(2)}(y) \\ a_1 = \dfrac{\alpha}{1-\alpha}\left(S_t^{(1)}(y) - S_t^{(2)}(y)\right). \end{cases} \quad (8.17)$$

Рассмотрим последовательность построения линейной модели на основе данных о выпуске продукции предприятием в 1994 г. (табл. 8.13).

Начальные величины $S^{(1)}{}_{t-1}(y)$ и $S^{(2)}{}_{t-1}(y)$ могут быть получены исходя из (8.17) подстановкой параметров b_0 и b_1, полученных при выравнивании динамического ряда по уравнению тренда с использованием МНК.

С помощью аналитического выравнивания по прямой получаем $b_0 = 7{,}49$, $b_1 = 0{,}549$ и, соответственно, уравнение тренда $\hat{y}_t = 7{,}49 + 0{,}549\,t$. Коэффициент сглаживания α выбирается после содержательного анализа исследуемого процесса в зависимости от относительной ценности прошлых данных. Если необходимо придать больший вес последним данным, то значение α выбирается близким к единице, если необходимо учесть большую часть имеющихся данных, то берутся небольшие значения коэффициента сглаживания. В качестве метода выбора оптимального значения α может быть использован следующий: динамический ряд делится на две части; по первой части ряда для различных значений α строится модель и осуществляется прогнозирование на период, соответствующий длине второй части. Оптимальное значение α выбирается по минимальной среднеквадратической ошибке уравнения.

Таблица 8.13

Месяц	Выпуск продукции, тыс. т	$S^{(1)}_t(y)$	$S^{(2)}_t(y)$	a_0	a_1	Выравненные значения выпуска продукции, тыс. т $\hat{y}_t$
1	2	3	4	5	6	7
Январь	4,2					
Февраль	5,3	5,115	5,684	4,546	-1,326	3,220
Март	6,3	5,245	5,337	5,113	-0,308	4,805
Апрель	6,4	5,984	5,802	6,166	0,424	6,590
Май	5,7	6,275	6,133	6,417	0,331	6,748
Июнь	6,5	5,873	5,951	5,795	-0,182	5,613
Июль	6,4	6,312	6,204	6,420	0,252	6,168
Август	7,6	6,374	6,323	6,425	0,119	6,544
Сентябрь	8,4	7,232	6,959	7,505	0,636	8,141
Октябрь	10,1	8,050	7,723	8,377	0,762	9,139
Ноябрь	11,2	9,485	8,956	10,014	1,233	11,247
Декабрь	11,8	10,686	10,167	11,205	1,209	12,414

Для дальнейших расчетов используем α=0,7. Система уравнений для определений $S^{(1)}_{t-1}(y)$ и $S^{(2)}_{t-1}(y)$ будет иметь следующий вид:

$$\begin{cases} 7{,}49 = 2S^{(1)}_{t-1}(y) - S^{(2)}_{t-1}(y) \\ 0{,}549 = \dfrac{0{,}7}{1-0{,}7}\left(S^{(1)}_{t-1}(y) - S^{(2)}_{t-1}(y)\right). \end{cases}$$

Из системы уравнений получаем $S^{(1)}_{t-1}(y)$=7,25; $S^{(2)}_{t-1}(y)$=7,01.

Для получения текущих значений скользящих средних используется формула (8.16), выведенная Р.Брауном. Для линейной модели значения скользящих средних определяются следующим образом:

$$\begin{cases} S^{(1)}_t(y) = \alpha y_t + (1-\alpha)S^{(1)}_{t-1}(y) \\ S^{(2)}_t(y) = \alpha S^{(1)}_t(y) + (1-\alpha)S^{(2)}_{t-1}(y) \end{cases} \qquad (8.18)$$

Например, для февраля $S^{(1)}_t(y)$ и $S^{(2)}_t(y)$ рассчитываются следующим образом:

$$S^{(1)}_t(y) = 0{,}7 \cdot 4{,}2 + 0{,}3 \cdot 7{,}25 = 5{,}115$$

$$S^{(2)}_t(y) = 0{,}7 \cdot 5{,}115 + 0{,}3 \cdot 7{,}01 = 5{,}684$$

Результаты расчетов скользящих средних представлены в гр. 3 и 4 табл. 8.13. Параметры a_0 и a_1, исходя из (8.17),определяются с использованием уже рассчитанных в гр. 3 и 4 табл. 8.13 значений $S^{(1)}_t(y)$ и $S^{(2)}_t(y)$. Параметры a_0 и a_1 (см. гр. 5 и 6 табл. 8.13) используются в качестве коэффициентов прямой для расчета выравненных уровней. Например, для февраля 4, 546-1, 326 (t=1)= 3,220 и т.д. (см. гр. 7 табл. 8.13).

Таким образом, мы видим, что одним из существенных преимуществ методов, основанных на экспоненциальном сглаживании, является возможность учета временной ценности информации и адаптация к изменяющимся условиям, что имеет большое практическое значение при нестабильном протекании экономических процессов.

Другим преимуществом экспоненциального сглаживания является то, что в отличие от метода скользящей средней при использовании экспоненциального сглаживания существует возможность построения прогнозных оценок уровней динамического ряда. Используя оценки параметров на момент времени t, могут быть получены прогнозы уровней ряда.

Рассчитаем прогнозную оценку выпуска продукции на январь 1995 г. Скользящие средние рассчитываются следующим образом:

$$S^{(1)}_t(y) = 0{,}7 \cdot 11{,}8 + 0{,}3 \cdot 10{,}686 = 11{,}466\,,$$

$$S^{(2)}_t(y) = 0{,}7 \cdot 11{,}466 + 0{,}3 \cdot 10{,}167 = 11{,}076\,.$$

Соответственно, значения параметров тренда a_0 = 11,856 и a_1 = 0,909. Подставляя значения параметров в уравнение прямой, получим прогнозное значение выпуска продукции $\hat{y}_t$=12,765 тыс. т.

Экспоненциальное сглаживание как метод выравнивания лежит в основе более сложных методов адаптивного моделирования. Например, Уинтерсом была предложена модель, учитывающая сезонную составляющую динамического ряда. Эти методы могут быть использованы для оценки тенденций развития различных общественных явлений.

8.5. Понятие сезонной неравномерности и ее характеристика

Слагаясь под совместным воздействием систематических и случайных факторов, уровень ряда динамики испытывает также воздействие причин, обусловленных периодичностью колебаний.

В рядах внутригодичной динамики, можно выделить три важнейшие составляющие колеблемости уровней временного ряда: тренд, сезонную и случайную компоненты. На рис. 8.5 показаны вышеназванные виды колебаний, а на графике 8.5в представлены случайные и сезонные колебания уровней временного ряда при наличии общей тенденции их повышения. При анализе деловой активности на протяжении значительного числа периодов можно выделить еще один вид - циклические колебания, отличающиеся от сезонных отсутствием регулярной модели динамики (см. рис. 8.6). На этом графике можно выделить цикл примерно с начала 1972 г. по 1976 г., когда процентная ставка возросла с 4,5% до 13%, а затем упала до 9,5%. Следующий цикл длился с 1978 по 1984 г. с пиком в 17% в 1980-1988 и 1988-1992 гг. На основе анализа фактических данных можно констатировать существование циклов процентных ставок, однако трудно предсказать как пик, так и время и направление изменений в процентных ставках. Например, наметившаяся тенденция снижения британской базовой процентной ставки в 1974 г. была прервана резким ее ростом в 1977 г. (до 14%), сменившимся падением до 6,5% в 1978 г.

Таким образом, при анализе колеблемости динамических рядов наряду с выделением случайных колебаний возникает и задача изучения периодических колебаний. Как правило, изучение периодических («сезонных») колебаний необходимо с целью исключения их влияния на общую динамику для выявления «чистой» (случайной) колеблемости.

В широком понимании к **сезонным** относят все явления, которые обнаруживают в своем развитии отчетливо выраженную закономерность внутригодичных изменений, т.е. более или менее устойчиво повторяющиеся из года в год колебания уровней. Часто эти колебания могут быть не связаны со сменой времен года. К сезонным явлениям относят, например, потребление электроэнергии; неравномерность производственной деятельности в отраслях пищевой промышленности, связанных с переработкой сельскохозяйственного сырья; перевозки пассажирским транспортом, спрос на многие виды продукции и услуг и т.д.

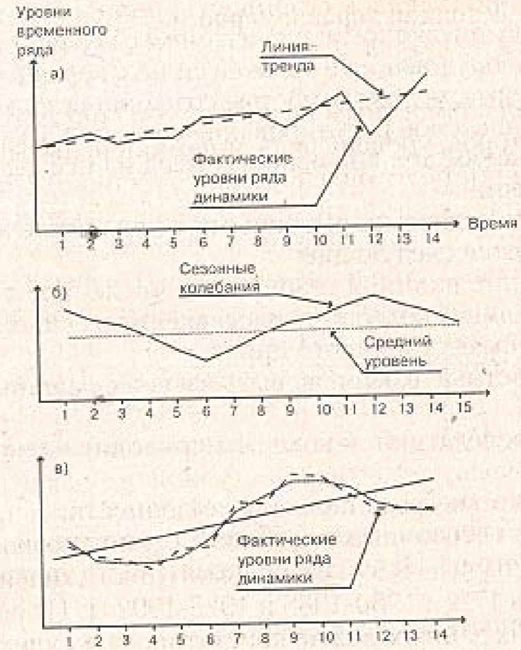

Рис. 8.5. **Виды колебаний уровней временного ряда**

Рис. 8.6. **Британская базовая процентная ставка**

Источник. Райс Т., Койли Б. Финансовые инвестиции и риск. К., Торгово-издательское бюро BHV, 1995. С. 394.

Как бы ни проявлялась сезонность, она наносит большой ущерб национальной экономике, связанный с неравномерным использованием оборудования и рабочей силы, с неравномерной загрузкой транспорта, необходимостью создания резервов мощностей и т.д. Комплексное регулирование сезонных изменений по отдельным отраслям должно основываться на исследовании сезонных отклонений.

Важнейшими задачами, решаемыми в ходе исследования сезонности, являются следующие:

1) определение наличия сезонности, численное выражение проявления сезонных колебаний и выявление их силы и характера в различных фазах годичного цикла;

2) характеристика факторов, вызывающих сезонные колебания;

3) оценка последствий, к которым приводит наличие сезонных колебаний;

4) математическое моделирование сезонности.

Для измерения **сезонных колебаний** статистикой предложены различные методы. Наиболее простые и часто употребляемые из них:

а) метод абсолютных разностей;

б) метод относительных разностей;

в) построение индексов сезонности.

Первые два способа предполагают нахождение разностей фактических уровней и уровней, найденных при выявлении основной тенденции развития.

Применяя способ абсолютных разностей, оперируют непосредственно размерами этих разностей, а при использовании метода относительных разностей определяют отношение абсолютных размеров указанных разностей к выравненному уровню. При выявлении основной тенденции используют либо метод скользящей средней, либо аналитическое выравнивание. В некоторых случаях в стационарных рядах можно пользоваться разностью фактических уровней и средним месячным уровнем за год. Рассмотрим применение методов абсолютных и относительных разностей для измерения сезонных колебаний отпуска электроэнергии с шин по ТЭЦ за трехлетний период. Использование данных за несколько лет связано с тем обстоятельством, что в отклонениях по отдельным годам сезонные колебания смешиваются со случайными. Чтобы элиминировать случайные колебания, берут средние отклонения за несколько лет.

Данные об отпуске электроэнергии с шин за 1993-1995 гг. (млн кВт-ч) представлены в табл. 8.14.

Таблица 8.14

Номер строки	Год	Месяц: январь	февраль	март	апрель	май	июнь	июль	август	сентябрь	октябрь	ноябрь	декабрь	Итого за все месяцы
1	1993	126,5	119,4	114,2	100,0	80,8	89,1	96,1	104,4	114,1	133,0	123,1	145,0	1345,7
2	1994	138,7	135,5	133,8	130,5	104,8	111,8	112,6	134,2	137,9	133,6	131,1	147,3	1551,8
3	1995	133,6	133,4	131,6	119,0	91,9	108,3	169,9	188,2	190,4	194,1	156,5	178,7	1796,6
4	Итого за весь период	398,8	389,3	379,6	349,5	277,5	309,2	378,6	426,8	442,4	460,7	410,7	471,0	4694,1
5	Средний уровень за месяц	132,9	129,8	126,5	116,5	92,5	103,1	126,2	142,3	147,5	153,6	136,9	157,0	130,4
6	Абсолютное отклонение от общей средней	+2,5	-0,6	-3,9	-13,9	-37,9	-27,3	-4,2	+31,9	+17,1	+23,2	+6,5	+26,6	
7	Относительное отклонение от общей ср. (в %)	+1,9	-0,5	-3,0	-10,7	-29,1	-20,9	-3,2	+9,1	+13,1	+17,8	+5,0	+20,4	
8	Индекс сезонности: $\frac{\bar{y}_i}{\bar{y}} \cdot 100\%$	101,9	99,5	97,0	89,3	70,9	79,1	96,8	109,1	113,1	117,8	105,0	120,4	

Для выделения сезонной волны надо определить средний уровень отпуска энергии за каждый месяц по трехлетним данным (см. строку 5 табл. 8.14) и общую среднюю за весь рассматриваемый период. Например, средний уровень отпуска энергии за январь получим делением суммы уровней на число лет, т.е.

$$\bar{y}_{янв.} = \frac{398,8}{3} = 132,9 \text{ млн кВт-ч},$$

$$\bar{y}_{февр.} = \frac{389,3}{3} = 129,8 \text{ млн кВт-ч и т.д.}$$

Общая средняя $\bar{y}_0$ получается делением суммы уровней отпуска за все три года на 36 (общее число месяцев), т.е.

$$\bar{y}_0 = \frac{4694,1}{36} = 130,4 \text{ млн кВт-ч.}$$

Затем определяется абсолютное отклонение средних месячных показателей от общей средней $\bar{y}_i - \bar{y}_0$ (см. строку 6 табл. 8.14). Например, за январь абсолютное отклонение составило 2,5 млн кВт-ч (132,9-130,4). Аналогичные расчеты сделаны для всех остальных месяцев в строке 6.

Метод относительных разностей является развитием метода абсолютных разностей. Для нахождения относительных разностей абсолютные отклонения делят на общую среднюю и выражают в процентах (см. строку 7 табл. 8.14). Например, за январь относительное отклонение от общей средней составило +1,8% ($\frac{+2,5}{130,4}\cdot 100$).

На рис. 8.7 сезонная волна выглядит достаточно отчетливо.

Вместо относительных разностей за каждый месяц может быть вычислен индекс сезонности, который рассчитывается как отношение среднего уровня соответствующего месяца к общей средней, т.е. $y_{сез.} = \frac{\bar{y}_i}{\bar{y}_0}\cdot 100\%$.

Значения индексов сезонности представлены в строке 8 табл. 8.14.

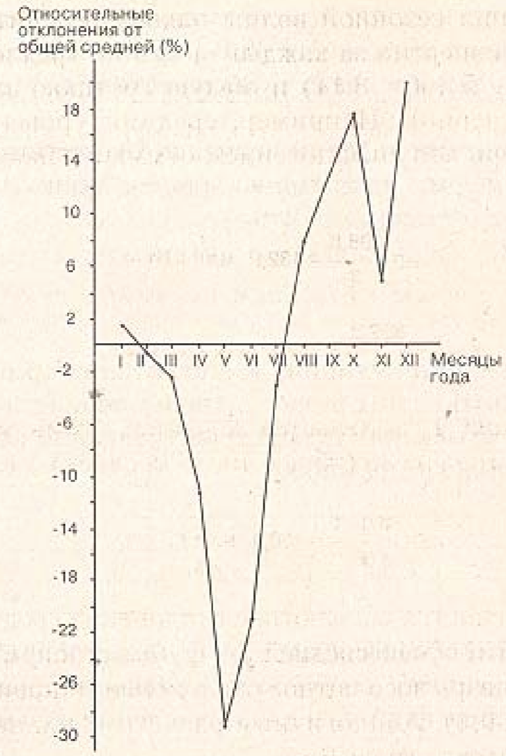

Рис. 8.7. **Относительные отклонения объема производства электроэнергии ТЭЦ по месяцам (в % от общей средней месячной)**

Индексы сезонности могут быть рассчитаны также как отношение фактического уровня соответствующего месяца к уровню, рассчитанному по методу скользящей средней или же определенному по уравнению тренда. (См. пример в табл. 8.18). В последнем случае предварительно для временного ряда рассчитывается уравнение тренда, на основании которого за каждый месяц определяется значение выравненного уровня.

Тогда индекс сезонности рассчитывается следующим образом:

$$y_{сез.} = \frac{y_i}{\hat{y}_i}\cdot 100\% \qquad (8.19)$$

где $\hat{y}_i$ - значение выравненного уровня i-того месяца.

Выделение сезонной волны можно выполнить на основе построения аналитической модели проявления сезонных колебаний.

Построение аналитической модели выявляет основной закон колеблемости данного временного ряда в связи с переходом от месяца к месяцу и дает среднюю характеристику внутригодичных колебаний.

При исследовании явлений периодического типа в качестве аналитической формы развития во времени принимается уравнение следующего типа (ряд Фурье):

$$\hat{y}_t = a_0 + \sum\left(a_k \cos kt + b_k \sin kt\right). \tag{8.20}$$

В этом уравнении величина k определяет гармонику ряда Фурье и может быть взята с разной степенью точности (чаще всего от 1 до 4). Для отыскания параметров уравнения используется метод наименьших квадратов, т.е.

$$\sum_1^n\left(y_i - \hat{y}_t\right)^2 = \min. \tag{8.20а}$$

Найдя частные производные этой функции и приравняв их к нулю, получим систему нормальных уравнений, решение которой дает следующие формулы для вычисления параметров:

$$a_0 = \frac{1}{n}\sum_1^n y_i;$$

$$a_k = \frac{2}{n}\sum_1^n y_i \cos kt_i; \tag{8.20б}$$

$$b_k = \frac{2}{n}\sum_1^n y_i \sin kt_i.$$

Как видим, параметры уравнения зависят от значений y и связанных с ними последовательных значений $\cos kt$ и $\sin kt$.

Для изучения сезонных колебаний на протяжении года необходимо взять n=12 (по числу месяцев в году). Тогда, представляя периоды как части длины окружности, ряд динамики можно записать в следующем виде:

Период	0	$\frac{\pi}{6}$	$\frac{\pi}{3}$	$\frac{\pi}{2}$	$\frac{2\pi}{3}$	$\frac{5\pi}{6}$	π	$\frac{7\pi}{6}$	$\frac{4\pi}{3}$	$\frac{3\pi}{2}$	$\frac{5\pi}{3}$	$\frac{11\pi}{6}$[1]
Уровень	y_0	y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8	y_9	y_{10}	y_{11}

При вычислениях надо иметь в виду, что в четырех квадрантах от 0 до 2 косинусы и синусы четыре раза принимают одни и те же абсолютные значения, а именно: 0; 0,5; 0,866; 1, взятые со знаком плюс или минус. Для вычисления синусов и косинусов разных гармоник лучше всего пользоваться табл. 8.15.

Таблица 8.15

t	$\cos t$	$\cos 2t$	$\cos 3t$	$\cos 4t$	$\sin t$	$\sin 2t$	$\sin 3t$	$\sin 4t$
0	1	1	1	1	0	0	0	0
π/6	0,866	0,5	0	-0,5	0,5	0,866	1	0,866
π/3	0,5	-0,5	-1	-0,5	0,866	0,866	0	-0,866
π/2	0	-1	0	1	1	0	-1	0
2π/3	-0,5	-0,5	1	-0,5	0,866	-0,866	0	0,866
5π/6	-0,866	0,5	0	-0,5	0,5	-0,866	1	-0,866
π	-1	1	-1	1	0	0	0	0
7π/6	-0,866	0,5	0	-0,5	-0,5	0,866	-1	0,866
4π/3	-0,5	-0,5	1	-0,5	-0,866	0,866	0	-0,866
3π/2	0	-1	0	1	-1	0	1	0
5π/3	0,5	-0,5	-1	-0,5	-0,866	-0,866	0	0,866
11π/6	0,866	0,5	0	-0,5	-0,5	-0,866	-1	-0,866

Так как t в годовой динамике означает номер соответствующего месяца, то $t=0$ соответствует январю, $t=\pi/6$ соответствует февралю и т.д. Для нахождения параметров a_k и b_k надо иметь произведение уровней данного месяца на синусы и косинусы соответствующих гармоник. Для $k=1$ уравнение примет вид: $\hat{y}_t=a_0+a_1\cos t+b_1\sin t$, в котором параметры a_0, a_1 и b_1 будут найдены из соотношений:

$$a_0 = \frac{\sum_1^n y_i}{12}; \quad a_1 = \frac{\sum_1^n y_i \cos t_i}{6}; \quad b_1 = \frac{\sum_1^n y_i \sin t_i}{6}.$$

[1] Величины 0, π/6, π/3 и т.д. получены так: $\frac{2\pi}{n}\cdot t$, т.е. при $t=0$ $\frac{2\pi}{12}\cdot 0 = 0$; при $t=1$ $\frac{2\pi}{12}\cdot 1 = \frac{\pi}{6}$ и т.д.

Рассмотрим построение модели сезонной волны на примере данных об условном расходе условного топлива на отпуск теплоэнергии (кг/Гкал), применив первую и вторую гармоники ряда Фурье для построения модели сезонной волны (все необходимые расчеты приведены в табл. 8.16).

Таблица 8.16

Месяцы	t	Удельные расходы условного топлива на отпуск теплоэнергии (кг/Гкал) y	$y \cos t$	$y \sin t$	$\hat{y}_t$	$y \cos 2t$	$y \sin 2t$	$\hat{y}_t$
1	2	3	4	5	6	7	8	9
Январь	0	168,46	168,46	0,00	169,41	168,46	0,00	169,34
Февраль	π/6	169,65	147,00	84,83	169,30	84,83	147,00	169,11
Март	π/3	168,52	84,26	146,00	168,39	-84,26	146,00	168,28
Апрель	π/2	166,12	0,00	166,12	166,92	-166,12	0,00	166,99
Май	2π/3	166,02	-83,01	143,81	165,29	-83,01	-143,81	165,48
Июнь	5π/3	164,23	-142,10	82,11	164,94	82,11	-142,10	164,05
Июль	π	163,04	-163,04	0,00	163,23	163,04	0,00	163,16
Август	7π/6	163,03	-141,10	-81,52	163,34	81,52	141,10	163,15
Сентябрь	4π/3	164,53	-82,27	-112,50	164,25	-82,27	142,50	164,14
Октябрь	3π/2	165,46	0,00	-165,46	165,72	-165,46	0,00	165,79
Ноябрь	5π/3	167,63	83,82	-145,20	167,35	-83,82	-145,20	167,54
Декабрь	11π/6	169,14	146,50	-81,57	168,70	84,57	-146,50	169,81
Итого		1995,83	18,52	3,62	1995,84	-0,41	-1,01	1998,84

Применяя первую гармонику ряда Фурье, получим следующие значения параметров уравнения:

$a_0 = 1995{,}83 : 12 = 166{,}32$; $a_1 = 18{,}52 : 6 = 3{,}09$;
$b_1 = 3{,}622 : 6 = 0{,}60$.

Уравнение модели будет иметь такой вид:

$\hat{y}_t = 166{,}32 + 3{,}09 \cos t + 0{,}60 \sin t$.

В табл. 8.16 (гр. 3 и 6) представлены фактические и расчетные уровни удельного расхода топлива на отпуск теплоэнергии, где наглядно видно, что модель достаточно точно отражает эмпирические уровни ряда динамики.

Применим к этим же данным вторую гармонику ряда Фурье для выражения модели сезонности (см. табл. 8.16, гр. 7-9).

Параметры a_2 и b_2 находим таким образом:

$$a_2 = \frac{\sum_{i=1}^{n} y_i \cos 2t}{6} = -\frac{0{,}41}{6} = -0{,}07,$$

$$b_2 = \frac{\sum_{i=1}^{n} y_i \sin 2t}{6} = -\frac{1{,}01}{6} = -0{,}17.$$

Подставляя полученные коэффициенты в уравнение ряда Фурье, будем иметь следующую модель сезонности данного ряда динамики:

$\hat{y}_t = 166{,}32 + 3{,}09 \cos t + 0{,}60 \sin t - 0{,}07 \cos 2t - 0{,}17 \sin 2t$.

Сопоставление выравненных уровней динамического ряда по первой и второй гармонике (см. табл. 8.16, гр. 6 и 9) приводит к выводу о достаточности использования для выравнивания только первой гармоники. Глубину сезонных колебаний месячных данных измеряют индексами сезонности, которые представляют собой отношение средних из фактических уровней одноименных месяцев за рассматриваемый период к средней из выравненных данных по тем же месяцам, т.е.

$$y_{\text{сез.}} = \frac{\bar{y}_i}{\hat{y}_{i\,\text{ср.}}}, \tag{8.21}$$

где $\bar{y}_i$ - средняя из фактических уровней i-того месяца за весь рассматриваемый период;

$\hat{y}_{i\,\text{ср}}$ - средний из выравненных уровней i-того месяца, полученный либо применением 12-месячной скользящей средней, либо аналитическим выравниваем.

Приведенная формула индекса сезонности показывает, что его величина различна для разных месяцев и зависит от способа выравнивания.

Расчет индексов сезонности проиллюстрируем по данным об удельных расходах условного топлива на отпуск теплоэнергии за 1986-1995 гг. по одной из ТЭЦ. За 10-летний период были рассчитаны средние из фактических уровней удельных расходов топлива за одноименные месяцы, затем производилось выравнивание ряда удельных расходов топлива по прямой и рассчитывались средние из выравненных данных по каждому месяцу (результаты расчетов приведены в табл. 8.17, гр. 2 и 3).

Таблица 8.17

Месяцы	Средний фактический удельный расход за 10-летний период $\bar{y}_i$	Средний удельный расход из выравненных по прямой уровней $\hat{y}_{t\,ср}$	Индекс сезонности, (%) $i_{сез.}$	$(i_{сез.}-100)^2$
1	2	3	4	5
Январь	168,46	166,65	101,1	1,21
Февраль	169,65	166,43	101,9	3,61
Март	168,52	166,40	101,3	1,69
Апрель	166,12	166,38	99,8	0,04
Май	166,02	166,36	99,8	0,04
Июнь	164,23	166,33	98,7	1,69
Июль	163,04	166,31	98,0	4,00
Август	163,03	166,38	98,0	4,00
Сентябрь	164,53	166,26	98,9	1,21
Октябрь	165,46	166,24	99,4	0,86
Ноябрь	167,63	166,21	100,9	0,81
Декабрь	169,14	166,15	101,9	3,61

Сопоставление индексов сезонности по месяцам (см. табл. 8.17, гр. 4) показывает, что минимальный удельный расход топлива имеет место в июле, августе, а максимальный - в декабре, феврале.

Обобщающим показателем силы колеблемости динамического ряда из-за сезонного характера производства служит среднее квадратическое отклонение индексов сезонности (в %) от 100%, т.е.

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

Расчет $\sigma_{сез.}$ основан на результатах, представленных в гр. 5 табл. 8.17, и составляет 1,36 %.

Сравнение средних квадратических отклонений, вычисленных за разные периоды, показывает сдвиги в сезонности. Так уменьшение $\sigma_{сез.}$ свидетельствует об уменьшении влияния сезонности на динамику анализируемого показателя.

Рассмотрим выявление всех типов колебаний внутригодичной динамики уровней временного ряда на следующем примере, характеризующем динамику квартального товарооборота специализированного магазина зарубежной фирмы за 1991-1995 гг.

Таблица 8.18

Год	Квартал	Фактический объем товарооборота, тыс. долл.	Скользящая средняя за четыре квартала, тыс. долл.	Центрированная скользящая средняя, тыс. долл.	Отношение фактического объема товарооборота к скользящей средней (гр.3:гр.5)	Индекс сезонности	Объем товарооборота, скорректированный на индекс сезонности, тыс. долл. (гр.3 : гр.7)
1	2	3	4	5	6	7	8
1991	I	16		·		0,9506	16,8
	II	21	16,00	-		1,2988	16,2
	III	9	15,75	15,875	0,567	0,6119	14,7
	IV	18	15,70	15,625	1,152	1,1387	15,8
1992	I	15	15,75	15,625	0,960	0,9506	15,8
	II	20	15,75	15,750	1,270	0,2988	15,4
	III	10	16,25	16,000	0,625	0,6119	16,3
	IV	18	17,25	16,750	1,075	1,1387	15,8
1993	I	17	18,00	17,625	0,965	0,9506	17,9
	II	24	19,00	18,500	1,297	1,2988	18,5
	III	13	19,00	19,000	0,684	0,6119	21,2
	IV	22	19,25	19,125	1,150	1,1387	19,3
1994	I	17	18,75	19,000	0,895	0,9506	17,9
	II	25	18,50	18,625	1,342	1,2988	19,2
	III	11	18,75	18,625	0,591	0,6119	18,0
	IV	21	19,00	18,875	1,113	1,1387	18,4
1995	I	18	19,75	19,375	0,929	0,9506	18,9
	II	26	20,75	20,250	1,284	1,2988	20,0
	III	14		-		0,6119	22,9
	IV	25		-		1,1387	21,9

Для выравнивания уровней ряда динамики товарооборота принимаем период сглаживания, равный четырем кварталам (m=4). Первая скользящая средняя равна 16,00 тыс. долл. ($\frac{16+21+9+18}{4}$), но отнесена она будет не к конкретному кварталу, а попадет в промежуток между вторым и третьим кварталом года. Точно также к промежутку между кварталами отнесены и все последующие скользящие средние (см. гр. 4 табл. 8.18). Для отнесения скользящей средней к определенному кварталу, находим средние из двух смежных скользящих средних, т.е. производим центрированные средних. Так, для третьего квартала 1991 г. центрированная скользящая средняя будет равна: $\frac{16{,}00+15{,}75}{2}$ =15,875 тыс. долл. Аналогичным образом определяются и все остальные средние, записанные в гр. 5 табл. 8.18.

Для выявления сезонной составляющей в колеблемости уровней ряда динамики рассчитываем отношения фактических объемов товарооборота каждого квартала к соответствующей ему скользящей средней. Эти соотношения представлены в гр. 6 табл. [гр.6=(гр.3:гр.5)]. Для примера находим относительный показатель для III квартала 1991 г.:

$$\frac{\text{Фактический объем товарооборота}}{\text{Центрированная скользящая средняя}} = \frac{9}{15{,}875} = 0{,}567.$$

На основании полученных соотношений выполним их группировку по кварталам.

Таблица 8.19

Соотношения между фактическим объемом товарооборота и скользящей средней

Год	Кварталы			
	I	II	III	IV
1991	-	-	0,567	1,152
1992	0,960	1,270	0,625	1,075
1993	0,965	1,297	0,684	1,150
1994	0,895	1,342	0,591	1,113
1995	0,929	1,284	-	-

Данные таблицы свидетельствуют о наличии различий в величине этих соотношений по кварталам разных лет, хотя порядок их примерно соблюдается. Для расчета индекса сезонности на основании сравнений фактических квартальных значений за ряд лет с соответствующей скользящей средней можно воспользоваться следующими приемами:

1) рассчитать для каждого квартала среднюю арифметическую из полученных соотношений, например, для I квартала среднее значение индекса составит 0,9373($\frac{0{,}960+0{,}965+0{,}895+0{,}929}{4}$);

2) определить медиану из значений индексов сезонности за каждый квартал. Например, ранжировав значения соотношений I квартала, получаем следующий ряд индексов: 0,895; 0,929; 0,960; 0,965.

Тогда медиана будет равна средней величине из второго и третьего значений: $Me = \frac{0{,}929+0{,}960}{2} = 0{,}9445$;

3) можно определить так называемую модифицированную среднюю. Величина этой средней находится путем исключения из рассмотрения по каждому кварталу наименьшего и наибольшего значений и определения средней арифметической из оставшихся значений индексов сезонности. Для расчета модифицированной средней первого квартала исключаем значения 0,895 и 0,965. Тогда $\bar{x}_{модиф.} = \frac{0{,}929+0{,}960}{2} = 0{,}9445$.

Поскольку в рассматриваемом примере для каждого квартала мы располагаем четырьмя значениями, расчет медианы приводит к такому же результату, что и расчет модифицированной средней. Поэтому в таблице 8.20 приведены только средние арифметические значения индексов и медианы.

Таблица 8.20

Расчет индексов сезонности

Кварталы	Средний арифметический индекс сезонности	Медиана	Скорректированное значение медианы
1	2	3	4
I	0,9373	0,9445	0,95056
II	1,2987	1,2905	1,29876
III	0,6168	0,6080	0,61189
IV	1,1138	1,1315	1,13874
Итого	3,9666	3,9745	3,99995 ≈ 4,00
Поправочный коэффициент	1,0084	1,0064	

В итоговой строке таблицы (см. гр. 2 и 3) сумма индексов сезонности хотя и незначительно, но отличается от 4 (для четырех кварталов сумма индексов должна быть равна 4, а их средняя равна 1,00). Для устранения этих расхождений определяется поправочный коэффициент как отношение теоретической суммы индексов (4,0) к фактической величине их суммы (3,9666 или 3,9745). Результаты расчета поправочного коэффициента привдены в последней строке табл. 8.20. Воспользуемся значениями медианы для расчета индексов сезонности, скорректированных на поправочный коэффициент. Так, для первого квартала скорректированное значение медианы составит 0,95056 (0,9445x1,0064). Аналогичным образом рассчитаны значения и для остальных кварталов (см. гр. 4 табл. 8.20). Округлив до четвертого знака после запятой, получим значение индексов сезонности (см.гр. 7 табл. 8.18).

Прежде чем анализировать основную тенденцию (тренд) или циклические колебания, необходимо исключить сезонную компоненту. Без учета сезонности объем квартального товарооборота представлен в гр. 8 табл. 8.18. Эти величины получены делением фактического объема товарооборота на индекс сезонности. Так, для первого квартала 1991 г. получим –16 : 0,9506 = 16,8 тыс. долл., для второго квартала 1991 г. - 21:1,2988=16,2 и т.д. После исключения сезонной составляющей из колеблемости уровней временного ряда, рассчитаем уравнение тренда, воспользовавшись линейной функцией

$$\hat{y}_t = a + bt.$$

В табл. 8.21 представлен расчет параметров линейного уравнения тренда. Напомним, что при условии $\Sigma t_i = 0$, $a = \dfrac{\sum_{i=1}^{n} y_i}{n}$, где n=20, и $b = \dfrac{\sum_{i=1}^{n} y_i t_i}{\sum_{i=1}^{n} t_i^2}$.

Используя итоги граф 3,5 и 6 табл. 8.21, находим: $a = \dfrac{360{,}9}{20} = 18{,}045$; $b = \dfrac{420{,}3}{2660} = 0{,}158$.

Следовательно, уравнение тренда для данного примера имеет следующий вид:

$$\hat{y}_t = 18{,}045 + 0{,}158t.$$

С помощью полученного уравнения тренда выполним экстраполяцию на I и II кварталы 1996 г. Подставив в уравнение t=21, получим $\hat{y}_t$ = 21,36 тыс. долл., при t=23, $\hat{y}_t$ = 21,68 тыс. долл.

Таблица 8.21

Год	Квартал	Объем товарооборота без учета сезонности, тыс. долл. (из гр.8 табл.8.18) y_i	Условные обозначения периодов, t_i	t_i^2	$y_i t_i$	$\hat{y}_i$	$\hat{y}_i$, скорректированный с учетом сезонности, $\hat{y}_i \cdot i_{сез.}$
1	2	3	4	5	6	7	8
1991	I	16,8	-19	361	-319,2	15,04	14,30
	II	16,2	-17	289	-275,4	15,36	19,95
	III	14,7	-15	225	-220,5	15,68	9,59
	IV	15,8	-13	169	-205,4	15,99	18,21
1992	I	15,8	-11	121	-173,8	16,31	15,50
	II	15,4	-9	81	-138,6	16,62	21,59
	III	16,3	-7	49	-114,1	16,94	10,37
	IV	15,8	-5	25	-79,0	17,26	19,65
1993	I	17,9	-3	9	-53,7	17,57	16,70
	II	18,5	-1	1	-18,5	17,89	23,24
	III	21,2	+1	1	+21,2	18,20	11,14
	IV	19,3	+3	9	57,9	18,52	21,09
1994	I	17,9	+5	25	89,5	18,84	17,91
	II	19,2	+7	49	134,4	19,15	24,87
	III	18,0	+9	81	162,0	19,47	11,91
	IV	18,4	+11	121	202,4	19,78	22,52
1995	I	18,9	+13	169	245,7	20,10	19,11
	II	20,0	+15	225	300,0	20,42	26,52
	III	22,9	+17	289	389,3	20,73	12,69
	IV	21,9	+19	361	416,1	21,05	23,97
Итого		360,9		2660	420,3		

Найденные таким образом значения не учитывают сезонные колебания в объеме товарооборота. Для учета сезонной составляющей уровень, полученный в результате экстраполяции, умножают на индекс сезонности, т.е.

$$\hat{y}' = \hat{y}_t \cdot i_{сез.},$$

где $\hat{y}'$ - экстраполируемый уровень с учетом сезонных колебаний.

Тогда для I квартала 1996 г. получим 20,30 тыс. долл. (21,36x0,9506), а для II квартала 1996 г. - 28,16 тыс. долл. (21,68x1,2988).

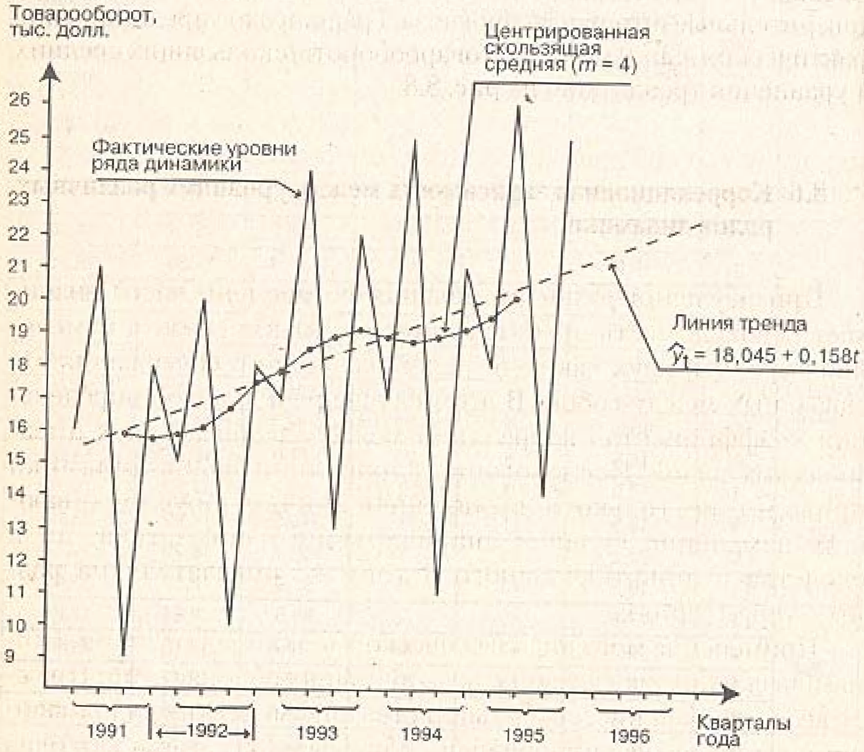

Рис. 8.8. **Динамика товарооборота специализированного магазина зарубежной фирмы за 1991-1995 гг.**

Выравненные по уравнению тренда значения товарооборота, скорректированные с учетом сезонности, представлены в гр. 8 табл. 8.21. При сопоставлении рассчитанных значений с фактическим объемом товарооборота по кварталам, приведенных в гр. 3 табл. 8.18, можно видеть, что величина расхождений между ними невелика (максимальное расхождение составляет 10,62%).

Следует подчеркнуть, что величины товарооборота, полученные в результате экстраполяции, не учитывают наличия отклонений фактических уровней от уровней, рассчитанных по урав-

нению тренда (сравните значения товарооборота в графах 3 и 7 табл. 8.21). Как было показано выше, для прогноза должна быть определена средняя квадратическая ошибка уравнения тренда, с помощью которой для заданного уровня значимости определяют доверительные интервалы прогноза. Графическое представление фактических данных объема товарооборота, скользящих средних и уравнения тренда дано на рис. 8.8.

8.6. Корреляционная зависимость между уровнями различных рядов динамики

При изучении развития явления во времени часто возникает необходимость оценивать степень взаимосвязи в изменениях уровней двух каких-то рядов различного содержания, но связанных между собой. В этом случае речь идет об определении коэффициентов корреляции между уровнями двух динамических рядов. Исследование корреляции рядов динамики приводит не только к выявлению причин, обусловливающих изменение уровней анализируемого показателя, но и сходства в динамике одного и того же показателя на разных территориях.

Применение методов классической теории корреляции в динамических рядах связано с некоторыми особенностями. Прежде всего это наличие для большинства динамических рядов зависимости последующих уровней от предыдущих, тогда как одна из предпосылок применения теории корреляции основана на независимости отдельных наблюдений.

Наличие зависимости между последующими и предшествующими уровнями динамического ряда в статистической литературе называют **автокорреляцией**. Существование или отсутствие автокорреляции всегда надо проверять и одним из условий применения методов корреляционного анализа для изучения взаимосвязей между рядами динамики является ее исключение из каждого из сравниваемых временных рядов.

Другим обстоятельством, которое нужно принимать во внимание при сопоставлении уровней двух динамических рядов, является существование лага, т.е. смещения во времени изменения одного показателя по сравнению с изменением другого. Поэтому при наличии отставания в развитии двух взаимосвязанных показателей нужно сдвинуть уровни одного ряда относительного другого на некоторый промежуток времени, что даст возможность получить более правильную оценку степени тесноты корреляционной связи.

И, наконец, третье обстоятельство связано с тем, что на разных отрезках времени изучаемого периода условия формирования уровня рассматриваемого показателя могли претерпеть существенные изменения. В таком случае мы имеем дело с так называемой **переменной корреляцией**, т.е. с изменяющейся во времени степенью тесноты корреляционной связи.

При оценке тесноты связи между динамическими рядами первостепенное значение приобретает логический анализ связи между показателями, так как в противном случае величина коэффициента корреляции может отражать не реальную зависимость, а чисто случайное совпадение изменений в величине уровней двух временных рядов.

Таким образом, при применении методов корреляции в динамических рядах встает двоякая задача: 1) измерить связь последовательных уровней одного и того же динамического ряда; 2) измерить связь между изменением двух параллельных рядов разного содержания, но так или иначе связанных друг с другом. В первом случае исчисляются коэффициенты автокорреляции и авторегрессии, показывающие зависимость между последовательными уровнями ряда, во втором - коэффициенты корреляции и регрессии.

Коэффициент автокорреляции вычисляется по непосредственным данным рядов динамики, когда фактические уровни одного ряда рассматриваются как значения факторного признака, а уровни этого же ряда со сдвигом на один период принимаются в качестве результативного признака.

Коэффициент корреляции рассчитывается по отклонениям от некоторого выравненного уровня в обоих коррелируемых рядах динамики. Как правило, определяют отклонения фактических уровней от тренда, представляющего основную тенденцию развития каждого ряда.

Коэффициент автокорреляции рассчитывается на основе формулы коэффициента корреляции для парной зависимости

$$r_{y_i \cdot y_{i+1}} = \frac{\sum_{i=1}^{n-1} y_i y_{i+1} - \frac{\sum_{i=1}^{n-1} y_i \sum_{i=1}^{n-1} y_{i+1}}{n-1}}{\sqrt{\left[\sum_{i=1}^{n-1} y_i^2 - \frac{\left(\sum_{i=1}^{n-1} y_i\right)^2}{n-1}\right]\left[\sum_{i=1}^{n-1} y_{i+1}^2 - \frac{\left(\sum_{i=1}^{n-1} y_{i+1}\right)^2}{n-1}\right]}}, \tag{8.22}$$

где n - число уровней первоначального ряда динамики.

Если по результатам расчета коэффициентов автокорреляции будет доказано наличие автокорреляции уровней исходного ряда динамики, то не следует коррелировать непосредственно уровни сравниваемых временных рядов, а предварительно необходимо исключить автокорреляцию уровней в рядах динамики.

Есть несколько способов исключения автокорреляции. Один из них связан с корреляцией отклонений фактических уровней от тренда. Исключение трендов позволяет ослабить автокорреляцию и привести данные в такой вид, который более пригоден для приложения классических методов теории корреляции.

При коррелировании отклонений фактических уровней от выравненных необходимо выполнить следующее:

1) произвести аналитическое выравнивание сопоставляемых рядов динамики;

2) определить величину отклонения каждого фактического уровня ряда динамики от соответствующего ему выравненного значения;

3) рассчитать коэффициент корреляции по полученным отклонениям.

К аналогичному результату в оценке степени тесноты связи рядов динамики мы придем, если будем коррелировать разности между последующим и предыдущим уровнями, т.е. величины

$$\Delta_{y_i} = y_i - y_{i-1}; \quad \Delta_{x_2} = x_i - x_{i-1},$$

где Δ_{y_i} и Δ_{x_i} - абсолютные приросты с переменной базой в рядах динамики показателей y и x.

Коэффициент корреляции первых разностей рассчитывается по такой формуле:

$$r = \frac{\sum_{i=1}^{n-1} \Delta_{x_i} \Delta_{y_i}}{\sqrt{\sum_{i=1}^{n-1} \Delta_{x_i}^2 \cdot \sum_{i=1}^{n-1} \Delta_{y_i}^2}}$$

Но при этом надо иметь в виду, что разность первого порядка исключает автокорреляцию только в тех рядах динамики, в которых изменение во времени происходит по прямой линии.

Контрольные вопросы к главе 8

1. С какой целью анализируются данные рядов динамики?
2. Какие показатели применяются для характеристики изменений уровней ряда динамики?
3. Какой вид средних величин используется для расчета среднего уровня моментного ряда динамики?
4. Охарактеризуйте роль графического представления временных рядов. Назовите наиболее распространенные виды графиков.
5. Как рассчитать средний темп роста и темп прироста уровней ряда динамики?
6. Назовите виды колебаний уровней временного ряда.
7. Как может быть выявлена основная тенденция в изменениях уровней ряда динамики?
8. Назовите преимущества и роль аналитического выравнивания уровней временного ряда.
9. Как выполнить прогноз на будущее с помощью уравнения тренда?
10. Какие факторы влияют на величину средней квадратической ошибки уравнения тренда?
11. Как рассчитать скользящую среднюю и для каких целей она может быть использована?
12. Какие методы можно использовать для выявления сезонных колебаний?
13. Как рассчитать индексы сезонности и осуществить экстраполяцию с учетом сезонной составляющей?
14. Какие особенности корреляции могут быть выделены в рядах динамики?

Глава 9
ИНДЕКСЫ И ИХ ИСПОЛЬЗОВАНИЕ В ЭКОНОМИКО-СТАТИСТИЧЕСКИХ ИССЛЕДОВАНИЯХ

9.1. Общее понятие об индексах и значение индексного метода анализа

В практике статистики индексы наряду со средними величинами являются наиболее распространенными статистическими показателями. С их помощью характеризуется развитие национальной экономики в целом и ее отдельных отраслей, анализируются результаты производственно-хозяйственной деятельности предприятий и организаций, исследуется роль отдельных факторов в формировании важнейших экономических показателей, выявляются резервы производства, индексы используются также в международных сопоставлениях экономических показателей, определении уровня жизни, мониторинге деловой активности в экономике и т.д.

Индекс представляет собой относительную величину, получаемую в результате сопоставления уровней сложных социально-экономических показателей во времени, в пространстве или с планом [1].

Обычно сопоставляемые показатели характеризуют явления, состоящие из разнородных элементов, непосредственное суммирование которых невозможно в силу их несоизмеримости. Например, промышленные предприятия выпускают, как правило, разнообразные виды продукции. Получить общий объем продукции предприятия в таком случае нельзя суммированием количества различных видов продукции в натуральном выражении. Здесь возникает проблема соизмерения разнородных элементов. Как писал К. Маркс в «Капитале», «различные виды становятся коли-

[1] Слово индекс (index) буквально означает указатель, показатель.

чественно сравнимыми лишь после того, как они сведены к одному и тому же единству. Только как выражения одного и того же единства они являются одноименными, а следовательно, соизмеримыми величинами» [1]. В качестве меры соизмерения разнородных продуктов можно использовать цену, себестоимость или трудоемкость единицы продукции.

В развитии индексной теории в нашей стране сложились два направления: обобщающее, или синтетическое, и аналитическое. Различие между этими направлениями обусловлено двумя возможностями интерпретации индексов в их приложении.

Обобщающее, или так называемое синтетическое, направление трактует индекс как показатель среднего изменения уровня изучаемого показателя. В **аналитической теории** индексы - это показатели изменения уровня результативной величины под влиянием изменения индексируемой величины [2].

Например, величина индекса цен продукции, равная 1,1536 может быть истолкована в обобщающем значении как величина, указывающая на то, что цены на продукцию возросли в целом на 15,36% или в 1,1536 раза, а в аналитическом значении как показатель того, что в связи с изменением цен стоимость продукции (или размер выручки) увеличилась на 15,36%. Развитие второго направления было обусловлено применением индексного метода в экономическом анализе. Аналитическое направление в развитии индексной теории в значительной степени сформировалось благодаря работам российский статистиков В.Н. Старовского, Н.М. Виноградовой, Л.В. Некраша, И.Ю. Писарева и др. В настоящее время индексный метод продолжает успешно развиваться в трудах многих отечественных статистиков и получил широкое применение в практике статистической работы.

Таким образом, с помощью индексных показателей решаются следующие основные задачи:

1) характеристика общего изменения сложного экономического показателя (например, затрат на производство продукции, стоимости произведенной продукции и т.д.) или формирующих его отдельных показателей-факторов;

2) выделение в изменении сложного показателя влияния одного из факторов путем элиминирования влияния других факторов (например, увеличение выручки от реализации продукции,

[1] *Маркс К., Энгельс Ф.* Соч. Т. 23. С. 58-59.

[2] Величина, изменение которой изучается в данном конкретном случае с помощью индекса, называется индексируемой величиной.

связанное с ростом цен или выпуска продукции в натуральном выражении). В качестве самостоятельной можно выделить задачу обособления влияния изменения структуры явления на индексируемую величину (например, при изучении динамики среднеотраслевой себестоимости продукции исследуется влияние изменения в распределении объемов выпуска продукции по предприятиям отрасли).

Способы построения индексов зависят от содержания изучаемых показателей, методологии расчета исходных статистических показателей, имеющихся в распоряжении исследователя статистических данных и целей исследования.

Для удобства восприятия индексов в теории статистики разработана определенная символика. Каждая индексируемая величина имеет свое символическое обозначение. Например, количество единиц данного вида продукции обозначается q_i, цена единицы изделия - p_i, себестоимость единицы изделия - z_i, трудоемкость единицы изделия - t_i и т.д.

По степени охвата элементов совокупности различают индивидуальные и сводные (общие) индексы. **Индивидуальными** называются индексы, характеризующие изменение только одного элемента совокупности (например, изменение выпуска легковых автомобилей определенной марки). Индивидуальный индекс обозначается i. **Сводный** индекс отражает изменение по всей совокупности элементов сложного явления. Если индексы охватывают не все элементы сложного явления, а лишь часть, то их называют **групповыми**, или **субиндексами**. Например, общий индекс характеризует динамику объема промышленной продукции. К субиндексам в этом случае могут быть отнесены индексы продукции по отдельным отраслям промышленности. Обозначают сводный (общий) индекс символом I.

Индексные показатели в статистике вычисляются на высшей ступени статистического обобщения и опираются на результаты сводки и обработки данных статистического наблюдения. Итоги по группам элементов в условиях их несоизмеримости получаются расчетным путем, являются производными. Например, объем продукции предприятия может быть представлен в стоимостном или трудовом выражении. В любом из этих случаев показатель объема продукции представляет собой сложный производный показатель, изменение которого синтезирует различный характер изменения отдельных элементов этого показателя и тех факторов, которые его формируют. В зависимости от содержания и характера индексируемой величины различают индексы **количественных** (объемных) показателей (например, индекс физического объема продукции) и индексы **качественных** показателей (например, индексы цен, себестоимости).

При вычислении индексов различают **сравниваемый** уровень и уровень, с которым производится сравнение, называемый **базисным**. Выбор базы сравнения определяется целью исследования. В индексах, характеризующих изменение индексируемой величины во времени, за базисную величину принимают размер показателя в каком-либо периоде, предшествующем отчетному. При этом возможны два способа расчета индексов - цепной и базисный. **Цепные индексы** получают сопоставлением текущих уровней с предшествующим. Таким образом, база сравнения непрерывно меняется. **Базисные индексы** получают сопоставлением с уровнем периода, принятого за базу сравнения.

При территориальных сравнениях за базу принимают данные по какой-либо одной части территории, например, при региональных сопоставлениях внутри России, или итоговый показатель по всей изучаемой территории в целом, как это имеет место в международных сопоставлениях.

При использовании индексов как показателей выполнения плана за базу сравнения принимаются плановые показатели.

В зависимости от методологии расчета различают **агрегатные индексы** и **средние из индивидуальных индексов**. Последние, в свою очередь, делятся на средние арифметические и средние гармонические индексы.

Агрегатные индексы качественных показателей могут быть рассчитаны как индексы **переменного состава** и **индексы фиксированного** (постоянного) **состава**. В индексах переменного состава сопоставляются показатели, рассчитанные на базе изменяющихся структур явлений, а в индексах фиксированного состава - на базе неизменной структуры явлений.

9.2. Индексы количественных показателей

Необходимость в применении особых приемов построения индексов количественных показателей возникает, когда итоги по отдельным элементам сложного явления непосредственно несоизмеримы. Например, предприятие экспортирует станки, металл, товары широкого потребления. Если имеются сведения об экспорте продукции только в натуральном выражении, то динамику экс-

спорта продукции предприятия в целом нельзя охарактеризовать показателем $\frac{\sum_{i=1}^{n} q_i^1}{\sum_{i=1}^{n} q_i^0}$, где q_i^1 - количество продукции данного вида в натуральном выражении, экспортируемой в отчетном периоде; q_i^0 - количество продукции того же вида, отправленной на экспорт в базисном периоде.

Различные виды продукции неравноценны по количеству затраченного на них общественного труда и имеют разные потребительные стоимости. Поэтому было бы неправильно непосредственно суммировать итоги по этим видам продукции. Для получения общего итога необходимо данные по различным видам продукции привести к единой, общей мере, например, использовать стоимостную оценку экспорта продукции. Тогда вместо $\sum_{i=1}^{n} q_i$ получим суммы вида $\sum_{i=1}^{n} p_i q_i$, где p_i - цена единицы продукции данного вида (при расчете экспорта это будет внешнеторговая цена). Такой переход от одних единиц измерения к другим в теории индексов называют **соизмерением**. При построении индексов объемных показателей в качестве соизмерителей применяют те или иные качественные показатели. Например, цену, себестоимость или трудоемкость единицы изделия. Выбор коэффициента соизмерения в каждом конкретном случае зависит от цели исследования. Универсальное значение в индексах физического объема имеют стоимостные соизмерители. Стоимость всей выработанной на предприятии продукции получаем умножением на цену количества выпущенной продукции каждого вида и суммированием произведений по всем видам продукции. Тогда стоимость продукции базисного периода будет определена так:

$$p_1^0 q_1^0 + p_2^0 q_2^0 + p_3^0 q_3^0 + \ldots + p_n^0 q_n^0 = \sum_{i=1}^{n} p_i^0 q_i^0 ,$$

а стоимость продукции отчетного периода составит:

$$p_1^1 q_1^1 + p_2^1 q_2^1 + p_3^1 q_3^1 + \ldots + p_n^1 q_n^1 = \sum_{i=1}^{n} p_i^1 q_i^1 ,$$

где q_i^0, q_i^1 - количество единиц отдельных видов продукции, соответственно в базисном и отчетном периодах; p_i^0, p_i^1 - цена единицы отдельных видов продукции соответственно в базисном и отчетном периодах; $i = 1, 2, 3, ..., n$ - количество отдельных видов продукции.

Если разделить стоимость продукции отчетного периода на стоимость продукции базисного периода, получим **индекс стоимости продукции.** В общем виде его можно записать:

$$I_{pq} = \frac{\sum_{i=1}^{n} p_i^1 q_i^1}{\sum_{i=1}^{n} p_i^0 q_i^0}. \qquad (9.1)$$

Приведенная формула характеризует изменение стоимости продукции, которая зависит от изменения уровня цен и количества выпускаемой продукции в отчетном периоде по сравнению с базисным. Поэтому индекс стоимости не дает количественного представления об изменении объема выпуска. Это представление мы получим, если элиминируем влияние изменения цен, для чего количество продукции, произведенной в отчетном и базисном периодах, умножим на одинаковые для обоих периодов цены:

$$I_q = \frac{\sum_{i=1}^{n} q_i^1 p_i}{\sum_{i=1}^{n} q_i^0 p_i}. \qquad (9.2)$$

Такой индекс называют **агрегатным индексом физического объема.**

При вычислении индекса физического объема продукции возможны разные решения в зависимости от выбора коэффициента соизмерения.

Если принять за коэффициент соизмерения цены базисного периода, то индекс физического объема продукции будет иметь следующий вид:

$$I_q = \frac{\sum_{i=1}^{n} q_i^1 p_i^0}{\sum_{i=1}^{n} q_i^0 p_i^0}. \qquad (9.3)$$

Такой вариант построения агрегатного индекса был предложен Э. Ласпейресом в 1864 г.

Внешней отличительной особенностью агрегатного индекса является то, что в числителе и в знаменателе меняется индексируемая величина, значения же другой, являющейся соизмерителем, остаются неизменными. В приведенном варианте индекса количественных показателей значения соизмерителей принимаются на уровне базисного периода. Используя коэффициенты соизмерения базисного периода, берут базисные соотношения по уровню цен, но зато полностью элиминируют влияние на изменение стоимости продукции изменения самих цен.

В практике планирования при проведении экономико-статистического анализа не ограничиваются исчислением отдельных, изолированных индексов, характеризующих изменение показателя за какой-то один период времени. Исчисляют, как правило, не один индекс, а несколько индексов за последовательные периоды времени. При таком исчислении обычно применяют во всех индексах в качестве соизмерителей цены одного и того же периода. Например, для динамических сопоставлений роста выпуска объема продукции в промышленности, строительстве и т.д. Такие цены называются **сопоставимыми (фиксированными или неизменными)**; в условиях стабильной экономики они применяются на протяжении длительного периода времени. При существенных различиях в соотношении уровней действующих и фиксированных цен производится пересмотр последних, и они меняются время от времени с изменением особенностей самого ценообразования. В настоящее время в странах СНГ, учитывая нестабильное состояние экономики, при расчетах динамики валового внутреннего продукта, национального богатства в качестве фиксированных используют цены предыдущего года.

Применяя в качестве соизмерителя неизменные цены, получим следующую формулу индекса физического объема продукции:

$$I_q = \frac{\sum_{i=1}^{n} q_i^1 p_i^{н}}{\sum_{i=1}^{n} q_i^0 p_i^{н}}, \qquad (9.3а)$$

где $p_i^{н}$ - неизменная оптовая цена единицы изделия.

Преимущество такого варианта соизмерения продукции состоит и в том, что путем суммирования может быть получен итоговый показатель за период любой продолжительности, т.е. на основе данных о стоимости продукции за каждый месяц можно получить стоимость продукции за квартал, полугодие, год. Использование неизменных цен в учете продукции дает возможность изучать динамику выпуска не только отдельных видов продукции, но и по предприятиям, отраслям промышленности и промышленности в целом.

Если поставить задачу характеристики изменения физического объема выпуска продукции, то правомерно в качестве соизмерителей использовать и цены отчетного периода, тогда индекс физического объема продукции будет записан так:

$$I_q = \frac{\sum_{i=1}^{n} q_i^1 p_i^1}{\sum_{i=1}^{n} q_i^0 p_i^1}. \qquad (9.4)$$

Агрегатный индекс с соизмерителями отчетного периода был предложен в 1874 г. Г. Пааше (Paashe).

Рассмотрим порядок вычисления индексов физического объема продукции на следующем условном примере. Имеются данные (табл. 9.1) о фактическом выпуске продукции машиностроительным предприятием за два года:

Таблица 9.1

Объем выпуска продукции предприятия по видам

Виды продукции	Выпуск продукции в натуральном выражении		Цена производителя за единицу, млн руб.		Индивидуальные индексы физического объема продукции	Индивидуальные индексы цен
	базисный период q_i^0	отчетный период q_i^1	базисного периода p_i^0	отчетного периода p_i^1	$i_{q_i} = \frac{q_i^1}{q_i^0}$	$i_{p_i} = \frac{p_i^1}{p_i^0}$
1	2	3	4	5	6	7
Оборудование, шт.	2100	2000	75,00	82,50	0,9524	1,1000
Литье, т	11 500	12 000	8,75	10,10	1,0435	1,1543

По данным о выпуске продукции в натуральном выражении можно рассчитать индексы, характеризующие динамику выпуска

отдельных видов продукции, или индивидуальные индексы. **Индивидуальный индекс** показывает, во сколько раз изменилось производство данного вида продукции в отчетном периоде по отношению к периоду, с которым проводилось сравнение.

Для вычисления индивидуальных индексов динамики определяют отношение объема выпуска продукции отчетного периода к объему выпуска в предшествующем периоде (см. гр. 6 табл. 9.1):

$$i_q = \frac{q_i^1}{q_i^0}.$$

Индекс динамики объема производства оборудования составляет 95,24%, что означает снижение его выпуска на 4,76% ($i_q \cdot 100\% - 100\%$). В динамике же выпуска литья наблюдается противоположная тенденция: выпуск литья возрос на 4,35% ($1{,}0435 \cdot 100\% - 100\%$). Общее изменение выпуска продукции предприятия может быть получено на основе определения агрегатной формы индекса физического объема продукции. Покажем расчет агрегатных индексов физического объема продукции в двух вариантах:

1) в качестве соизмерителей используются цены базисного периода - формула (9.3) и 2) соизмерителями разнородной продукции предприятия являются текущие цены (цены отчетного периода) - формула (9.4). Стоимостные показатели выпуска продукции, необходимые для расчета индексов, приведены в табл. 9.2.

Таблица 9.2

Расчет стоимости выпуска продукции (млн руб.)

Виды продукции	Стоимость выпуска		Условная стоимость выпуска		Доля изделий в стоимости продукции предприятия	
	базисного периода $p_i^0 q_i^0$	отчетного периода $p_i^1 q_i^1$	базисного периода в текущих ценах $p_i^1 q_i^0$	отчетного периода в базисных ценах $p_i^0 q_i^1$	базисного периода d_i^0	отчетного периода d_i^1
1	2	3	4	5	6	7
Оборудование	157 500	165 000	173 250	150 000	0,6102	0,5765
Литье	100 625	121 200	116 150	105 000	0,3898	0,4235
Итого	258 125	286 200	289 400	255 000	1,0000	1,0000

Агрегатный индекс динамики физического объема продукции, рассчитанный по формуле Ласпейреса, составит 0,9879:

[1] Индексы могут быть выражены в долях единицы или в процентах.

$$I_q^{Л} = \frac{\sum_{i=1}^{n} p_i^0 q_i^1}{\sum_{i=1}^{n} p_i^0 q_i^0} = \frac{255\,000}{258\,125} = 0{,}9879, \text{ или } 98{,}79\%,$$

т.е. физический объем выпуска продукции предприятия снизился в отчетном периоде на 1,21%.

$I_q^{Л}$- агрегатный индекс физического объема продукции, рассчитанный по формуле Ласпейреса.

Величина агрегатного индекса физического объема, рассчитанная по формуле Пааше $I_q^{П}$, равна 0,98894:

$$I_q^{П} = \frac{\sum_{i=1}^{n} q_i^1 p_i^1}{\sum_{i=1}^{n} q_i^0 p_i^1} = \frac{286\,200}{289\,400} = 0{,}98894,$$

т.е. физический объем выпуска продукции предприятия уменьшился на 1,106%.

Если сопоставить величины двух индексов $I_q^{Л}$ и $I_q^{П}$, то несмотря на некоторые отличия в величине, они отражают одну и ту же тенденцию - снижение физического объема выпуска продукции предприятия.

Величина агрегатного индекса физического объема зависит от индивидуальных индексов, так как общее изменение объема производимой продукции (при неизменности ассортимента) есть результат изменения объема выпуска каждого отдельного вида. Общий результат изменения определяется также удельным весом стоимости отдельных видов продукции в общей стоимости продукции.

Общий индекс физического объема, построенный на базе индивидуальных индексов, принимает форму среднего арифметического или гармонического индекса. Например, известна стоимость продукции каждого вида в базисном периоде ($p_i^0 q_i^0$) и индивидуальные индексы физического объема $i_q = \frac{q_i^1}{q_i^0}$. Исходной базой построения среднего из индивидуальных индексов служит агрегатная форма индекса Ласпейреса

$$I_q = \frac{\sum_{i=1}^{n} p_i^0 q_i^1}{\sum_{i=1}^{n} p_i^0 q_i^0}$$

Из имеющихся данных непосредственно суммированием можно получить только знаменатель формулы. Числитель же может быть получен перемножением стоимости отдельного вида продукции базисного периода на индивидуальный индекс:

$$q_i^1 p_i^0 = p_i^0 q_i^0 \times i_{q_i} = p_i^0 q_i^0 \times \frac{q_i^1}{q_i^0}.$$

Тогда формула агрегатного индекса физического объема принимает вид:

$$I_q = \frac{\sum_{i=1}^{n} i_{q_i} \cdot p_i^0 q_i^0}{\sum_{i=1}^{n} p_i^0 q_i^0}, \tag{9.5}$$

т.е. получим средний арифметический индекс физического объема, где весами служит стоимость отдельных видов продукции в базисном периоде. При выборе весов следует иметь в виду, что средний индекс должен быть тождествен агрегатному, который является основной формой индекса.

Учитывая, что отношение $\frac{p_i^0 q_i^0}{\sum_{i=1}^{n} p_i^0 q_i^0}$ характеризует долю данного вида продукции в общей стоимости продукции базисного периода (d_i^0), средний арифметический индекс физического объема будет иметь вид:

$$I_q = \sum_{i=1}^{n} i_{q_i} \cdot d_i^0 \tag{9.5а}$$

Воспользуемся данными гр. 6 табл. 9.1 и гр. 2 табл. 9.2 для расчета среднего арифметического индекса физического объема продукции:

$$I_q = \frac{0{,}9524 \cdot 157\,500 + 1{,}0435 \cdot 100\,625}{157\,500 + 100\,625} = 0{,}9879,$$

т.е. получим такой же результат, как и при расчете агрегатного индекса физического объема по формуле Ласпейреса.

Снижение общего объема выпуска продукции предприятия на 1,21% объясняется тем, что превалирующее влияние на величину агрегатного индекса оказывает изменение физического объема выпуска оборудования, поскольку доля оборудования в стоимости продукции предприятия в базисном периоде составляла 61,02% (см.гр. 6 табл. 9.2).

Допустим, что в наличии имеется информация о динамике объема выпуска каждого вида продукции (i_q) и стоимость каждого вида продукции в отчетном периоде ($p_i^1 q_i^1$).

Для определения общего изменения выпуска продукции предприятия в этом случае удобно воспользоваться формулой Пааше, так как числитель формулы можно получить суммированием величин $p_i^1 q_i^1$, а знаменатель - делением фактической стоимости каждого вида продукции на соответствующий индивидуальный индекс физического объема продукции, т.е. делением $\frac{p_i^1 q_i^1}{i_q}\left(\frac{p_i^1 q_i^1}{\frac{q_i^1}{q_i^0}} = p_i^1 q_i^0\right)$.

Тогда

$$I_q^{\Pi} = \frac{\sum_{i=1}^{n} p_i^1 q_i^1}{\sum_{i=1}^{n} p_i^1 q_i^0} = \frac{\sum_{i=1}^{n} p_i^1 q_i^1}{\sum_{i=1}^{n} \frac{p_i^1 q_i^1}{i_{q_i}}} \tag{9.6}$$

Таким образом, в этом случае расчет выполняется по формуле среднего взвешенного гармонического индекса физического объема и величина его будет равна 0,98896 (используем данные гр. 6 табл. 9.1 и гр. 3 табл. 9.2):

$$I_q = \frac{165\,000 + 121\,200}{\frac{165\,000}{0{,}9524} + \frac{121\,200}{1{,}0435}} = 0{,}98896.$$

Можно использовать также данные о стоимости продукции в неизменных (или базисных) ценах (см. гр. 5 табл. 9.2). В этом случае, располагая индивидуальными индексами физического объема продукции и величинами $p_i^0 q_i^1$, расчет общего изменения фи-

зического объема продукции производим по формуле среднего взвешенного гармонического индекса физического объема продукции:

$$I_q = \frac{\sum_{i=1}^{n} p_i^0 q_i^1}{\sum_{i=1}^{n} \frac{p_i^0 q_i^1}{i_{q_i}}} \tag{9.6а}$$

Из сказанного следует, что применение той или иной формулы индекса физического объема (агрегатного, среднего арифметического и среднего гармонического) зависит от имеющейся в нашем распоряжении информации.

Вместе с тем общий индекс физического объема продукции не всегда может быть представлен средней величиной из индивидуальных индексов. Этого нельзя сделать в том случае, когда перечень (номенклатура) изделий в текущем периоде не совпадает с их перечнем в базисном периоде, т.е. средние индексы могут быть рассчитаны лишь по сравнимому кругу изделий. По несравнимой продукции нельзя определить индивидуальные индексы, а потому становится невозможным преобразование агрегатного индекса в адекватные ему средние индексы.

В промышленности наблюдается непрерывное обновление ассортимента выпускаемой продукции, в связи с чем объем выпуска ряда новых видов изделий не может быть сопоставлен ни с одним из предшествующих периодов. Если строго придерживаться формулы агрегатного индекса, то пришлось бы определить индексы физического объема не по всей продукции, а только по тем ее видам, которые вырабатывались на протяжении всех изучаемых периодов времени. Индекс же физического объема продукции должен отразить изменение в общем объеме выпуска, которое происходит как вследствие увеличения (уменьшения) выпуска изделий в отчетном периоде по сравнению с базисным, так и в результате появления новых видов изделий или исключения старых, ранее изготовляемых изделий. Чтобы индекс продукции мог отразить указанные изменения, числитель индекса должен состоять из двух слагаемых: стоимости сравнимой продукции, т.е. продукции, которая изготавливалась и в предшествующие периоды, и стоимости несравнимой продукции, т.е. тех новых изделий, которые ранее не вырабатывались. В знаменателе индекса физического объема продукции приводится стоимость всей продукции базисного периода, включая стоимость и той продукции, которая в отчетном периоде уже не выпускается.

И, наконец, расчет агрегатных индексов может производиться на основе данных о стоимостных (а не натуральных) объемах выпуска каждого вида продукции и индивидуальных индексах цен. В условиях рыночной экономики мониторинг цен имеет первостепенное значение.

Пусть мы располагаем данными о стоимости выпуска продукции в отчетном периоде $p_i^1q_i^1$ и базисном периоде $p_i^0q_i^0$ и индивидуальными индексами цен по отдельным видам продукции $i_{p_i} = \frac{p_i^1}{p_i^0}$. Если воспользоваться формулой агрегатного индекса физического объема Ласпейреса, то следует рассчитать числитель $p_i^0q_i^1$ путем деления стоимости продукции отчетного периода на индекс цен.

Тогда получим следующую формулу:

$$I_q^Л = \frac{\sum_{i=1}^{n}\left(p_i^1q_i^1 : \frac{p_i^1}{p_i^0}\right)}{\sum_{i=1}^{n} p_i^0q_i^0} = \frac{\sum_{i=1}^{n}\frac{p_i^1q_i^1}{i_{p_i}}}{\sum_{i=1}^{n} p_i^0q_i^0}. \qquad (9.7)$$

Если в расчетах динамики выпуска продукции предприятия опираться на индекс Пааше, то следует произвести пересчет знаменателя формулы умножением стоимости продукции базисного периода на индекс цен, т.е. рассчитать величины $p_i^0q_i^0 \times \frac{p_i^1}{p_i^0} = p_i^1q_i^0$. В этом случае формула общего индекса физического объема продукции имеет следующий вид:

$$I_q^П = \frac{\sum_{i=1}^{n} p_i^1q_i^1}{\sum_{i=1}^{n}\left(p_i^0q_i^0 \cdot \frac{p_i^1}{p_i^0}\right)} = \frac{\sum_{i=1}^{n} p_i^1q_i^1}{\sum_{i=1}^{n} p_i^0q_i^0 \cdot i_{p_i}} \qquad (9.8)$$

Воспользуемся данными гр. 7 табл. 9.1 и граф 2 и 3 табл. 9.2 для определения индексов физического объема продукции. На

основе известных индивидуальных индексов цен и стоимостных объемов выпуска продукции

$$I_1 = \frac{\frac{165\,000}{1{,}10} + \frac{121\,200}{1{,}1543}}{157\,500 + 100\,625} = \frac{254\,998{,}7}{258\,125} = 0{,}9879,$$

$$I_2 = \frac{165\,000 + 121\,200}{157\,500 \cdot 1{,}1000 + 100\,625 \cdot 1{,}1543} = \frac{286\,200}{289\,401{,}4} = 0{,}98894.$$

Результаты расчетов аналогичны ранее полученным по формулам агрегатных индексов физического объема Ласпейреса и Пааше.

9.3. Индексы качественных показателей

Наряду с индексами физического объема продукции в планировании и статистико-экономическом анализе деятельности предприятий и отраслей широко применяются индексы качественных показателей: цен, себестоимости, производительности труда, средней заработной платы и т.д. Качественный показатель характеризует уровень изучаемого результативного показателя в расчете на количественную единицу и определяется как отношение данного результативного показателя к связанному с ним количественному показателю (фактору), на единицу которого он определяется. Например, себестоимость единицы продукции определяется как отношение суммы затрат на производство этого вида продукции к количеству единиц продукции данного вида; средняя заработная плата определяется делением фонда заработной платы на численность работников и т.д.

Формулы индексов качественных показателей рассмотрим на примере расчета индексов цен по данным табл. 9.1.

Индивидуальные индексы цен $i_{p_i} = \frac{p_i^1}{p_i^0}$ характеризуют относительное изменение уровня цен единицы каждого вида продукции в отчетном периоде по сравнению с базисным. Приведенные в гр. 7 табл. 9.1 значения индивидуальных индексов цен показывают, что на оборудование цены выросли в 1,10 раза, или на 10%, а цены на литье - в 1,1543 раза, или на 15,43%.

Для определения общего изменения уровня цен на продукцию предприятия, включающую различные виды, нужно рассчитать агрегатный индекс цен. Непосредственное суммирование уровня цен одного станка и одной тонны литья не имеет экономического содержания. Несоизмеримость уровней в таком случае преодолевается путем взвешивания цены каждого вида продукции на количество произведенных единиц, т.е. для отчетного и базисного периода определяются величины вида $\sum_{i=1}^{n} p_i q_i$, которые и сравниваются между собой. Чтобы это сравнение отражало только изменение цен, необходимо, чтобы величина q_i фиксировалась в числителе и знаменателе индекса цен на уровне одного из периодов.

Общая формула агрегатного индекса цен записывается так:

$$I_p = \frac{\sum_{i=1}^{n} p_i^1 q_i}{\sum_{i=1}^{n} p_i^0 q_i}. \qquad (9.9)$$

Очевидно, что как и в случае построения агрегатных индексов физического объема, возможен выбор в качестве веса количества продукции отчетного периода q_i^1 (формула Пааше) или количества продукции базисного периода q_i^0 (формула Ласпейреса).

Формула агрегатного индекса цен Ласпейреса:

$$I_p^{Л} = \frac{\sum_{i=1}^{n} p_i^1 q_i^0}{\sum_{i=1}^{n} p_i^0 q_i^0} \qquad (9.10)$$

Воспользуемся данными табл. 9.1 и итогами гр. 4 и 2 табл. 9.2 для расчета этого индекса:

$$I_p^{Л} = \frac{82{,}5 \cdot 2100 + 10{,}10 \cdot 11\,500}{75{,}0 \cdot 2100 + 8{,}75 \cdot 11\,500} = \frac{289\,400}{258\,125} = 1{,}1212.$$

Полученная величина индекса означает, что цены на продукцию предприятия возросли в отчетном периоде на 12,12%.

Формула агрегатного индекса цен Пааше:

$$I_p^{П} = \frac{\sum_{i=1}^{n} p_i^1 q_i^1}{\sum_{i=1}^{n} p_i^0 q_i^1} \qquad (9.11)$$

Используя данные табл. 9.1 и 9.2, получим величину агрегатного индекса цен Пааше 1,1224:

$$I_p^{П} = \frac{82{,}5 \cdot 2000 + 10{,}10 \cdot 12\,000}{75{,}0 \cdot 2000 + 8{,}75 \cdot 12\,000} = \frac{286\,200}{255\,000} = 1{,}1224.$$

По результатам расчета можно констатировать, что цены на всю продукцию предприятия возросли на 12,24%. Значения агрегатных индексов цен, рассчитанные по формулам Ласпейреса и Пааше, совпадают лишь в случае полного совпадения состава продукции отчетного и базисного периодов. Различия в соотношении этих индексов определяются относительной вариацией индивидуальных индексов цен (v_{i_p}), индивидуальных индексов физического объема (v_{i_q}) и коэффициентом корреляции, оценивающим степень тесноты корреляционной связи между названными индивидуальными индексами ($r_{i_p i_q}$):

$$\frac{I_p^{П}}{I_p^{Л}} = 1 + r_{i_p i_q} \cdot v_{i_p} \cdot v_{i_q}, \qquad (9.12)$$

где $I_p^{П}$ и $I_p^{Л}$ - индексы цен Пааше и Ласпейреса.

v_{i_p} - коэффициент вариации индивидуальных индексов цен:

$$v_{i_p} = \frac{\sigma_{i_p}}{I_p^{Л}}; \quad \sigma_{i_p} = \sqrt{\frac{\sum_{i=1}^{n} \left(i_{p_i} - I_p^{Л}\right)^2 p_i^0 q_i^0}{\sum_{i=1}^{n} p_i^0 q_i^0}},$$

v_{i_q} - коэффициент вариации индивидуальных индексов физического объема продукции:

$$v_{i_q} = \frac{\sigma_{i_q}}{I_q^{Л}}; \quad \sigma_{i_q} = \sqrt{\frac{\sum_{i=1}^{n} \left(i_{q_i} - I_q^{Л}\right)^2 p_i^0 q_i^0}{\sum_{i=1}^{n} p_i^0 q_i^0}}.$$

Степень тесноты связи между индивидуальными индексами цен и физического объема продукции определится так:

$$r_{i_q i_p} = \frac{\sum\limits_{i=1}^{n}\left(i_{p_i} - I_p^{П}\right)\left(i_{q_i} - I_q^{П}\right)p_i^0 q_i^0}{\sum\limits_{i=1}^{n} p_i^0 q_i^0 \cdot \sigma_{i_p} \cdot \sigma_{i_q}}.$$

На практике в целом ряде случаев могут быть известны не абсолютные значения индексируемых показателей, а их относительные изменения. Например, могут быть известны изменения уровней цен или себестоимости по отдельным видам продукции, изменение средней заработной платы по отдельным категориям персонала, изменение рентабельности на отдельных предприятиях отрасли и т.д. В таких случаях агрегатный индекс можно рассчитать косвенным путем, используя взвешенную среднюю из индивидуальных индексов, если нам известен размер результативного показателя за отчетный или базисный период.

Широко применяется средний взвешенный гармонический индекс в статистике торговли при определении индексов розничных цен. Учет товарооборота ведется в денежном выражении по группам товаров, данные же о количестве проданных товаров в натуральном выражении во многих случаях отсутствуют. Поэтому непосредственно определить условную сумму товарооборота $\sum\limits_{i=1}^{n} p_i^0 q_i^1$ невозможно и тогда вместо агрегатной формы индекса вычисляется средний гармонический индекс с текущими весами

$$I_p = \frac{\sum\limits_{i=1}^{n} p_i^1 q_i^1}{\sum\limits_{i=1}^{n} \frac{p_i q_i^1}{i_p}}. \qquad (9.13)$$

Он алгебраически тождествен формуле Пааше и имеет точно такое же экономическое содержание.

Рассмотрим вычисление средних взвешенных индексов качественных показателей на примере. За отчетный месяц цена единицы изделия А возросла на 5% по сравнению с предыдущим меся-

цем, изделия Б - на 3%, изделия В - на 11%. Нужно определить общий (средний) процент роста цен по всем изделиям в отчетном месяце, если известно, что объем товарооборота в отчетном месяце составил (млн руб): по изделию А- 780, по изделию Б - 520, по изделию В - 340. Имеющиеся данные представим в табл. 9.3 (гр. 3 и 4).

Таблица 9.3

Динамика и структура товарооборота магазина

Изделия	Объем товарооборота, млн руб.		Индивидуальные индексы цен	Условный объем товарооборота, млн руб.		Удельный вес стоимости изделия в общем объеме товарооборота	
	предшествующий месяц $p_i^0 q_i^0$	отчетный месяц $p_i^1 q_i^1$	i_{p_i}	отчетного месяца по ценам предыдущего $\frac{p_i^1 q_i^1}{i_{p_i}}$	предыдущего месяца по ценам отчетного $p_i^0 q_i^0 \times i_{p_i}$	предыдущего месяца d_i^0	отчетного месяца d_i^1
1	2	3	4	5	6	7	8
А	750	780	1,05	742,86	787,5	47,02	47,56
Б	530	520	1,03	504,85	545,9	33,23	31,71
В	315	340	1,11	306,31	349,65	19,75	20,73
Итого	1595	1640		1554,02	1683,05	100,00	100,00

Определяем агрегатный индекс цен (по формуле Пааше)

$$I_{p_i}^{П} = \frac{\sum\limits_{i=1}^{n} p_i^1 q_i^1}{\sum\limits_{i=1}^{n} p_i^0 q_i^1} = \frac{\sum\limits_{i=1}^{n} p_i^1 q_i^1}{\sum\limits_{i=1}^{n} \frac{p_i^1 q_i^1}{i_{p_i}}}.$$

Числитель формулы приведен в итоговой строке гр. 3 табл. 9.3 и равен 1640 млн руб. Слагаемые знаменателя можно определить делением товарооборота данного вида продукции в отчетном году на индивидуальный индекс цен:

$$\frac{p_i^1 q_i^1}{i_{p_i}} = p_i^1 q_i^1 : \frac{p_i^1}{p_i^0} = p_i^0 q_i^1.$$

Так, по изделию А $p_i^0 q_i^1 =$ 742,86 млн руб. и т.д. (см. гр. 5 табл. 9.3).

Таким образом, получен общий объем товарооборота по ценам базисного периода. Общая его сумма $\sum_{i=1}^{n} \frac{p_i^1 q_i^1}{i_{p_i}}$ стоит в знаменателе формулы. Разделив итог гр. 3 на итог гр. 5, получим, что в среднем цены возросли на 5,53%

$$I_p^{П} = \frac{1640}{1554{,}02} = 1{,}0553\,.$$

В данном случае агрегатный индекс цен представлен в форме среднего гармонического взвешенного индекса. В качестве весов используются фактические объемы товарооборота в отчетном месяце.

Поставим ту же задачу определения общего изменения цен на все изделия, но при условии , что известен товарооборот предыдущего месяца, т.е. величины $p_i^0 q_i^0$. Тогда при имеющейся информации об индивидуальных индексах цен $i_{p_i} = \frac{p_i^1}{p_i^0}$ и товарообороте предыдущего месяца (данные гр. 4 и 2 табл. 9.3) рассчитать общий индекс цен можно с использованием агрегатного индекса Ласпейреса:

$$I_p^{Л} = \frac{\sum_{i=1}^{n} p_i^1 q_i^0}{\sum_{i=1}^{n} p_i^0 q_i^0} = \frac{\sum_{i=1}^{n} \frac{p_i^1}{p_i^0} \times p_i^0 q_i^0}{\sum_{i=1}^{n} p_i^0 q_i^0} \qquad (9.14)$$

В данном случае агрегатный индекс цен представлен формой среднего арифметического индекса, а в качестве весов используются фактические объемы товарооборота предыдущего месяца.

Учитывая, что $\frac{p_i^0 q_i^0}{\sum_{i=1}^{n} p_i^0 q_i^0}$ представляет собой удельный вес стоимости i-го изделия в общем объеме товарооборота предыдущего месяца, в качестве веса могут использоваться и показатели структуры товарооборота предыдущего месяца (см. гр. 7 табл. 9.3). Используем итоги гр. 6 и 2 табл. 9.3 для расчета агрегатного индекса цен и установим, что в среднем цены возросли на 5,52%:

$$I_p^{Л} = \frac{1683{,}05}{1595} = 1{,}0552\,.$$

Значения индексов $I_q^{П}$ и $I_q^{Л}$ достаточно близки по величине, поскольку в структуре товарооборота в отчетном месяце не произошло значительных изменений, хотя нет и полного совпадения структуры товарооборота текущего и предшествующего месяца (сравните данные по строкам гр. 7 и 8 табл. 9.3). Близость величины агрегатного индекса цен к величине индивидуального индекса цен изделия А объясняется тем, что на долю этого изделия приходится более 47% общего объема товарооборота. «Вклад» отдельных изделий в общий рост цен зависит от структуры товарооборота.

Приведенные варианты исчисления индексов отражают практику отечественной статистики. Во многих странах индексы физического объема и цен также исчисляются аналогичным образом. Вместе с тем в международной статистике для расчетов индексов рекомендуются и другие формы индексов [1].

Если подходить к принципам построения индексов с формально-математических позиций, то ориентируясь на принцип элиминирования влияния других факторов, кроме изучаемого, возможно при исчислении индексов опираться на веса базисного периода (формула Ласпейреса) или же на веса отчетного периода (формула Пааше). Основываясь на этих двух вариантах построения индексов, Фишер предложил рассчитывать среднюю геометрическую из двух агрегатных индексов, назвав ее «идеальной формулой». В табл. 9.4 представлены варианты расчета агрегатных индексов физического объема и цен, наиболее часто используемых в отечественной и зарубежной практике для характеристики временных или пространственных изменений в уровнях анализируемых показателей.

[1] См. например: *Иванов Н.* Обзор аксиоматической теории индексов/ Вопросы статистики, 1995. № 10. С. 25-39.

Таблица 9.4

Название индекса	Агрегатные индексы	
	физического объема	цен
Индекс с базисными «весами» (формула Ласпейреса)	$\frac{\sum_{i=1}^{n} q_i^1 p_i^0}{\sum_{i=1}^{n} q_i^0 p_i^0}$	$\frac{\sum_{i=1}^{n} p_i^1 q_i^0}{\sum_{i=1}^{n} p_i^0 q_i^0}$
Индекс с «весами» отчетного периода (формула Пааше)	$\frac{\sum_{i=1}^{n} q_i^1 p_i^1}{\sum_{i=1}^{n} q_i^0 p_i^1}$	$\frac{\sum_{i=1}^{n} p_i^1 q_i^1}{\sum_{i=1}^{n} p_i^0 q_i^1}$
«Идеальная» формула Фишера	$\sqrt{\frac{\sum_i q_i^1 p_i^0}{\sum_i q_i^0 p_i^0} \times \frac{\sum_i q_i^1 p_i^1}{\sum_i q_i^0 p_i^1}}$	$\sqrt{\frac{\sum_i p_i^1 q_i^0}{\sum_i p_i^0 q_i^0} \times \frac{\sum_i p_i^1 q_i^1}{\sum_i p_i^0 q_i^1}}$

Однако обзор далеко не полного перечня теоретически возможных методов построения агрегатных индексов количественных и качественных показателей свидетельствует о многовариантности возможных расчетов этих показателей. Как подходить к выбору наиболее оптимального варианта, что влияет на выбор того или иного варианта индекса? Основные факторы, влияющие на принятие решения по методу расчета агрегатного индекса, определены целями исследования, возможностями получения статистической информации, а также потребностями удешевления затрат на ее получение.

С этих позиций, если ориентироваться на синтетическое направление в использовании индексов, т.е. поставить задачу характеристики общего изменения уровня анализируемого показателя, предпочтение может быть отдано индексу Ласпейреса. Например, при исчислении агрегатного индекса физического объема продукции в этом случае достаточно вести мониторинг за изменением физических объемов продукции, тогда как при использовании варианта агрегатного индекса Пааше должно учитываться изменение и физического объема продукции и цен. Расчет агрегатного индекса физического объема продукции по формуле Ласпейреса получил наибольшее распространение в мировой практике. Опора на неизменную структуру потребления при расчете агрегатного индекса цен также обусловила применение формулы Ласпейреса при расчете индекса потребительских цен (ИПЦ), величина которого используется при индексации доходов населения (подробнее см. параграф 9.5).

Еще одно преимущество формулы агрегатного индекса Ласпейреса связано с возможностями перехода от ряда индексов с переменной базой сравнения к ряду индексов с постоянной базой сравнения и обратно.

Как уже указывалось, для изучения динамики показателя за ряд периодов возможно вычисление системы цепных и базисных индексов. Расчет такой системы индексов осуществляется в двух вариантах:

1) сравнивают размер показателя в различные периоды с уровнем того же показателя в какой-то определенный период (в этом случае говорят о системе индексов с постоянной базой сравнения - **базисные индексы**);

2) оценивают относительное изменение уровня изучаемого явления по сравнению с предшествующим периодом (получают систему индексов с переменной базой сравнения - **цепные индексы**).

Рассмотрим системы цепных и базисных индексов цен, физического объема и стоимости продукции (см. табл. 9.5). Допустим, имеются данные об объеме продукции в натуральном выражении и уровне цен на эту продукцию за пять периодов (номера периодов обозначим 0, 1, 2, 3, 4).

Для индивидуальных индексов цен, физического объема и индексов стоимости продукции справедливо следующее правило:

1) произведение промежуточных по периодам цепных индексов дает базисный индекс последнего периода, т.е.

$$i_{4/0} = i_{1/0} \cdot i_{2/1} \cdot i_{3/2} \cdot i_{4/3};$$

2) отношение базисного индекса отчетного периода к базисному индексу предшествующего периода дает цепной индекс отчетного периода, т.е.

$$i_{4/3} = i_{4/0} : i_{3/0}.$$

Это правило позволяет применить так называемый цепной метод, т.е. находить неизвестный ряд базисных индексов по известным цепным и обратно.

Рассмотрим возможность применения цепного метода исчисления для агрегатных индексов. Имея два базисных агрегатных индекса физического объема с постоянным соизмерителем (цены базисного периода p^0), можно получить цепной индекс отчетного периода. Действительно,

$$I_{4/3} = \frac{\sum p^0 q^4}{\sum p^0 q^3} = \frac{\sum p^0 q^4}{\sum p^0 q^0} : \frac{\sum p^0 q^3}{\sum p^0 q^0}.$$

Таблица 9.5

Индексы			Базисные				Цепные			
физического объема продукции[1]	индивидуальные		$\frac{q^1}{q^0}$	$\frac{q^2}{q^0}$	$\frac{q^3}{q^0}$	$\frac{q^4}{q^0}$	$\frac{q^1}{q^0}$	$\frac{q^2}{q^1}$	$\frac{q^3}{q^2}$	$\frac{q^4}{q^3}$
	агрегатные	в ценах базисного периода	$\frac{\sum p^0q^1}{\sum p^0q^0}$	$\frac{\sum p^0q^2}{\sum p^0q^0}$	$\frac{\sum p^0q^3}{\sum p^0q^0}$	$\frac{\sum p^0q^4}{\sum p^0q^0}$	$\frac{\sum p^0q^1}{\sum p^0q^0}$	$\frac{\sum p^0q^2}{\sum p^0q^1}$	$\frac{\sum p^0q^3}{\sum p^0q^2}$	$\frac{\sum p^0q^4}{\sum p^0q^3}$
		в ценах отчетного периода	$\frac{\sum p^1q^1}{\sum p^1q^0}$	$\frac{\sum p^2q^2}{\sum p^2q^0}$	$\frac{\sum p^3q^3}{\sum p^3q^0}$	$\frac{\sum p^4q^4}{\sum p^4q^0}$	$\frac{\sum p^1q^1}{\sum p^1q^0}$	$\frac{\sum p^2q^2}{\sum p^2q^1}$	$\frac{\sum p^3q^3}{\sum p^3q^2}$	$\frac{\sum p^4q^4}{\sum p^4q^3}$
цен	индивидуальные		$\frac{p^1}{p^0}$	$\frac{p^2}{p^0}$	$\frac{p^3}{p^0}$	$\frac{p^4}{p^0}$	$\frac{p^1}{p^0}$	$\frac{p^2}{p^1}$	$\frac{p^3}{p^2}$	$\frac{p^4}{p^3}$
	агрегатные	с весами базисного периода	$\frac{\sum p^1q^0}{\sum p^0q^0}$	$\frac{\sum p^2q^0}{\sum p^0q^0}$	$\frac{\sum p^3q^0}{\sum p^0q^0}$	$\frac{\sum p^4q^0}{\sum p^0q^0}$	$\frac{\sum p^1q^0}{\sum p^0q^0}$	$\frac{\sum p^2q^0}{\sum p^1q^0}$	$\frac{\sum p^3q^0}{\sum p^2q^0}$	$\frac{\sum p^4q^0}{\sum p^3q^0}$
		с весами отчетного периода	$\frac{\sum p^1q^1}{\sum p^0q^1}$	$\frac{\sum p^2q^2}{\sum p^0q^2}$	$\frac{\sum p^3q^3}{\sum p^0q^3}$	$\frac{\sum p^4q^4}{\sum p^0q^4}$	$\frac{\sum p^1q^1}{\sum p^0q^1}$	$\frac{\sum p^2q^2}{\sum p^1q^2}$	$\frac{\sum p^3q^3}{\sum p^2q^3}$	$\frac{\sum p^4q^4}{\sum p^3q^4}$

Продолжение табл. 9.5

Индексы		Базисные				Цепные			
стоимости	индивидуальные	$\frac{p^1q^1}{p^0q^0}$	$\frac{p^2q^2}{p^0q^0}$	$\frac{p^3q^3}{p^0q^0}$	$\frac{p^4q^4}{p^0q^0}$	$\frac{p^1q^1}{p^0q^0}$	$\frac{p^2q^2}{p^1q^1}$	$\frac{p^3q^3}{p^2q^2}$	$\frac{p^4q^4}{p^3q^3}$
	агрегатные	$\frac{\sum p^1q^1}{\sum p^0q^0}$	$\frac{\sum p^2q^2}{\sum p^0q^0}$	$\frac{\sum p^3q^3}{\sum p^0q^0}$	$\frac{\sum p^4q^4}{\sum p^0q^0}$	$\frac{\sum p^1q^1}{\sum p^0q^0}$	$\frac{\sum p^2q^2}{\sum p^1q^1}$	$\frac{\sum p^3q^3}{\sum p^2q^2}$	$\frac{\sum p^4q^4}{\sum p^3q^3}$

[1] При исчислении цепных агрегатных индексов физического объема может быть применена и другая система соизмерителей. Например, при исчислении отдельных индексов используются цены периода, предшествующего отчетному. Тогда получают следующий ряд цепных индексов:

$$I_{1/0} = \frac{\sum q^1p^0}{\sum q^0p^0}\ ;\ I_{2/1} = \frac{\sum q^2p^1}{\sum q^1p^1}\ ;\ I_{3/2} = \frac{\sum q^3p^2}{\sum q^2p^2}\ ;\ I_{4/3} = \frac{\sum q^4p^3}{\sum q^3p^3}.$$

То же самое может иметь место и при определении цепных агрегатных индексов цен с весами базисного периода. В качестве веса могут использоваться физические объемы продукции периода, предшествующего отчетному.

Базисный индекс отчетного периода может быть получен перемножением соответствующих цепных индексов, если соизмеритель (цена) принимается на уровне одного и того же периода (например, p^0):

$$I_{3/0} = \frac{\sum q^3 p^0}{\sum q^0 p^0} = \frac{\sum p^0 q^1}{\sum p^0 q^0} \cdot \frac{\sum p^0 q^2}{\sum p^0 q^1} \cdot \frac{\sum p^0 q^3}{\sum p^0 q^2}.$$

Использование цепного метода для агрегатных индексов физического объема оказалось возможным благодаря применению в качестве соизмерителя фиксированных (сопоставимых) цен.

Однако, если воспользоваться другим вариантом построения индекса физического объема (см. примечание к табл. 9.5), нельзя получить точного значения базисного индекса путем перемножения соответствующих цепных индексов, так как при исчислении каждого следующего индекса цены меняются: получается ряд индексов с меняющимся соизмерителем. Действительно,

$$I_{3/0} = \frac{\sum p^0 q^3}{\sum p^0 q^0} \neq \frac{\sum p^0 q^1}{\sum p^0 q^0} \cdot \frac{\sum p^1 q^2}{\sum p^1 q^1} \cdot \frac{\sum p^2 q^3}{\sum p^2 q^2}.$$

Таким образом, при использовании переменных соизмерителей цепной метод для расчета базисных агрегатных индексов физического объема применить нельзя.

Агрегатные индексы качественных показателей, рассчитанные по формуле Пааше, всегда являются индексами с меняющимися весами, так как количество продукции каждый раз принимается на уровне отчетного периода (см. индексы цен с весами отчетного периода в табл. 9.5). Поэтому цепной метод расчета индексов неприменим к таким агрегатным индексам качественных показателей.

Если же воспользоваться формулой Ласпейреса для расчета агрегатных индексов цен при условии постоянных весов q^0 для всех периодов, то базисные индексы могут быть определены на основе цепных и цепные индексы рассчитывают с помощью базисных индексов (см. табл. 9.5 - индексы цен с весами базисного периода q^0).

Сформулированное выше правило взаимосвязи цепных и базисных индивидуальных индексов в полном объеме применимо к агрегатным индексам стоимости.

В случае применения индексов в задачах экономического анализа индексы количественных и качественных показателей не рассматриваются изолированно, а исследуются как факторы, определяющие изменение определенного результативного показателя.

Это важное требование увязки индексов в систему объясняет применяемые на практике правила построения агрегатных индексов количественных и качественных показателей.

Допустим, фирма реализует один продукт. Очевидно, что индекс выручки от реализации продукта равен произведению индекса цен и индекса физического объема продукции:

$$i_{pq} = \frac{p_i^1 q_i^1}{p_i^0 q_i^0} = i_{p_i} \cdot i_{q_i}.$$

Такой же должна быть взаимосвязь и между агрегатными индексами, т.е. $I_{pq} = I_q \cdot I_p$, где

$$I_{pq} = \frac{\sum_{i=1}^{n} p_i^1 q_i^1}{\sum_{i=1}^{n} p_i^0 q_i^0}.$$

Какие же агрегатные индексы физического объема и цен будут удовлетворять этому условию? Обратимся к табл. 9.4, где представлены формулы агрегатных индексов Ласпейреса и Пааше. Сопоставив варианты приведенных в таблице индексов, можно видеть, что равенство индекса стоимости произведению агрегатных индексов цен и физического объема продукции соблюдается в двух вариантах сочетания этих индексов:

1) индекса физического объема Ласпейреса и индекса цен Пааше:

$$I_{pq} = \frac{\sum_{i=1}^{n} p_i^0 q_i^1}{\sum_{i=1}^{n} p_i^0 q_i^0} \times \frac{\sum_{i=1}^{n} p_i^1 q_i^1}{\sum_{i=1}^{n} p_i^0 q_i^1} = \frac{\sum_{i=1}^{n} p_i^1 q_i^1}{\sum_{i=1}^{n} p_i^0 q_i^0},$$

2) индекса физического объема Пааше и индекса цен Ласпейреса:

$$I_{pq} = \frac{\sum_{i=1}^{n} p_i^1 q_i^1}{\sum_{i=1}^{n} p_i^1 q_i^0} \times \frac{\sum_{i=1}^{n} p_i^1 q_i^0}{\sum_{i=1}^{n} p_i^0 q_i^0} = \frac{\sum_{i=1}^{n} p_i^1 q_i^1}{\sum_{i=1}^{n} p_i^0 q_i^0}.$$

И опять возникает проблема выбора одного из указанных вариантов. Здесь на первый план выдвигается экономическое содержание решаемых с помощью индекса задач. Так, в числителе индекса цен Пааше записана фактическая стоимость продукции отчетного периода, которая сравнивается со стоимостью фактического выпуска по базисным ценам, т.е. в случае роста цен речь идет о реальной дополнительной выручке от реализации произведенной продукции, обусловленной увеличением цен производителя. Именно это обстоятельство и обусловливает выбор варианта агрегатного индекса цен Пааше. Тогда, чтобы получить индекс стоимости, агрегатный индекс физического объема должен быть рассчитан по формуле Ласпейреса. Использование индексов в задачах экономического анализа рассмотрено в следующем параграфе.

9.4. Использование индексов в экономическом анализе

Как было показано выше, индексы применяются для характеристики изменения уровня сложных экономических показателей. Их можно использовать также и в аналитических целях для оценки влияния на результативный показатель изменения факторов, его формирующих. Предпосылкой для проведения анализа в индексной форме является возможность представления результативного экономического показателя произведением двух или более определяющих его величину показателей (факторов) или суммой таких произведений. Например, стоимостной объем экспорта может быть представлен произведением уровня внешнеторговых цен p_j на объем экспортных поставок в натуральном выражении q_i[1].

Таким образом, стоимость экспорта зависит от изменения внешнеторговых цен или объема поставок продукции в натуральном выражении, либо от одновременного изменения указанных факторов. Поэтому при анализе динамики или выполнения пла-

[1] p_j и q_j соответственно внешнеторговая цена единицы продукции и объем экспорта продукции в натуральном выражении, поставляемого в j-ую страну.

на по экспорту продукции необходимо показать, в какой мере изменение стоимости экспорта продукции вызвано изменением каждого из этих факторов. С экономической точки зрения небезразлично, какой из этих факторов оказал решающее влияние на увеличение объема экспорта.

Оценивать роль отдельных факторов изменения результативного показателя статистика может путем построения системы взаимосвязанных индексов. Задача состоит в том, чтобы рассчитать изменение сложного показателя при изменении величины только одного фактора так, чтобы величина других факторов была бы сохранена на определенном постоянном уровне. В основе приема аналитических индексных расчетов лежит принцип элиминирования изменений величины всех факторов, кроме изучаемого. При построении индексов, оценивающих влияние отдельных факторов на изменение сложного показателя, необходимо иметь в виду, что общий результат изменения этого показателя представляет собой сумму изменения за счет влияния всех исследуемых факторов, формирующих этот показатель.

Поэтому сформулируем два дополнительных правила, позволяющих обеспечить выполнение этих условий:

1) при расчете индексов количественных показателей соизмерители принимаются на уровне базисного периода, т.е. расчет ведется по формуле Ласпейреса;

2) при расчете индексов качественных показателей веса в числителе и знаменателе фиксируются на уровне, относящемся к текущему периоду, т.е. используется формула Пааше.

Формулы агрегатных индексов позволяют осуществить разложение абсолютного прироста результативного показателя по факторам, например:

$$\Delta pq = \Delta pq(q) + \Delta pq(p), \tag{9.15}$$

где Δpq - абсолютный прирост стоимости продукции;

$\Delta pq(q)$ - абсолютный прирост стоимости продукции, обусловленный изменением физического объема продукции;

$\Delta pq(p)$ - абсолютный прирост стоимости продукции, обусловленный изменением уровня цен на продукцию.

Каждая из названных величин абсолютного прироста рассчитывается как разность числителя и знаменателя соответствующего агрегатного индекса:

$$\underbrace{\sum_{i=1}^{n} p_i^1 q_i^1 - \sum_{i=1}^{n} p_i^0 q_i^0}_{\Delta pq} = \underbrace{\left(\sum_{i=1}^{n} p_i^0 q_i^1 - \sum_{i=1}^{n} p_i^0 q_i^0\right)}_{\Delta pq(q)} + \underbrace{\left(\sum_{i=1}^{n} p_i^1 q_i^1 - \sum_{i=1}^{n} p_i^0 q_i^1\right)}_{\Delta pq(p)}. \qquad (9.15а)$$

Для иллюстрации воспользуемся данными условного примера о поставках одного вида продукции «А» на экспорт (табл. 9.6)

Таблица 9.6

Динамика экспорта продукции «А» предприятия

Страна – импортер	Объем поставок, шт.		Внешнеторговая цена, долл.		Индексы динамики цен	Стоимость экспорта, долл.		Удельный вес поставок в натуральном выражении	
	базисного периода q_j^0	отчетного периода q_j^1	базисного периода p_j^0	отчетного периода p_j^1	i_{p_j}	базисного периода $p_j^0 q_j^0$	отчетного периода $p_j^1 q_j^1$	базисного периода	отчетного периода
1	2	3	4	5	6	7	8	9	10
Болгария	700	450	100	102	1,0200	70 000	45 900	0,7	0,39
Германия	200	250	95	97	1,02105	19 000	24 250	0,2	0,22
Китай	100	450	97	98	1,0103	9700	44 100	0,1	0,39
Итого	1000	1150				98 700	114 250	1,0	1,00

Знак j соответствует индексу страны-импортера.

Можно ли положительно оценить работу отдела маркетинга данного предприятия? Начнем отвечать на поставленный вопрос с определения динамики общего объема поставок. Количество изделий, поставляемых на экспорт, выросло в отчетном периоде на 150 шт. (1150–1000), т.е. на 15%. С этих позиций отделу маркетинга должна быть поставлена положительная оценка.

Далее анализируем динамику внешнеторговых цен со всеми странами-импортерами продукции предприятия (см. гр. 5 и 4 табл. 9.6). Индивидуальные индексы цен показывают их рост в отчетном периоде в торговле со всеми странами (см.гр. 6 табл. 9.6), что также следует оценить положительно.

Выручка предприятия от поставок продукции на экспорт возросла в отчетном периоде по сравнению с базисным на 15 550 долл, или на 15,755% (см. итоги гр. 8 и 7 табл. 9.6).

Однако более детальный анализ динамики стоимости экспорта приводит и к несколько иным выводам. По самым высоким ценам осуществляются поставки продукции в Болгарию, тем не менее в отчетном периоде объем поставок (в натуральном выражении) в Болгарию снизился на 35,71% ($\frac{450}{700}\cdot 100\% - 100\%$). Одновременно на 25% возросли поставки в Германию и на 350% в Китай (индексы физического объема продукции равны соответственно 1,25 и 4,50), где внешнеторговые цены ниже. Поскольку речь идет об изменении распределения объема поставок в натуральном выражении между странами, это не могло не отразиться на изменении среднего уровня внешнеторговой цены. Используя данные о количестве единиц продукта, поставленных в каждую страну (см. гр. 2 и 3 табл. 9.6), и уровне внешнеторговых цен (гр. 4 и 5 табл. 9.6), рассчитаем среднюю цену поставок продукта «А» на экспорт в отчетном и базисном периодах по формуле средней арифметической взвешенной:

$$\bar{p}^{\,0} = \frac{\sum_{j=1}^{3} p_j^0 q_j^0}{\sum_{j=1}^{3} q_j^0} = \frac{98\,700}{1000} = 98{,}7 \text{ долл.}$$

$$\bar{p}^{\,1} = \frac{\sum_{j=1}^{3} p_j^1 q_j^1}{\sum_{j=1}^{3} q_j^1} = \frac{114\,250}{1150} = 99{,}3478 \text{ долл.}$$

Индекс средних цен экспорта продукта «А» определится из соотношения средней цены отчетного и средней цены базисного периодов:

$$I_{\bar{p}} = \frac{\bar{p}^{\,1}}{\bar{p}^{\,0}} = \frac{\sum_{j=1}^{k} p_j^1 q_j^1}{\sum_{j=1}^{k} q_j^1} : \frac{\sum_{j=1}^{k} p_j^0 q_j^0}{\sum_{j=1}^{k} q_j^0} \qquad (9.16)$$

где k - число стран–импортеров продукта.

Поскольку величины $\frac{q_j^1}{\sum_{j=1}^{k} q_j^1}$ и $\frac{q_j^0}{\sum_{j=1}^{k} q_j^0}$ отражают распределение экспорта продукта по странам, т.е. географическую структуру экспорта, формула индекса средних цен экспорта может быть записана так:

$$I_{\bar{p}} = \frac{\sum_{j=1}^{k} p_j^1 d_{q_j}^1}{\sum_{j=1}^{k} p_j^0 d_{q_j}^0} \qquad (9.16а)$$

Величина индекса средних цен экспорта составит 1,00656:

$$I_{\bar{p}} = \frac{99{,}3478}{98{,}7} = 1{,}00656,$$

т.е. прирост средних цен экспорта был 0,656%. Эта величина значительно ниже, чем прирост цен в торговле с отдельными странами: его величина варьирует от 1,03% до 2,105% (см. гр. 6 табл. 9.6). Чем можно объяснить это обстоятельство?

При исчислении средней цены экспорта единицы продукции отчетного периода весами служит количество продукции отчетного периода; при определении средней цены экспорта единицы продукции базисного периода - количество продукции базисного периода. Отношение двух взвешенных средних с меняющимися (переменными) весами, показывающее изменение индексируемой величины, в статистике принято называть **индексом переменного состава**. Величина этого индекса будет зависеть от изменения уровня внешнеторговых цен при поставках продукции в отдельные страны и от изменения в распределении физического объема экспорта между странами.

Чтобы элиминировать влияние изменений в структуре весов на показатель изменения уровня цен, рассчитывают отношение средних взвешенных цен с одними и теми же весами, т.е. исчисляют индекс цен фиксированного состава. Для этого среднюю внешнеторговую цену единицы продукции в базисном периоде корректируют на структуру фактического объема экспорта продукции. Тогда формула индекса средних цен экспорта фиксированного состава будет записана так:

$$I_{\bar{p}} = \frac{\sum_{j=1}^{n} p_j^1 q_j^1}{\sum_{j=1}^{n} q_j^1} : \frac{\sum_{j=1}^{n} p_j^0 q_j^1}{\sum_{j=1}^{n} q_j^1} = \frac{\sum_{j=1}^{n} p_j^1 d_{q_j}^1}{\sum_{j=1}^{n} p_j^0 d_{q_j}^1} \qquad (9.17)$$

Отношение (9.17) показывает, каково было бы изменение среднего уровня внешнеторговых цен по группе стран, в которые данный вид продукции экспортируется, если бы удельный вес экспорта в разные страны в базисном периоде был таким же, как и в отчетном. Но, как указывалось выше, величина взвешенной средней зависит от двух факторов - изменения отдельных уровней цен (величины p_j по отдельным странам) и от изменения в структуре весов. Поэтому, если веса не остаются постоянными, индекс фиксированного состава будет отличаться от индекса переменного состава в меру отношения:

$$\frac{I_{\text{перем.состава}}}{I_{\text{фикс.состава}}} = \left(\frac{\sum_{j=1}^{k} p_j^1 q_j^1}{\sum_{j=1}^{k} q_j^1} : \frac{\sum_{j=1}^{k} p_j^0 q_j^0}{\sum_{j=1}^{k} q_j^0} \right) : \frac{\sum_{j=1}^{k} p_j^1 q_j^1}{\sum_{j=1}^{k} p_j^0 q_j^1}$$

Указанное отношение принято называть **индексом влияния структурных сдвигов**. После несложных преобразований получим:

$$I_{\text{влияния структурных сдвигов}} = \frac{\sum_{j=1}^{k} p_j^0 q_j^1}{\sum_{j=1}^{k} q_j^1} : \frac{\sum_{j=1}^{k} p_j^0 q_j^0}{\sum_{j=1}^{k} q_j^0} = \frac{\sum_{j=1}^{k} p_j^0 d_{q_j}^1}{\sum_{j=1}^{k} p_j^0 d_{q_j}^0} \qquad (9.18)$$

Из приведенной формулы видно, что **индекс влияния структурных сдвигов** представляет собой отношение среднего уровня внешнеторговой цены базисного периода, рассчитанного на структуру экспорта определенного вида продукции отчетного периода, и фактической средней внешнеторговой цены в базисном периоде.

Покажем вычисление индексов цен фиксированного состава и влияния структурных сдвигов на примере (см. данные табл. 9.6).

Агрегатный индекс цен фиксированного состава определится следующим образом:

$$I_{\bar{p}} = \frac{\sum\limits_{j=1}^{k} p_j^1 q_j^1}{\sum\limits_{j=1}^{k} p_j^0 q_j^1} = \frac{\sum\limits_{j=1}^{k} p_j^1 d_{q_j}^1}{\sum\limits_{j=1}^{k} p_j^0 d_{q_j}^1},$$

т.е. расчет индекса может производиться при выборе в качестве веса как объема поставок продукции в натуральном выражении (q_j), так и доли поставок в каждую страну (d_j).

В первом случае величина агрегатного индекса цен рассчитывается следующим образом:

$$1)\ I_{\bar{p}} = \frac{\sum\limits_{j=1}^{k} p_j^1 q_j^1}{\sum\limits_{j=1}^{k} p_j^0 q_j^1} = \frac{102 \cdot 450 + 97 \cdot 250 + 98 \cdot 450}{100 \cdot 450 + 95 \cdot 250 + 97 \cdot 450} = \frac{114\,250}{112\,400} = 1{,}01646.$$

При использовании в качестве веса показателей структуры экспорта в отчетном периоде получим то же значение агрегатного индекса:

$$2)\ I_{\bar{p}} = \frac{\sum\limits_{j=1}^{k} p_j^1 d_{q_j}^1}{\sum\limits_{j=1}^{k} p_j^0 d_{q_j}^1} = \frac{102 \cdot 0{,}39 + 97 \cdot 0{,}22 + 98 \cdot 0{,}39}{100 \cdot 0{,}39 + 95 \cdot 0{,}22 + 97 \cdot 0{,}39} = \frac{99{,}34}{97{,}73} = 1{,}01646.$$

Следует обратить внимание на то, что разность числителя и знаменателя двух вариантов индекса цен фиксированного состава будет характеризовать изменение разных показателей. В первом случае эта разность отражает влияние цен на изменение стоимости экспорта, а во втором – влияние этого же фактора на изменение среднего уровня внешнеторговых цен экспорта.

Агрегатный индекс цен фиксированного состава будет равен 101,646%, а это означает, что в среднем по всем странам внешнеторговые цены возросли на 1,646%. Величина этого индекса получилась выше, чем индекса переменного состава, что свидетельствует о неблагоприятных структурных сдвигах в экспорте продукции.

Величина индекса влияния структурных сдвигов будет равна 99,02%:

$$I_{\text{влияния стр.сдвигов}} = \frac{100 \cdot 0{,}39 + 95 \cdot 0{,}22 + 97 \cdot 0{,}39}{100 \cdot 0{,}70 + 95 \cdot 0{,}20 + 97 \cdot 0{,}10} = \frac{39 + 20{,}9 + 37{,}83}{70 + 19 + 9{,}7} =$$

$$= \frac{97{,}73}{98{,}70} = 0{,}9902\ (99{,}02\%)$$

Следовательно, за счет изменения в распределении объемов экспорта между странами средняя внешнеторговая цена в отчетном периоде снизилась на 0,98%. В абсолютном выражении - на 0,97 долл. Какое влияние оказывает перераспределение физического объема экспорта между странами на величину экспортной выручки? Общая стоимость экспорта $\sum\limits_{j=1}^{k} p_j q_j$ может быть представлена в данном случае величиной $\bar{p} \sum\limits_{j=1}^{k} q_j$, где $\bar{p}$ - средняя внешнеторговая цена единицы продукции.

Общее изменение выручки от экспорта продукта будет равно:

$$\Delta pq = \bar{p}^1 \sum_{j=1}^{k} q_j^1 - \bar{p}^0 \sum_{j=1}^{k} q_j^0. \tag{9.19}$$

Абсолютный прирост выручки, обусловленный изменением средних цен экспорта, составит:

$$\Delta pq(\bar{p}) = (\bar{p}^1 - \bar{p}^0) \sum_{j=1}^{k} q_j^1 = (99{,}3478 - 98{,}7) \cdot 1150 = 745 \text{ долл.}$$

Величина $\sum\limits_{j=1}^{k} q_j$ выступает как вес для качественного показателя – средней цены, а потому фиксируется на уровне отчетного периода. Абсолютный прирост выручки, связанный с изменением физического объема поставок, будет равен

$$\Delta pq\left(\sum_{j=1}^{k} q_j\right) = \left(\sum_{j=1}^{k} q_j^1 - \sum q_j^0\right) \bar{p}^0 = (1150 - 1000) 98{,}7 = 14\,805 \text{ долл.}$$

Поскольку в данном случае оценивается влияние изменения физического объема, соизмеритель фиксируется на уровне базисного периода.

Общий абсолютный прирост экспортной выручки составит 15 550 долл. (745+14805). (Сравните с результатом непосредственного расчета по итогам гр. 8 и 7 табл. 9.6.)

Учитывая, что средняя внешнеторговая цена зависит от уровня цен экспорта в торговле с каждой страной и от распределения общего физического объема между странами, величину Δpq можно записать так:

$$\Delta pq = \sum_{j=1}^{k} p_j^1 d_j^1 \cdot \sum_{j=1}^{k} q_j^1 - \sum_{j=1}^{k} p_j^0 d_j^0 \sum_{j=1}^{k} q_j^0 . \qquad (9.20)$$

Такая запись выявляет влияние уже трех факторов на изменение экспортной выручки: изменения внешнеторговых цен (p_j), географической структуры экспорта (d_j) и объема поставок (q_j).

Общее изменение экспортной выручки можно представить в виде суммы трех слагаемых:

$$\Delta pq = \Delta pq\left(p_j\right) + \Delta pq\left(d_j\right) + \Delta pq\left(q_j\right). \qquad (9.20а)$$

В случае трехфакторной модели при оценке влияния одного из факторов изменения анализируемого показателя два других фактора фиксируются на уровне одного и того же периода. Выбор периода, на уровне которого фиксируются значения факторов (их влияние элиминируется), подчиняется правилу, сформулированному на с. 367. При этом каждый раз необходимо выяснить, каким показателем (количественным или качественным) по отношению к фактору, влияние которого оценивается, являются все остальные. Так, например, при оценке влияния изменения уровня внешнеторговых цен на изменение экспортной выручки распределение физического объема экспорта по странам (d_{q_j}) и объем поставок продукта в натуральном выражении (q_j) являются количественными факторами по отношению к уровню внешнеторговой цены, а потому их величина фиксируется на уровне отчетного периода. Тогда

$$\Delta pq\left(p_j\right) = \left(\sum_{j=1}^{k} p_j^1 d_j^1 - \sum_{j=1}^{k} p_j^0 d_j^1\right) \sum_{j=1}^{k} q_j^1 = (99{,}3478 - 97{,}73)\,1150 = 1860{,}5 \text{ долл.}$$

Полученная величина показывает, что за счет увеличения внешнеторговых цен экспорта продукции в отдельные страны, экспортная выручка в отчетном периоде возросла на 1860,5 долл.

При построении **многофакторных моделей** целесообразно соблюдать определенную последовательность в записи факторов: в основу должно быть положено экономическое содержание произведений двух смежных факторов. В результате может быть осуществлено преобразование (свертывание) сложной многофакторной модели в модель, содержащую меньшее число факторов, что облегчает обоснование выбора весов или соизмерителей соответствующих показателей. Так, в приводимой модели сумма произ-

ведений внешнеторговой цены на долю поставок в j-ую страну $\sum_{j=1}^{k} p_j q_j$ характеризуют среднюю экспортную цену соответствующего периода. В этой двухфакторной модели p_j является качественным показателем, а d_j - количественным. Следовательно, при определении влияния структуры физического объема товарооборота на изменение экспортной выручки p_j будет фиксироваться на уровне базисного периода. Величина физического объема товарооборота по отношению к средней цене экспорта является количественным фактором, а потому будет фиксироваться на уровне отчетного периода (фактор d_j, влияние которого оценивается, входит в величину средней цены экспорта).

Таким образом,

$$\Delta pq\left(d_j\right) = \left(\sum_{j=1}^{k} p_j^0 d_j^1 - \sum_{j=1}^{k} p_j^0 d_j^0\right) \sum_{j=1}^{k} q_j^1 = (97{,}73 - 98{,}70)\,1150 = -1115{,}5 \text{ долл.}$$

Полученная величина означает, что неблагоприятные изменения в географической структуре физического объема экспорта привели к снижению экспортной выручки на 1115,5 долл.

И, наконец, изменение третьего фактора – физического объема экспорта – привело к увеличению экспортной выручки на 14805 долл.:

$$\Delta pq\left(q_j\right) = \sum_{j=1}^{k} p_j^0 d_j^0 \left(\sum_{j=1}^{k} q_j^1 - \sum_{j=1}^{k} q_j^0\right) = 98{,}7\,(1150 - 1000) = 14\,805 \text{ долл.}$$

Средняя внешнеторговая цена по отношению к физическому объему экспорта является качественным фактором, а потому фиксируется на уровне базисного периода. Общее изменение стоимости экспорта составит:

$$\Delta pq = 1860{,}5 - 1115{,}5 + 14\,805 = 15\,550 \text{ долл.},$$

т.е. общая величина экспортной выручки под влиянием всех факторов возросла на 15 550 долл. Однако выполненное разложение абсолютного прироста выручки от экспорта показало, что упущенная выгода в связи с изменением географической структуры экспорта составила 1115,5 долл., а основным фактором, повлиявшим на изменение стоимостного объема экспорта, было увеличение его физического объема.

Рассмотрим по данным табл. 9.7 разложение темпа прироста стоимости экспорта по определяющим его факторам (расчеты выполнены по данным, приведенным в табл. 9.6).

Таблица 9.7

Динамика показателей экспорта продукции

Показатель	Условное обозначение	Темп роста, %	Темп прироста, %
Физический объем экспорта	Q	115,000	15, 000
Средняя внешнеторговая цена	$\overline{p}$	100,656	0,656
Стоимость экспорта	V	115,755	15,755

Основываясь на приведенных в табл. 9.7 данных, определим, на сколько процентов возрастает стоимость экспорта за счет увеличения физического объема экспорта и на сколько - за счет роста уровня средних цен экспорта. Для этого преобразуем абсолютный прирост стоимости экспорта под влиянием изменений каждого из факторов, опираясь на двухфакторную модель взаимосвязи названных выше показателей:

$$V = Q \cdot \overline{p}\ .$$

Тогда

$$\Delta V = V^1 - V^0 = Q^1\overline{p}^1 - Q^0\overline{p}^0 = Q^0\overline{p}^0\left(I_Q \cdot I_p - 1\right) = V^0\left(I_V - 1\right) \qquad (9.21)$$

$$\Delta V_{(Q)} = Q^1\overline{p}^0 - Q^0\overline{p}^0 = \overline{p}^0\left(Q^0 I_Q - Q^0\right) = \overline{p}^0 Q^0\left(I_Q - 1\right) = V^0\left(I_Q - 1\right) \qquad (9.22)$$

$$\begin{aligned}\Delta V\left(\overline{p}\right) &= Q^1\overline{p}^1 - Q^1\overline{p}^0 = Q^1\left(\overline{p}^0 I_p - \overline{p}^0\right) = Q^0 I_Q \overline{p}^0\left(I_p - 1\right) = \\ &= V^0 I_Q\left(I_p - 1\right) = V^0\left(I_V - I_Q\right)\end{aligned} \qquad (9.23)$$

Полученные выражения используем для расчета темпов прироста стоимости экспорта, вспомнив, что темп прироста получают делением абсолютного прироста на уровень базисного периода (см. гл. 8). Темп прироста выручки общий и за счет каждого из указанных факторов определится так:

$$\frac{\Delta V}{V^0}\cdot 100\% = \left(I_V - 1\right)\cdot 100\% = 115{,}755 - 100{,}000 = 15{,}755\%\ .$$

$$\frac{\Delta V_{(Q)}}{V^0}\cdot 100\% = \left(I_Q - 1\right)\cdot 100\% = 115{,}0 - 100{,}0 = 15{,}0\%\ .$$

$$\frac{\Delta V\left(\overline{p}\right)}{V^0}\cdot 100\% = \left(I_V - I_Q\right)\cdot 100\% = 115{,}755 - 115{,}0 = 0{,}755\%\ .$$

Таким образом, прирост стоимости экспорта за счет роста физического объема экспорта на 15% составил 15%, а за счет роста средних цен экспорта на 0,656% стоимость экспорта возросла на 0,755%.

При однонаправленном изменении факторов, формирующих результативный показатель, может быть рассмотрена структура темпа прироста результативного показателя, т.е. определена доля каждого фактора в общем приросте:

$$\frac{\Delta V(Q)}{\Delta V} \quad \text{и} \quad \frac{\Delta V\left(\overline{p}\right)}{\Delta V}\ .$$

Используя выражения (9.21) - (9.23), определим относительный размер влияния каждого фактора:

$$\frac{\Delta V(Q)}{\Delta V} = \frac{V^0\left(I_Q - 1\right)}{V^0\left(I_V - 1\right)} = \frac{I_Q - 1}{I_V - 1}, \qquad (9.24)$$

$$\frac{\Delta V\left(\overline{p}\right)}{\Delta V} = \frac{V^0\left(I_V - I_Q\right)}{V^0\left(I_V - 1\right)} = \frac{I_V - I_Q}{I_V - 1}. \qquad (9.25)$$

По данным табл. 9.7 подсчитаем, что благодаря изменению физического объема экспорта получено 95,21% всего прироста стоимости экспорта [$\frac{1{,}15-1}{1{,}15755-1}\cdot 100\%$].

За счет изменения средних цен экспорта имеем 4,79% общего прироста стоимости экспорта [$\frac{1{,}15755-1{,}15}{1{,}15755-1{,}00}\cdot 100\%$].

Результаты расчетов свидетельствуют о том, что преобладающее влияние на увеличение выручки от экспорта продукции оказал рост экспорта в натуральном выражении, а на долю роста цен приходится 4,79% общего прироста экспортной выручки.

9.5. Использование индексов в макроэкономических исследованиях

Важным направлением статистических исследований является **сопоставление макроэкономических показателей** различных стран. Проблемы, возникающие при международных сопоставлениях, обусловлены тем, что сравниваемые объекты могут иметь свою структуру показателей и свою систему соизмерителей.

Так, при сопоставлении уровней промышленного производства двух стран А и В могут быть рассчитаны два индекса физического объема: один – с использованием соизмерителей страны А, второй – с соизмерителями страны В. Два названных индекса физического объема будут выглядеть следующим образом:

1) при структуре цен страны А

$$I_q = \frac{\sum_i q_{iB} p_{iA}}{\sum_i q_{iA} p_{iA}} = \frac{\sum_i \frac{q_{iB}}{q_{iA}}\left(q_{iA} p_{iA}\right)}{\sum_i q_{iA} p_{iA}} \qquad (9.26)$$

2) при структуре цен страны В

$$I_q = \frac{\sum_i q_{iB} p_{iB}}{\sum_i q_{iA} p_{iB}} = \frac{\sum_i q_{iB} p_{iB}}{\sum_i \frac{q_{iA}}{q_{iB}}\left(q_{iB} p_{iB}\right)} \qquad (9.27)$$

где q_{iA} и q_{iB} - количество i-го продукта соответственно в стране A и в стране B.

p_{iA} и p_{iB} - цена i-го продукта соответственно в стране A и стране B.

Таким образом может быть два результата, как правило, заметно отличающихся один от другого. Поэтому для получения единого вывода предлагается в качестве возможного варианта использовать среднюю геометрическую из двух территориальных индексов физического объема (т.е. предлагается использовать «идеальную» формулу индекса Фишера (см. табл. 9.4).

Другим вариантом является применение стандартизованной структуры показателя. Преобразуем агрегатный индекс физического объема продукции страны A к продукции страны B следующим образом:

$$I_{q(A/B)} = \frac{\sum_{i=1}^{M} q_{iA} p_i}{\sum_{i=1}^{M} q_{iB} p_i} = \frac{\sum_{i=1}^{M}\left(\frac{q_{iA}}{Q_i}\right) p_i Q_i}{\sum_{i=1}^{M}\left(\frac{q_{iB}}{Q_i}\right) p_i Q_i} = \frac{\sum_{i=1}^{M} d_{iA} D_i}{\sum_{i=1}^{M} d_{iB} D_i},$$

где d_{iA} и d_{iB} - удельные веса стран A и B в объеме продукта i;

$D_i = \frac{p_i Q_i}{\sum_{i=1}^{M} p_i Q_i}$ - общая для стран A и B структура показателя, рассчитанная на основе выбранных соизмерителей p и значений физического объема продукта Q ($\sum_{i=1}^{M} D_i = 1$).

Величины $\sum_{i=1}^{M} d_{iA} D_i$ и $\sum_{i=1}^{M} d_{iB} D_i$ представляют собой средние взвешенные удельные веса стран A и B в производстве продуктов, входящих в сопоставляемый показатель.

Величины D_i можно определить как на основе цен страны A (D_{iA}), так и цен страны B (D_{iB}), т.е. можно рассчитать величины:

$$D_{iA} = \frac{p_{iA} Q_i}{\sum_{i=1}^{M} p_{iA} Q_i} \quad \text{и} \quad D_{iB} = \frac{p_{iB} Q_i}{\sum_{i=1}^{M} p_{iB} Q_i}.$$

В качестве стандартизованной структуры может быть принята средняя арифметическая из показателей D_{iA} и D_{iB}, т.е. величина $\overline{D}_i = \frac{D_{iA} + D_{iB}}{2}$. Тогда индекс физического объема будет представлен так:

$$I_{q(A/B)} = \frac{\sum_{i=1}^{M} d_{iA} \overline{D}_i}{\sum_{i=1}^{M} d_{iB} \overline{D}_i} \qquad (9.28)$$

Традиционным направлением использования индексного метода в статистике развитых стран является анализ **состояния рынка ценных бумаг**. Индикаторы рынка ценных бумаг рассчитываются по их видам: акции, облигации, опционы и др.

По степени обобщения исследуемой информации можно выделить следующие показатели рынка ценных бумаг:

1) интегральные, характеризующие состояние исследуемого рынка в целом одним обобщенным показателем. Например, известный сводный индекс Доу-Джонса «Индекс-65», рассчитывающийся по акциям 30 крупнейших промышленных корпораций, 20 транспортных и 15 коммунальных;

2) частные, которые дополняют интегральный показатель характеристикой отдельных составных частей (элементов) исследуемого рынка. Например, в дополнение к индексу Доу-Джонса могут быть даны характеристики изменения курса акций отдельных промышленных компаний.

По составу изучаемых объектов интегральные индексы могут охватывать как весь мировой рынок акций, так и его географические секторы и рынок акций отдельных государств. Таковы индексы MSCI (Morgan Stanly Capital International) - например Europe 13, North America, World Index.

Секторные интегральные индексы характеризуют состояние внутринационального рынка. Например, индекс Нью-йоркской фондовой биржи характеризует «движение» акций всех компаний, котируемых на этой бирже.

Субсекторные интегральные индексы являются составной частью секторного индекса и характеризуют динамику акций, например промышленных, транспортных или финансовых компаний США.

Таким образом, интегральные индексы могут характеризовать различные сегменты глобального рынка ценных бумаг.

Индексы рынка ценных бумаг - это индексы цен акций, обращающихся на рынке и определяющие динамику их изменения на этом рынке. Индексы могут рассчитываться ежедневно, еженедельно, ежемесячно, ежеквартально, по полугодиям, ежегодно. Изменения значений индексов рассматривается как показатель спроса на рынке.

Индекс рынка ценных бумаг может использоваться для различных сопоставлений:

1) изменение цен определенных акций можно сравнить с индексом определенного сегмента рынка или с индексом всего рынка и прогнозировать будущее движение цен на акции, а также оценивать спрос на данный вид акций;

2) можно сопоставить изменения цен в различных сегментах рынка и делать выводы о том, какой сектор из них является наиболее прибыльным для инвесторов в данный момент;

3) для сравнения в аналогичных целях цен на акции в разных странах;

4) для сравнения изменения цен акций мелких и крупных компаний.

Биржевые индексы являются ключевыми показателями для прогнозирования общего положения дел на фондовом рынке и в

отдельных отраслях. С помощью индексов ставится задача изучить общие тенденции на рынке, чтобы служить основным руководством действий инвесторов.

На практике используют четыре методических приема для построения интегральных индексов:

1) рассчитывается темп роста (снижения) средней цены акции, определенной по формуле простой средней арифметической;

2) рассчитывается темп роста (снижения) средневзвешенной цены акций (в качестве веса наиболее часто используется количество обращающихся акций);

3) рассчитывается средний арифметический темп прироста (снижения) цены акций;

4) рассчитывается средний геометрический темп прироста (снижения) цены акций.

Рассмотрим эти приемы на условном примере, показывающем расчет индексов цен акций трех компаний. Курс их акций, а также количество выпущенных акций приведены в табл. 9.8.

Таблица 9.8

Наименование компании	Курс акций, долл.		Количество выпущенных акций, млн ед.	Прирост курса акций, %	Рыночная стоимость всех акций, млн долл.	
	базисный период p_i^0	текущий период p_i^1	q_i		базисный период	текущий период
1	2	3	4	5	6	7
Центр	20	22	20	+10	400	440
Омега	40	38	10	-5	400	380
Плюс	100	120	1	+20	100	120
Итого					900	940

Если воспользоваться первым методом расчета индекса цен на акции, то получим следующий результат:

средняя арифметическая (простая) цена акций - в базисном периоде

$$\bar{p}^{0} = \frac{20+40+100}{3} = 53{,}3 \text{ долл.},$$

в отчетном периоде:

$$\bar{p}^{1} = \frac{22+38+120}{3} = 60 \text{ долл.}$$

Индекс цен акций будет равен 1,1257:

$$I_p = \frac{60}{53,3} = 1,1257,$$

т.е. прирост цен на акции компаний в среднем составил 12,57%.

При втором варианте расчета индексов должно учитываться количество выпущенных акций (q_i).

$$I_p = \frac{\sum_{i=1}^{3} p_i^1 q_i}{\sum_{i=1}^{3} p_i^0 q_i} = \frac{940}{900} = 1,0444,$$

т.е. прирост цен составил 4,44%.

По аналогичной методике рассчитывается индекс «Standard and Poors-500», включающий акции 400 промышленных, 20 транспортных, 40 коммунальных и 40 финансовых корпораций [1]. Эта же методология построения индексов используется при расчетах с опционами и фьючерсами. Также считают индекс «Wilshire-5000», включающий акции всех корпораций, котируемых на Нью-йоркской, Американской фондовой биржах и во внебиржевом обороте.

Используя третий метод, рассчитаем средний арифметический темп прироста цен акций (см. гр. 5 табл. 9.8).

$$\overline{T}_{\text{прир.}} = \frac{0,10+(-0,05)+0,20}{3} = 0,0833.$$

Это означает, что средний прирост цен акций составил 8,33%.

И, наконец, с помощью средней геометрической получим следующий результат:

$$\overline{k} = \sqrt[3]{(1+0,10)(1-0,05)(1+0,20)} = 1,0784,$$

а, следовательно, можно сделать вывод о среднем темпе прироста цен акций на 7,84%.

Если отвлечься от числовых различий в величине полученных индексов, то можно видеть, что общую тенденцию расчета курса акций названных компаний в текущем периоде они характеризуют однонаправленно, хотя имеет место довольно значительная вариация в величине индексов. Главное назначение интегральных показателей - выявить основное направление, основную тенденцию (тренд) движения курса акций на рынке для характеристики деловой активности в экономике.

[1] Акции этих корпораций зарегистрированы в основном на Нью-йоркской фондовой бирже.

На условном примере мы показали основные принципы и методические подходы к построению индексов рынка ценных бумаг. Если рассматривать эту проблему с формально-математических позиций, то для сравнения за отдельные периоды можно выбрать один из вариантов рассмотренных средних, с помощью которых и осуществлять мониторинг цен акций. Однако на практике приходится учитывать целый ряд дополнительных обстоятельств, вносящих существенные коррективы в расчет индексов. К такого рода обстоятельствам относится изменение списка компаний, акции которых формируют индекс; дробление акций компаний; выбор весовых коэффициентов и др. Для примера рассмотрим особенности расчета индекса Франкфуртской фондовой биржи - индекса DAX.

В основе построения этого индекса лежит формула Ласпейреса, однако в отличие от общепринятой международной практики в качестве веса выступает размер основного капитала, а не число выпущенных акций. Размер основного капитала каждой компании получают произведением номинальной цены акции на количество выпущенных акций. Для каждой из 30 компаний, включенных в расчет индекса, определяется удельный вес основного капитала в составе портфеля акций. При расчете индекса DAX даже при одинаковых абсолютных изменениях курса акций большее влияние на величину индекса окажет акция той компании, у которой больше доля основного капитала, т.е. выше весовой коэффициент [1].

В условиях перехода к рыночной экономике большое практическое значение приобретает **анализ динамики цен на товары и услуги** [2]. Это обусловлено тем, что показатели, характеризующие инфляционные процессы на рынке потребительских товаров, применяются при решении многих актуальных экономических задач. Для оценки динамики цен на товары используется индекс потребительских цен (ИПЦ), который иногда называют индексом стоимости жизни. Индекс потребительских цен характеризует изменение во времени общего уровня цен на товары и услуги, приобретаемые населением для непроизводственного потребления. ИПЦ ориентирован на решение следующих основных задач:

[1] Подробная формула индекса DAX приведена в журнале: «Вопросы статистики», 1995. № 12. С. 23.

[2] Этот раздел параграфа написан совместно с доц. Саблиной Е.А.

а) оценку инфляции;
б) индексацию доходов;
в) определение текущих издержек производства;
г) регулирование реального курса национальной валюты.

Известны два основных источника информации по ценам. Первым источником является наблюдение за изменением цен и тарифов на потребительском рынке, которое проводится с 1992 г. специальной государственной службой при Госкомстате РФ. Эта служба занимается сбором и обработкой данных о розничной торговле, объеме производства, обследованием покупок, используя данные статистики производства, статистики налогов и торговой статистики. Вторым важным источником информации являются бюджетные обследования как одна из форм выборочного статистического исследования доходов, расходов и потребления населения. На основании этих источников рассчитывается ИПЦ по фиксированному набору основных потребительских товаров и услуг. Методология расчета индекса является единой для многих стран, что позволяет производить международные сопоставления ИПЦ.

При проведении выборочного наблюдения за ценами для определения индекса необходимо учитывать следующие факторы: группы обследуемого населения по уровню доходов и по объему тех благ, на которые население расходует средства; виды торговых точек; состав потребительской корзины, т.е. тех товаров, которые используются для расчета индекса.

Потребительский набор, на основании которого рассчитывается ИПЦ, состоит из трех групп: продовольственные товары, непродовольственные товары и платные услуги, оказываемые населению. Каждая группа представлена конкретными товарами, услугами или малыми товарными подгруппами. Отбор товаров-представителей производится с учетом их относительной важности для потребления населением, устойчивого наличия их в продаже, представительности с точки зрения отражения динамики цен на однородные товары.

С 1992 г. потребительская корзина в России охватывает 409 групп товаров-представителей и платных услуг, в том числе по продовольственным товарам - 103 позиции, непродовольственным товарам - 222 позиции и платным услугам – 84 позиции. Наблюдение за ценами по этим товарам и услугам проводится во всех столицах республик, входящих в состав Российской Федерации, краев, областей, округов и выборочно – в районных центрах (15% районных центров каждой области). Всего в 1994 г. наблюдение проводилось в 834 городах РФ.

Индекс потребительских цен определяется с недельной, месячной, квартальной периодичностью. Расчет за месяц (квартал, период с начала года) производится путем перемножения недельных (месячных, квартальных) индексов потребительских цен.

Оперативный индекс (еженедельный) рассчитывается по 122 позициям товаров и услуг на основе наблюдения за ценами в 132 городах России и характеризует изменение цен по сравнению с предыдущей неделей, с последней неделей предыдущего месяца, а также с последней неделей декабря предыдущего года.

Учитывая различие в структуре потребления населением товаров и услуг в различных регионах Российской Федерации, ИПЦ на региональном уровне разрабатывается на основе структуры потребительских расходов населения данной территории. Принципиальная схема, иллюстрирующая порядок расчета индексов потребительских цен, представлена на рис. 9.1.

Расчет индекса потребительских цен (индекса стоимости жизни) выполняется в несколько этапов. Первоначально определяются индивидуальные индексы по отдельным товарам (услугам) для данного города. На основе индивидуальных индексов цен и территориальных весов определяются агрегатные индексы цен отдельных товаров, товарных групп и услуг в целом по региону, экономическому району, Российской Федерации. В качестве территориального веса используется численность населения этой территории или удельный вес численности населения обследуемой территории в общей численности населения РФ, т.е.

$$I_p = \frac{\sum_{j=1}^{k} i_{p_{ij}} N_j}{\sum_{j=1}^{k} N_j} = \sum_{j=1}^{k} i_{p_{ij}} d_{N_j} \qquad (9.29)$$

где $i_{p_{ij}}$ - индивидуальный индекс цен i-того товара, зарегистрированный на j-той территории;
N_j - численность населения i-той территории;
k - число территориальных единиц.

Используя агрегатные индексы по товарам и услугам по региону, Российской Федерации и привлекая данные об удельном весе расходов на их приобретение в потребительских расходах на-

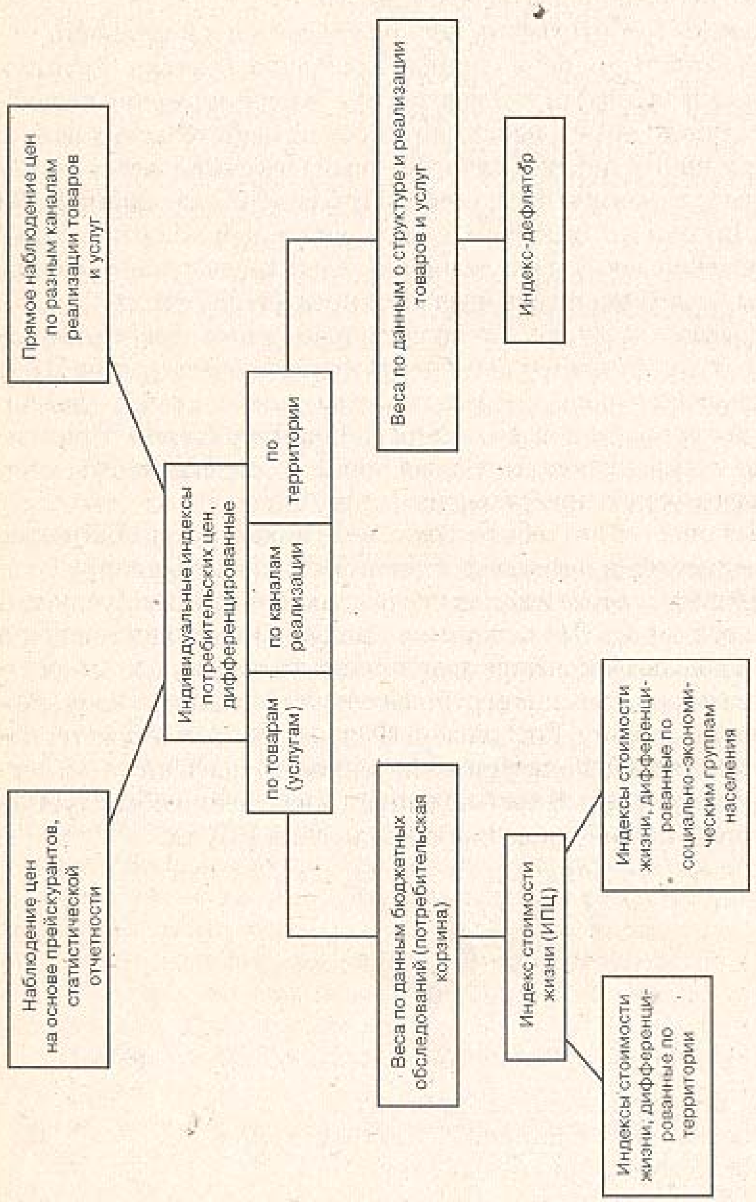

Рис. 9.1. Принципы построения системы индексов потребительских цен

селения, определяют сводные индексы цен по группам продовольственных, непродовольственных товаров и услуг, а затем и сводный индекс потребительских цен по региону и Российской Федерации в целом.

Определение расчетной формулы сводного индекса потребительских цен связано с задачей выбора между весами текущего и базисного периода. Удобство практического расчета индекса цен по формуле Ласпейреса сделали ее основной формулой для расчета ИПЦ (см. формулу 9.10), т.е. определяется отношение стоимости потребительской корзины базисного периода в текущих ценах к ее стоимости в базисном периоде. Следовательно, изменение потребительских цен представлено в виде переоценки физических объемов материальных благ и услуг, входящих в потребительскую корзину.

Как показано на рис. 9.1, практическая реализация методики расчета ИПЦ основана на использовании средней арифметической взвешенной по данным об изменении цен по отдельному товару и о размере потребительских расходов, а не физических объемов потребления. Следовательно, формула (9.10) может быть представлена следующим образом:

$$I_p = \frac{\sum_{i=1}^{n} \frac{p_i^t}{p_i^0}\cdot p_i^0\cdot q_i^0}{\sum_{i=1}^{n} p_i^0\cdot q_i^0} = \frac{\sum_{i=1}^{n} i_p\cdot Q_i^0}{\sum_{i=1}^{n} Q_i^0} \qquad (9.30)$$

где P_i^t - цена i-того товара за t-тый период времени;

$Q_i^0 = p_i^0\cdot q_i^0$ - стоимость потребительских расходов базисного периода по i-тому товару.

Формула (9.30) может быть преобразована и иметь следующий вид:

$$I_p = \sum_{i=1}^{n} i_p\cdot d_i^0, \qquad (9.30а)$$

где $d_i^0 = \frac{Q_i^0}{\sum_{i=1}^{n} Q_i^0}$, т.е. представляет собой долю расходов по конкретному i-тому товару в общем объеме потребительских расходов базисного периода.

Формула Ласпейреса для расчета ИПЦ используется в большинстве стран. Преимуществом этой формулы является постоянная структура потребительских расходов, а это означает, что для расчета этого индекса не требуется постоянный пересмотр системы весов. Но на практике при сравнении цен по каждому товару за длительные периоды времени нельзя не учитывать такие факторы, как изменение стереотипов потребления, изменение в предложении товаров и услуг и изменение доходов.

Например, в Соединенных Штатах Америки ИПЦ (consumer price index - CPI) рассчитывается по следующей формуле:

$$I_p = \frac{\sum_{i=1}^{n} p_i^t \cdot q_i^t}{\sum_{i=1}^{n} p_i^0 \cdot q_i^t}, \qquad (9.31)$$

где q_i^t - физический объем потребления i-того товара, фиксированный на каком-либо конкретном уровне t, или средний уровень потребления за несколько периодов времени. Использование этой формулы позволяет анализировать изменения в текущем периоде по сравнению с любым другим периодом, а не только с тем годом, когда проводилось обследование потребительских цен.

В России в условиях высокой инфляции структура потребления товаров и услуг используется как постоянная в течение года и пересчитывается каждый год. ИПЦ определяется за каждый месяц и нарастающим итогом с начала года по модифицированной (базисно-взвешенной) формуле Ласпейреса:

$$I_p = \frac{\sum_{i=1}^{n} \frac{p_i^t}{p_i^{t-1}} \cdot p_i^{t-1} \cdot q_i^0}{\sum_{i=1}^{n} p_i^0 \cdot q_i^0}, \qquad (9.32)$$

где $p_i^{t-1} \cdot q_i^0 = q_i^0 \cdot p_i^0 \cdot \frac{p_i^1}{p_i^0} \cdot \frac{p_i^2}{p_i^1} \cdot \ldots \cdot \frac{p_i^{t-1}}{p_i^{t-2}}$.

Например, для третьего месяца

$$I_{p(2/0)} = \frac{\sum \frac{p_i^2}{p_i^1} \cdot \frac{p_i^1}{p_i^0} \cdot p_i^0 \cdot q_i^0}{\sum p_i^0 \cdot q_i^0}.$$

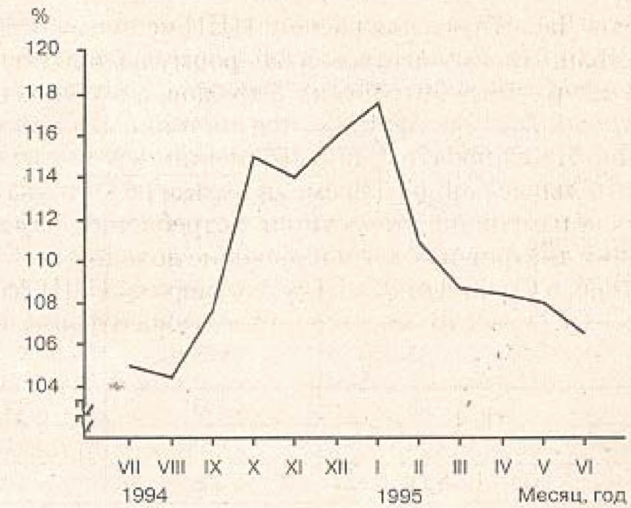

Рис. 9.2. **Индекс потребительских цен с июля 1994 г. по июнь 1995 г.** (в % к предыдущему месяцу)

Бюллетень банковской статистики, вып. 1(20) М., 1995. С. 4; вып. 6(25) М., 1995. С. 4.

На рис. 9.2 представлена динамика потребительских цен с июля 1994 г. по июнь 1995 г. Как видно из графика, в течение рассматриваемого периода цены изменялись крайне неравномерно. Так, максимальное их увеличение наблюдалось в январе 1995 г. (на 17,8%). К концу первого полугодия темпы роста цен постепенно снижались.

Рассмотрим на примере (табл. 9.9) анализ ИПЦ по основным товарным группам, включающим продукты питания.

Сводный индекс по товарным группам:

$$I_p = \sum_{i=1}^{5} I_{p_i} \cdot d_i^0 =$$

$$= 1{,}58 \cdot 0{,}3717 + 1{,}99 \cdot 0{,}3109 + 1{,}89 \cdot 0{,}0650 + 1{,}08 \cdot 0{,}1631 + 2{,}34 \cdot 0{,}0893 = 1{,}7139.$$

Следовательно, цены по основным товарным группам возросли в июне 1994 г. по сравнению с декабрем 1993 г. на 71,39%.

Как было сказано выше, для определения ИПЦ по каждой территории рассчитываются индексы по продовольственным, непродовольственным товарам и платным услугам и строится сводный индекс. Покажем расчет сводного индекса по всем потребительским товарам на примере (табл. 9.10).

Таблица 9.9

Расчет индекса цен по основным товарным группам в России (на основе фактических данных за декабрь 1993 г. и июнь 1994 г.)

Товарные группы	Индексы цен за июнь 1994 г. к декабрю 1993 г.	Удельный вес каждой товарной группы	
		декабрь 1993 г.	июнь 1994 г.
Мясо и мясопродукты	1,58	0,3717	0,2875
Молоко и молочные продукты	1,99	0,3109	0,2769
Рыба и рыбопродукты	1,89	0,0650	0,0646
Сахар	1,08	0,1631	0,1106
Хлебные продукты	2,34	0,0893	0,2604
Итого	1,71	1,0000	1,0000

Составлено по: Российский статистический ежегодник. М., Госкомстат России, 1994.

Таблица 9.10

Расчет индекса потребительских цен в России (на основе фактических данных за декабрь 1993 г. и июнь 1994 г.)

Товарные группы	Индексы цен за июнь 1994 г. к декабрю 1993 г.	Удельный вес каждой товарной группы, декабрь 1993 г.
Продовольственные товары	1,69	0,494
Непродовольственные товары	1,54	0,424
Платные услуги	3,23	0,082
Итого	1,75	1,000

Составлено по: Российский статистический ежегодник М., Госкомстат России. 1994.

Сводный индекс потребительских цен:

$$I_p = \sum_{i=1}^{3} I_{p_i} \cdot d_i^0 = 1{,}69 \cdot 0{,}494 + 1{,}54 \cdot 0{,}424 + 3{,}23 \cdot 0{,}082 = 1{,}75.$$

Следовательно, цены на потребительские товары возросли в июне 1994 г. по сравнению с декабрем 1993 г. на 75%.

Для макроэкономического анализа процессов в экономике используется индекс-дефлятор, который оценивает степень инфляции по всей совокупности благ, производимых и потребляемых в государстве, т.е. с учетом инвестиций, экспорта и импорта товаров и услуг. В российской статистике применяют дефлятор валового внутреннего продукта с квартальной периодичностью его расчета. Поскольку расчет индекса-дефлятора должен быть максимально приближен к совокупности товаров, произведенных в отчетном периоде, в большинстве стран мира его определяют по формуле индекса Пааше.

В этой главе мы показали основные формы расчета индексов и направления их использования. Чрезвычайно разнообразны индексы, используемые в анализе различных экономических показателей, их приложение к решению конкретных экономических задач рассматриваются в курсах статистики отраслей и макроэкономической статистики.

Контрольные вопросы к главе 9

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 формы индексов используют при территориальных сопоставлениях?

ПРИЛОЖЕНИЕ I

Таблица значений функции $f(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}$.

t	0	1	2	3	4	5	6	7	8	9
0,0	3989	3989	3989	3988	3986	3984	3982	3980	3977	3973
0,1	3970	3965	3961	3956	3951	3945	3939	3932	3925	3918
0,2	3910	3902	3894	3885	3876	3867	3857	3847	3836	3825
0,3	3814	3802	3790	3778	3765	3752	3739	3725	3712	3697
0,4	3683	3668	3653	3637	3621	3605	3589	3572	3555	3538
0,5	3521	3503	3485	3467	3448	3429	3410	3391	3372	3352
0,6	3332	3312	3292	3271	3251	3230	3209	3187	3166	3144
0,7	3123	3101	3079	3056	3034	3011	2989	2966	2943	2920
0,8	2897	2874	2850	2827	2803	2780	2756	2732	2709	2685
0,9	8661	2637	2613	2589	2565	2541	2516	2492	2468	2444
1,0	2420	2396	2371	2347	2323	2299	2275	2251	2227	2203
1,1	2179	2155	2131	2107	2083	2059	2036	2012	1989	1965
1,2	1942	1919	1895	1872	1849	1826	1804	1781	1758	1736
1,3	1714	1691	1669	1647	1626	1604	1582	1561	1539	1518
1,4	1497	1476	1456	1435	1415	1394	1374	1354	1334	1315
1,5	1295	1276	1257	1238	1219	1200	1182	1163	1145	1127
1,6	1109	1092	1074	1057	1040	1023	1006	0989	0973	0957
1,7	0940	0925	0909	0893	0878	0863	0848	0833	0818	0804
1,8	0790	0775	0761	0748	0734	0721	0707	0694	0681	0669
1,9	0656	0644	0632	0620	0608	0596	0584	0573	0562	0551
2,0	0540	0529	0519	0508	0498	0488	0478	0468	0459	0449
2,1	0440	0431	0422	0413	0404	0396	0387	0379	0371	0363
2,2	0855	0347	0339	0332	0325	0317	0310	0303	0297	0290
2,3	0283	0277	0270	0264	0258	0252	0246	0241	0235	0229
2,4	0224	0219	0213	0203	0203	0198	0194	0189	0184	0180
2,5	0175	0171	0167	0163	0158	0154	0151	0147	0143	0139
2,6	0136	0132	0129	0126	0122	0119	0116	0113	0110	0107
2,7	0104	0101	0099	0096	0093	0091	0088	0086	0084	0081
2,8	0079	0077	0075	0073	0071	0069	0067	0065	0063	0061
2,9	0060	0058	0056	0055	0053	0051	0050	0048	0047	0046
3,0	0044	0043	0042	0040	0039	0038	0037	0036	0035	0034
4,0	0001	0001	0001	0000	0000	0000	0000	0000	0000	0000

ПРИЛОЖЕНИЕ II

Таблица значений фукнции Лапласа при разных значениях t

$$\Phi(t) = \frac{2}{\sqrt{2\pi}} \int_0^t e^{-\frac{t^2}{2}} dt$$

t	Φ(t)	t	Φ(t)	t	Φ(t)	t	Φ(t)
0,00	0,00000	0,30	0,23582	0,60	0,45149	0,90	0,63188
01	00798	31	24344	61	45814	91	63718
02	01596	32	25103	62	46474	92	64243
03	02393	33	25860	63	47131	93	64763
04	03191	34	26614	64	47783	94	65278
05	03988	35	27366	65	48431	95	65789
06	04784	36	28115	66	49075	96	66294
07	05581	37	28862	67	49714	97	66795
08	06376	38	29605	68	50350	98	67291
09	07171	39	30346	69	50981	99	67783
0,10	0,07966	0,40	0,31084	0,70	0,51607	1,00	0,68269
11	08759	41	31819	71	52230	01	68750
12	09552	42	32552	72	52848	02	69227
13	10348	43	33280	73	53461	03	69699
14	11134	44	34006	74	54070	04	70166
15	11924	45	34729	75	54675	1,05	70628
16	12712	46	35448	76	55275	06	71086
17	13499	47	36164	77	55870	07	71538
18	14285	48	36877	78	56461	08	71986
19	15069	49	37587	79	57047	09	72429
0,20	0,15852	0,50	0,38292	0,80	0,57629	1,10	0,72867
21	16633	51	38995	81	58206	11	73300
22	17413	52	39694	82	58778	12	73729
23	18191	53	40389	83	59346	13	74152
24	18967	54	41080	84	59909	14	74571
25	19741	55	41768	85	60468	15	74986
26	20514	56	42452	86	61021	16	75395
27	21284	57	43132	87	61570	17	75800
28	22052	58	43809	88	62114	18	76200
29	22818	59	44481	89	62953	19	76595

Продолжение

t	Φ(t)	t	Φ(t)	t	Φ(t)	t	Φ(t)
1,20	0,76986	1,55	0,87886	1,90	0,94257	2,25	0,97555
21	77372	56	88124	91	94387	26	97618
22	77754	57	88358	92	94514	27	97679
23	78130	58	88589	93	94639	28	97739
24	78502	59	88817	94	94792	29	97798
25	78870	1,60	0,89040	95	94882	2,30	0,97855
26	79233	61	89260	96	95000	31	97911
27	79592	62	89477	97	95116	32	97966
28	79945	63	89690	98	95230	33	98019
29	80295	64	89899	99	95341	34	98072
1,30	0,80640	65	90106	2,00	0,95450	35	98123
31	80980	66	90309	01	95557	36	98172
32	81316	67	90508	02	95662	37	98221
33	81948	68	90704	03	95764	38	98269
34	81975	69	90897	04	95865	39	98315
35	82298	1,70	0,91087	05	95964	2,40	0,98360
36	82617	71	91273	06	96060	41	98405
37	82931	72	91457	07	96155	42	98448
38	83241	73	91637	08	96247	43	98490
39	83547	74	91814	09	96338	44	98531
1,40	0,83849	75	91988	2,10	0,96427	45	98571
41	84146	76	92159	11	96514	46	98611
42	84439	77	92327	12	96599	47	98649
43	84728	78	92492	13	96683	48	98686
44	85013	79	92655	14	96765	49	98723
45	85294	1,80	0,92814	15	96844	2,50	0,98758
46	85571	81	92970	16	96923	51	98793
47	85844	82	93124	17	96999	52	98826
48	86113	83	93275	18	97074	53	98859
49	86328	84	93423	19	97148	54	98891
1,50	0,86639	85	93569	2,20	0,97219	55	98923
51	86696	86	93711	21	97289	56	98953
52	87140	87	93852	22	97358	57	98983
53	87398	88	93989	23	97425	58	99012
54	87644	89	94124	24	97491	59	99040

Продолжение

t	$\Phi(t)$	t	$\Phi(t)$	t	$\Phi(t)$	t	$\Phi(t)$
2,60	0,99068	2,95	0,99682	3,30	0,99903	3,65	0,99974
61	99095	96	99692	31	99907	66	99975
62	99121	97	99702	32	99910	67	99976
63	99146	98	99712	33	99913	68	99977
64	99171	99	99721	34	99916	69	99978
65	99195	3,00	0,99730	35	99919	3,70	0,99978
66	99219	01	99739	36	99922	71	99979
67	99241	02	99747	37	99925	72	99980
68	99283	03	99755	38	99928	73	99981
69	99285	04	99763	39	99930	74	99982
2,70	0,99307	05	99771	3,40	0,99933	75	99982
71	99327	06	99779	41	99935	76	99983
72	99347	07	99786	42	99937	77	99984
73	99367	08	99793	43	99940	78	99984
74	99386	09	99800	44	99942	79	99985
75	99404	3,10	0,99806	45	99944	3,80	0,99986
76	99422	11	99813	46	99946	81	99986
77	99439	12	99819	47	99948	82	99987
78	99456	13	99825	48	99950	83	99987
79	99473	14	99831	49	99952	84	99988
2,80	0,99489	15	99837	3,50	0,99953	85	99988
81	99505	16	99842	51	99955	86	99989
82	99520	17	99848	52	99957	87	99989
83	99535	18	99853	53	99958	88	99990
84	99549	19	99858	54	99960	89	99990
85	99563	3,20	0,99863	55	99961	3,90	0,99990
86	99576	21	99867	56	99963	91	99991
87	99590	22	99872	57	99964	92	99991
88	99502	23	99876	58	99966	93	99992
89	99615	24	99880	59	99967	94	99992
2,90	0,99627	25	99855	3,60	0,99968	95	99992
91	99639	26	99889	61	99969	96	99992
92	99650	27	99892	62	99971	97	99993
93	99661	28	99896	63	99972	98	99993
94	99672	29	99800	64	99973	99	99993

ПРИЛОЖЕНИЕ III

Таблица случайных чисел

5489	5583	3156	0835	1988	3912	0938	7460	0869	4420
3522	0935	7877	5665	7020	9555	7375	7124	7878	5544
7555	7579	2550	2487	9477	0864	2349	1012	8250	2633
5759	3554	5080	9074	7001	6249	3224	6368	9102	2672
6303	6895	3371	3196	7231	2918	7380	0438	7547	2644
7351	5634	5323	2623	7803	8374	2191	0464	0696	9529
7068	7803	8832	5119	6350	0120	5026	3684	5657	0304
3613	1428	1796	8447	0503	5654	3254	7336	9536	1944
5143	4534	2105	0368	7890	2473	4240	8652	9435	1422
9815	5144	7649	8638	6137	8070	5345	4865	2456	5708
5780	1277	6816	1013	2867	9938	3930	3203	5696	1769
1187	0951	5991	5245	5700	5564	7352	0891	6249	6568
4184	2179	4554	9083	2254	2435	2965	5154	1209	7069
2916	2972	9885	0275	0144	8034	8122	3213	7666	0230
5524	1341	9860	6565	6981	9842	0171	2284	2707	3008
0146	5291	2354	5694	0377	5336	6460	9585	3415	2358
4920	2826	5238	5402	7937	1993	4332	2327	6875	5230
7978	1947	6380	3425	7267	7285	1130	7722	0164	8573
7453	0653	3645	7497	5969	8682	4191	2976	0361	9334
1473	6938	4899	5348	1641	3652	0852	5296	4538	4456
8162	8797	8000	4707	1880	9660	8446	1883	9768	0881
5645	4219	0807	3301	4279	4168	4305	9937	3120	5547
2042	1192	1175	8851	6432	4635	5757	6656	1660	5389
5470	7702	6958	9080	5925	8519	0127	9233	2452	7341
4045	1730	6005	1704	0345	3275	4738	4862	2556	8333
5880	1257	6163	4439	7276	6353	6912	0731	9033	5294
9083	4260	5277	4998	4298	5204	3965	4028	8936	5148
1762	8713	1189	1090	8989	7273	3213	1935	9321	4820
2023	2589	1740	0424	8924	0005	1969	1636	7237	1227
7965	3855	4765	0703	1678	0841	7543	0308	9732	1289
7690	0480	8098	9629	4819	7219	7241	5128	3853	1921
9292	0426	9573	4903	5916	6576	8368	3270	6641	0033
0867	1656	7016	4220	2533	6345	8227	1904	5138	2537
0505	2127	8255	5276	2233	3956	4118	8199	6380	6340
6295	9795	1112	5761	2575	6837	3336	9322	7403	8345
6323	2615	3410	3365	1117	2417	3176	2434	5240	5455
8672	8536	2966	5773	5412	8114	0930	4697	6919	4569
1422	5507	7596	0670	3013	1351	3886	3268	9469	2584
2653	1472	5113	5735	1469	9545	9331	5303	9914	6394
0438	4376	3328	8649	8327	0110	4549	7955	5275	2890
2851	2157	0047	7085	1129	0460	6821	8323	2572	8962
7962	2753	3077	8718	7418	8004	1425	3706	8822	1494
3837	4098	0220	1217	4732	0150	1637	1097	1040	7372
8542	4126	9274	2251	0607	4301	8730	7690	6235	3477
0139	0765	8039	9484	2577	7859	1976	0623	1418	6685
6687	1943	4307	0579	8171	8224	8641	7034	3595	3875
6242	5582	5872	3197	4919	2792	5991	4058	9769	1918
6859	9606	0522	4993	0345	8958	1289	8825	6941	7685
6590	1932	6043	3623	1973	4112	1795	8465	2110	8045
3482	0478	0221	6738	7323	5643	4767	0106	2272	9862

ПРИЛОЖЕНИЕ IV

Значения α-процентных пределов $t_{\alpha k}$ в зависимости от k степеней свободы и заданного уровня значимости α для распределения Стьюдента

k \ α	10,0	5,0	2,5	2,0	1,0	0,5	0,3	0,2	0,1
1	6,314	12,706	25,452	31,821	63,657	127,3	212,2	318,3	636,6
2	2,920	4,303	6,205	6,965	9,925	14,089	18,216	22,327	31,600
3	2,353	3,182	4,177	4,541	5,841	7,453	8,891	10,214	12,922
4	2,132	2,776	3,495	3,747	4,604	5,597	6,435	7,173	8,610
5	2,015	2,571	3,163	3,365	4,032	4,773	5,376	5,893	6,869
6	1,943	2,447	2,969	3,143	3,707	4,317	4,800	5,208	5,959
7	1,895	2,365	2,841	2,998	3,499	4,029	4,442	4,785	5,408
8	1,860	2,306	2,752	2,696	3,355	3,833	4,199	4,501	5,041
9	1,833	2,262	2,685	2,821	3,250	3,690	4,024	4,297	4,781
10	1,812	2,228	2,634	2,764	3,169	3,581	3,892	4,144	4,587
12	1,782	2,179	2,560	2,681	3,055	3,428	3,706	3,930	4,318
14	1,761	2,145	2,510	2,624	2,977	3,326	3,583	3,787	4,140
16	1,746	2,120	2,473	2,583	2,921	3,252	3,494	3,686	4,015
18	1,734	2,101	2,445	2,552	2,878	3,193	8,428	3,610	3,922
20	1,725	2,086	2,423	2,528	2,845	3,153	3,376	3,552	3,849
22	1,717	2,074	2,405	2,508	2,819	3,119	3,335	3,505	3,792
24	1,711	2,064	2,391	2,492	2,797	3,092	3,302	3,467	3,745
26	1,706	2,056	2,379	2,479	2,779	3,067	3,274	3,435	3,704
28	1,701	2,048	2,369	2,467	2,763	3,047	3,250	3,408	3,674
30	1,697	2,042	2,360	2,457	2,750	3,030	3,230	3,386	3,646
∞	1,645	1,960	2,241	2,326	2,576	2,807	2,968	3,090	3,291

ПРИЛОЖЕНИЕ V

Значения верхнего α % предела $\chi^2\alpha$ в зависимости от вероятности $P(\chi^2 > \chi^2\alpha)$ и числа степеней свободы χ^2-распределения

Число степеней свободы k	Вероятность $P(\chi^2 > \chi^2\alpha)$							
	0,99	0,98	0,95	0,90	0,80	0,70	0,50	0,30
1	1,64	2,7	3,8	5,4	6,6	7,9	9,5	10,83
2	3,22	4,6	6,0	7,8	9,2	11,6	12,3	13,8
3	4,64	6,3	7,8	9,8	11,3	12,8	14,8	16,3
4	6,0	7,8	9,5	11,7	13,3	13,9	16,9	18,5
5	7,3	9,2	11,1	13,4	15,1	16,3	18,9	20,5
6	8,6	10,6	12,6	15,0	16,8	18,6	20,7	22,5
7	9,8	12,0	14,1	16,6	18,5	20,3	22,6	24,3
8	11,0	13,4	15,5	18,2	20,1	21,9	24,3	26,1
9	12,2	14,7	16,9	19,7	21,7	23,6	26,1	27,9
10	13,4	16,0	18,3	21,2	23,2	25,2	27,7	29,6
11	14,6	17,3	19,7	22,6	24,7	26,8	29,4	31,3
12	15,8	18,5	21,0	24,1	26,2	28,3	31,0	32,9
13	17,0	19,8	22,4	25,5	27,7	29,8	32,5	34,5
14	18,2	21,1	23,7	26,9	29,1	31,0	34,0	36,1
15	19,3	22,3	25,0	28,3	30,6	32,5	35,5	37,7
16	20,5	23,5	26,3	29,6	32,0	34,0	37,0	39,2
17	21,6	24,8	27,6	31,0	33,4	35,5	38,5	40,8
18	22,8	26,0	28,9	32,3	34,8	37,0	40,0	42,3
19	23,9	27,2	30,1	33,7	36,2	38,5	41,5	43,8
20	25,0	28,4	31,4	35,0	37,6	40,0	43,0	45,3
21	26,2	29,6	32,7	36,3	38,9	41,5	44,5	46,8
22	27,3	30,8	33,9	37,7	40,3	42,5	46,0	48,3
23	28,4	32,0	35,2	39,0	41,6	44,0	47,5	49,7
24	29,6	33,2	36,4	40,3	43,0	45,5	48,5	51,2
25	30,7	34,4	37,7	41,6	44,3	47,0	50,0	52,6
26	31,8	35,6	38,9	42,9	45,6	48,0	51,5	54,1
27	32,9	36,7	40,1	44,1	47,0	49,5	53,0	55,5
28	34,0	37,9	41,3	45,4	48,3	51,0	54,5	56,9
29	35,1	39,1	42,6	46,7	49,9	52,5	56,0	58,3
30	36,3	40,3	43,8	48,0	50,9	54,0	57,5	59,7

Значения пяти- и однопроцентных
в зависимости от степеней свободы k_1 и k_2
набраны обычным шрифтом,

k_2 - степени свободы для меньшей дисперсии	k_1 - степени свободы: 1	2	3	4	5	6	7	8	9	10	11	12
1	161	200	216	225	230	234	237	239	241	242	243	244
	4 052	**4 999**	**5 403**	**5 625**	**5 764**	**5 889**	**5 928**	**5 981**	**6 022**	**6 056**	**6 082**	**6 106**
2	18,51	19,00	19,16	19,25	19,30	19,33	19,36	19,37	19,38	19,39	19,40	19,41
	98,49	**99,01**	**99,17**	**99,25**	**99,30**	**99,33**	**99,34**	**99,36**	**99,38**	**99,40**	**99,41**	**99,42**
3	10,13	9,55	9,28	9,12	9,01	8,94	8,88	8,84	8,81	8,78	8,76	8,74
	34,12	**30,81**	**29,46**	**28,71**	**28,24**	**27,91**	**27,67**	**27,49**	**27,34**	**27,23**	**27,13**	**27,05**
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,96	5,93	5,91
	21,20	**18,00**	**16,69**	**15,98**	**15,52**	**15,21**	**14,98**	**14,80**	**14,66**	**14,54**	**14,45**	**14,37**
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,78	4,74	4,70	4,68
	16,26	**13,27**	**12,06**	**11,39**	**10,97**	**10,67**	**10,45**	**10,27**	**10,15**	**10,05**	**9,96**	**9,89**
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06	4,03	4,00
	13,74	**10,92**	**9,78**	**9,15**	**8,75**	**8,47**	**8,26**	**8,10**	**7,98**	**7,87**	**7,79**	**7,72**
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,63	3,60	3,57
	12,25	**9,55**	**8,45**	**7,85**	**7,46**	**7,19**	**7,00**	**6,84**	**6,71**	**6,62**	**6,54**	**6,47**
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,34	3,31	3,28
	11,26	**8,65**	**7,59**	**7,01**	**6,63**	**6,37**	**6,19**	**6,03**	**5,91**	**5,82**	**5,74**	**5,67**
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,13	3,10	3,07
	10,56	**8,02**	**6,99**	**6,42**	**6,06**	**5,80**	**5,62**	**5,47**	**5,35**	**5,26**	**5,18**	**5,11**
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,97	2,94	2,91
	10,04	**7,56**	**6,55**	**5,99**	**5,64**	**5,39**	**5,21**	**5,06**	**4,95**	**4,85**	**4,78**	**4,71**
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,86	2,82	2,79
	9,85	**7,20**	**6,22**	**5,67**	**5,32**	**5,07**	**4,88**	**4,74**	**4,63**	**4,54**	**4,46**	**4,40**
12	4,75	3,88	3,49	3,26	3,11	3,00	2,92	2,85	2,80	2,76	2,72	2,69
	9,33	**6,93**	**5,95**	**5,41**	**5,06**	**4,82**	**4,65**	**4,50**	**4,39**	**4,30**	**4,22**	**4,16**
13	4,67	3,80	3,41	3,18	3,02	2,92	2,84	2,77	2,72	2,67	2,63	2,60
	9,07	**6,70**	**5,74**	**5,20**	**4,86**	**4,62**	**4,44**	**4,30**	**4,19**	**4,10**	**4,02**	**3,96**
14	4,60	3,74	3,34	3,11	2,96	2,85	2,77	2,70	2,65	2,60	2,56	2,53
	8,86	**6,51**	**5,56**	**5,03**	**4,69**	**4,46**	**4,28**	**4,14**	**4,03**	**3,94**	**3,86**	**3,80**
15	4,54	3,68	3,29	3,06	2,90	2,79	2,70	2,64	2,59	2,55	2,51	2,48
	8,68	**6,36**	**5,42**	**4,89**	**4,56**	**4,32**	**4,14**	**4,00**	**3,89**	**3,80**	**3,73**	**3,67**
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54	2,49	2,45	2,42
	8,53	**6,23**	**5,29**	**4,77**	**4,44**	**4,20**	**4,03**	**3,89**	**3,78**	**3,69**	**3,61**	**3,55**
17	4,45	3,59	3,20	2,96	2,81	2,70	2,62	2,55	2,50	2,45	2,41	2,38
	8,40	**6,11**	**5,18**	**4,67**	**4,34**	**4,10**	**3,93**	**3,79**	**3,68**	**3,59**	**3,52**	**3,45**

ПРИЛОЖЕНИЕ VI

верхних пределов уклонения величины F
(пятипроцентные пределы $F_{5,}\ k_1k_2$
однопроцентные $F_{1,}\ k_1k_2$ - жирным)

для большей дисперсии: 14	16	20	24	30	40	50	75	100	200	500	∞	k_2
245	246	248	249	250	251	252	253	253	254	254	254	1
6 142	**6 169**	**6 208**	**6 234**	**6 258**	**6 286**	**6 302**	**6 323**	**6 334**	**6 352**	**6 361**	**6 366**	
19,42	19,43	19,44	19,45	19,46	19,47	19,47	19,48	19,49	19,49	19,50	19,50	2
99,43	**99,44**	**99,45**	**99,46**	**99,47**	**99,48**	**99,48**	**99,49**	**99,49**	**99,49**	**99,50**	**99,50**	
8,71	8,69	8,66	8,64	8,62	8,60	8,58	8,57	8,56	8,54	8,54	8,53	3
26,92	**26,83**	**26,69**	**26,60**	**26,50**	**26,41**	**26,35**	**26,27**	**26,23**	**26,18**	**26,14**	**26,12**	
5,87	5,84	5,80	5,77	5,74	5,71	5,70	5,68	5,66	5,65	5,64	5,63	4
14,24	**14,15**	**14,02**	**13,93**	**13,83**	**13,74**	**13,69**	**13,61**	**13,57**	**13,52**	**13,48**	**13,46**	
4,64	4,60	4,56	4,53	4,50	4,46	4,44	4,42	4,40	4,38	4,37	4,36	5
9,77	**9,68**	**9,55**	**9,47**	**9,38**	**9,29**	**9,24**	**9,17**	**9,13**	**9,07**	**9,04**	**9,02**	
3,96	3,92	3,87	3,84	3,81	3,77	3,75	3,72	3,71	3,69	3,68	3,67	6
7,60	**7,52**	**7,39**	**7,31**	**7,23**	**7,14**	**7,09**	**7,02**	**6,99**	**6,94**	**6,90**	**6,88**	
3,52	3,49	3,44	3,41	3,38	3,34	3,32	3,29	3,28	3,25	3,24	3,23	7
6,35	**6,27**	**6,15**	**6,07**	**5,98**	**5,90**	**5,85**	**5,78**	**5,75**	**5,70**	**5,67**	**5,65**	
3,23	3,20	3,15	3,12	3,08	3,05	3,03	3,00	2,98	2,96	2,94	2,93	8
5,56	**5,48**	**5,36**	**5,28**	**5,20**	**5,11**	**5,06**	**5,00**	**4,96**	**4,91**	**4,88**	**4,86**	
3,02	2,98	2,93	2,90	2,86	2,82	2,80	2,77	2,76	2,73	2,72	2,71	9
5,00	**4,92**	**4,80**	**4,73**	**4,64**	**4,56**	**4,51**	**4,45**	**4,41**	**4,36**	**4,33**	**4,31**	
2,86	2,82	2,77	2,74	2,70	2,67	2,64	2,61	2,59	2,56	2,55	2,54	10
4,60	**4,52**	**4,41**	**4,33**	**4,25**	**4,17**	**4,12**	**4,05**	**4,01**	**3,96**	**3,93**	**3,91**	
2,74	2,70	2,65	2,61	2,57	2,53	2,50	2,47	2,45	2,42	2,41	2,40	11
4,29	**4,21**	**4,10**	**4,02**	**3,94**	**3,86**	**3,80**	**3,74**	**3,70**	**3,66**	**3,62**	**3,60**	
2,64	2,60	2,54	2,50	2,46	2,42	2,40	2,36	2,35	2,32	2,31	2,30	12
4,05	**3,98**	**3,86**	**3,78**	**3,70**	**3,61**	**3,56**	**3,49**	**3,46**	**3,41**	**3,38**	**3,36**	
2,55	2,51	2,46	2,42	2,38	2,34	2,32	2,28	2,26	2,24	2,22	2,21	13
3,85	**3,78**	**3,67**	**3,59**	**3,51**	**3,42**	**3,37**	**3,30**	**3,27**	**3,21**	**3,18**	**3,16**	
2,48	2,44	2,39	2,35	2,31	2,27	2,24	2,21	2,19	2,16	2,14	2,13	14
3,70	**3,62**	**3,51**	**3,43**	**3,34**	**3,26**	**3,21**	**3,14**	**3,11**	**3,06**	**3,02**	**3,00**	
2,43	2,39	2,33	2,29	2,25	2,21	2,18	2,15	2,12	2,10	2,08	2,07	15
3,56	**3,48**	**3,36**	**3,29**	**3,20**	**3,12**	**3,07**	**3,00**	**2,97**	**2,92**	**2,89**	**2,87**	
2,37	2,33	2,28	2,24	2,20	2,16	2,13	2,09	2,07	2,04	2,02	2,01	16
3,45	**3,37**	**3,25**	**3,18**	**3,10**	**3,01**	**2,96**	**2,89**	**2,86**	**2,80**	**2,77**	**2,75**	
2,33	2,29	2,23	2,19	2,15	2,11	2,08	2,04	2,02	1,99	1,97	1,96	17
3,35	**3,27**	**3,16**	**3,08**	**3,00**	**2,92**	**2,86**	**2,79**	**2,76**	**2,70**	**2,67**	**2,65**	

ПРИЛОЖЕНИЕ VII

Соотношение между r и z' для значений z' от 0 до 5[1]

z'	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,0000	0,0100	0,0200	0,0300	0,0400	0,0500	0,0599	0,0699	0,0798	0,0898
0,1	0,0997	0,1096	0,1194	0,1293	0,1391	0,1489	0,1587	0,1684	0,1781	0,1878
0,2	0,1974	0,2070	0,2165	0,2260	0,2355	0,2449	0,2543	0,2636	0,2729	0,2821
0,3	0,2913	0,3004	0,3095	0,3185	0,3275	0,3364	0,3452	0,3540	0,3627	0,3714
0,4	0,3800	0,3885	0,3969	0,4053	0,4136	0,4219	0,4301	0,4382	0,4462	0,4542
0,5	0,4621	0,4700	0,4777	0,4854	0,4930	0,5005	0,5080	0,5154	0,5227	0,5299
0,6	0,5370	0,5441	0,5511	0,5581	0,5649	0,5717	0,5784	0,5850	0,5915	0,5980
0,7	0,6044	0,6107	0,6169	0,6231	0,6291	0,6352	0,6411	0,6469	0,6527	0,6584
0,8	0,6640	0,6696	0,6751	0,6805	0,6858	0,6911	0,6963	0,7014	0,7064	0,7114
0,9	0,7163	0,7211	0,7259	0,7306	0,7352	0,7398	0,7443	0,7487	0,7531	0,7574
1,0	0,7616	0,7658	0,7699	0,7739	0,7779	0,7818	0,7857	0,7895	0,7932	0,7969
1,1	0,8005	0,8041	0,8076	0,8110	0,8144	0,8178	0,8210	0,8243	0,8275	0,8306
1,2	0,8337	0,8367	0,8397	0,8426	0,8455	0,8483	0,8511	0,8538	0,8565	0,8591
1,3	0,8617	0,8643	0,8668	0,8693	0,8717	0,8741	0,8764	0,8787	0,8810	0,8832
1,4	0,8854	0,8875	0,8896	0,8917	0,8937	0,8957	0,8977	0,8996	0,9015	0,9033
1,5	0,9052	0,9069	0,9087	0,9104	0,9121	0,9138	0,9154	0,9170	0,9186	0,9202
1,6	0,9217	0,9232	0,9246	0,9261	0,9275	0,9289	0,9302	0,9316	0,9329	0,9342
1,7	0,9354	0,9367	0,9379	0,9391	0,9402	0,9414	0,9425	0,9436	0,9447	0,9458
1,8	0,9468	0,9478	0,9498	0,9488	0,9508	0,9518	0,9527	0,9536	0,9545	0,9554
1,9	0,9562	0,9571	0,9579	0,9587	0,9595	0,9603	0,9611	0,9619	0,9626	0,9633
2,0	0,9640	0,9647	0,9654	0,9661	0,9668	0,9674	0,9680	0,9687	0,9693	0,9699
2,1	0,9705	0,9710	0,9716	0,9722	0,9727	0,9732	0,9738	0,9743	0,9748	0,9753
2,2	0,9757	0,9762	0,9767	0,9771	0,9776	0,9780	0,9785	0,9789	0,9793	0,9797
2,3	0,9801	0,9805	0,9809	0,9812	0,9816	0,9820	0,9823	0,9827	0,9830	0,9834
2,4	0,9837	0,9840	0,9843	0,9846	0,9849	0,9852	0,9855	0,9858	0,9861	0,9863
2,5	0,9866	0,9869	0,9871	0,9874	0,9876	0,9879	0,9881	0,9884	0,9886	0,9888
2,6	0,9890	0,9892	0,9895	0,9897	0,9899	0,9901	0,9903	0,9905	0,9906	0,9908

[1] Цифры таблицы являются значениями коэффициента корреляции r, соответствующими значениям z', указанными слева и сверху таблицы.

Продолжение

z'	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
2,7	0,9910	0,9912	0,9914	0,9915	0,9917	0,9919	0,9920	0,9922	0,9923	0,9925
2,8	0,9926	0,9928	0,9929	0,9931	0,9932	0,9933	0,9935	0,9936	0,9937	0,9938
2,9	0,9940	0,9941	0,9942	0,9943	0,9944	0,9945	0,9946	0,9947	0,9949	0,9950
3,0	0,9951									
4,0	0,9993									
5,0	0,9999									

ПРИЛОЖЕНИЕ VIII

Значения коэффициента корреляции рангов Спирмэна для двухсторонних пределов уровня значимости α

n \ α	0,20	0,10	0,05	0,02	0,01	0,002
4	0,8000	0,8000				
5	0,7000	0,8000	0,9000	0,9000		
6	0,6000	0,7714	0,8286	0,8857	0,9429	
7	0,5357	0,6786	0,7450	0,8571	0,8929	0,9643
8	0,5000	0,6190	0,7143	0,8095	0,8571	0,9286
9	0,4667	0,5833	0,6833	0,7667	0,8167	0,9000
10	0,4424	0,5515	0,6364	0,7333	0,7818	0,8667
11	0,4182	0,5273	0,6091	0,7000	0,7455	0,8364
12	0,3986	0,4965	0,5804	0,6713	0,7273	0,8182
13	0,3791	0,4780	0,5549	0,6429	0,6978	0,7912
14	0,3626	0,4593	0,5341	0,6220	0,6747	0,7670
15	0,3500	0,4429	0,5179	0,6000	0,6536	0,7464
16	0,3382	0,4265	0,5000	0,5824	0,6324	0,7265
17	0,3260	0,4118	0,4853	0,5637	0,6152	0,7083
18	0,3148	0,3994	0,4716	0,5480	0,5975	0,6904
19	0,3070	0,3895	0,4579	0,5333	0,5825	0,6737
20	0,2977	0,3789	0,4451	0,5203	0,5684	0,6586
21	0,2909	0,3688	0,4351	0,5078	0,5545	0,6455
22	0,2829	0,3597	0,4241	0,4963	0,5426	0,6318
23	0,2767	0,3518	0,4150	0,4852	0,5306	0,6186
24	0,2704	0,3435	0,4061	0,4748	0,5200	0,6070
25	0,2646	0,3362	0,3977	0,4654	0,5100	0,5962
26	0,2588	0,3299	0,3894	0,4564	0,5002	0,5856
27	0,2540	0,3236	0,3822	0,4481	0,4915	0,5757
28	0,2490	0,3175	0,3749	0,4401	0,4828	0,5660
29	0,2443	0,3113	0,3685	0,4320	0,4744	0,5567
30	0,2400	0,3059	0,3620	0,4251	0,4665	0,5479

ПРИЛОЖЕНИЕ IX

Общая структура номенклатуры Гармонизированной системы описания и кодирования товаров (ГС)

Код	Наименование разделов	Группа	Подгруппа	Товарная позиция	Товарная подпозиция	Товарная субпозиция
1	2	3	4	5	6	7
I	Живые животные и продукция животноводства	5	-	44	115	194
II	Продукты растительного происхождения	9	-	79	220	270
III	Жиры и масла животного или растительного происхождения; продукты их расщепления; приготовленные пищевые жиры; воски животного или растительного происхождения	1	-	22	44	53
IV	Продукция пищевой промышленности; алкогольные и безалкогольные напитки и уксус; табак и его заменители	9	-	56	160	181
V	Минеральные продукты	3	-	67	131	151
VI	Продукция химической и связанных с ней отраслей промышленности	11	19	176	505	759
VII	Пластмассы и изделия из них; каучук и резиновые изделия	2	2	43	138	189
VIII	Кожевенное сырье, кожа, пушнина, меховое сырье и изделия из них; шорно-седельные изделия и упряжь; дорожные принадлежности, сумки и подобные им товары; изделия из кишок (за исключением кетгута)	3	-	21	50	74
IX	Древесина и изделия из древесины; древесный уголь; пробка и изделия из нее; изделия из соломы, альфы и прочих материалов для плетения; корзиночные изделия и другие плетеные изделия	3	-	27	57	79

Продолжение

1	2	3	4	5	6	7
X	Бумажная масса из древесины или из других волокнистых растительных материоалов; бумажные и картонные отходы и макулатура; бумага, картон и изделия из них	3	-	41	117	149
XI	Текстиль и текстильные изделия	14	3	149	432	809
XII	Обувь, головные уборы, зонты, трости, трости-сидения, хлысты, кнуты и их части; обработанные перья и изделия их них; искусственные цветы; изделия из волоса	4	-	20	42	55
XII	Изделия из камня, гипса, цемента, асбеста, слюды и из подобных материалов; керамические изделия; стекло и изделия из него	3	2	49	112	138
XIV	Жемчуг природный или культивированный, драгоценные или полудрагоценные камни, драгоценные металлы, неблагородные металлы, покрытые и плакированные драгоценными металлами, и изделия из них; бижутерия; монеты	1	3	18	40	52
XV	Черные и цветные металлы и изделия из них	11	4	157	418	587
XVI	Машины, оборудование и механизмы; электрическое оборудование; их части; звукозаписывающая и воспроизводящая аппаратура; аппаратура для записи и воспроизведения телевизионного изображения и звука; их части и принадлежности	2	-	133	561	762
XVII	Средства наземного, воздушного и водяного транспорта и их части и принадлежности	4	-	38	100	132
XVIII	Приборы и аппараты оптические, фотографические, кинематографические, контрольные, прецизионные, медицинские и хирургические; часы; музыкальные инструменты, их части и принадлежности	3	-	56	184	230
XIX	Оружие и боеприпасы: их части и принадлежности	1	-	7	15	17

Продолжение

1	2	3	4	5	6	7
XX	Разные промышленные товары	3	-	32	108	131
XXI	Произведения искусства, предметы коллекционирования и антиквариат	1	-	6	7	7
	Итого	96	33	1241	3558	5010

Список рекомендуемой литературы

1. *Адамов В.Е.* Факторный индексный анализ. Методология и проблемы. - М.: Статистика, 1977.

2. *Аллен Р.* Экономические индексы. Пер. с англ. Л.С. Кучаева. Предисл. В.В. Мартынова. - М.: Статистика, 1980.

3. *Виноградова Н.М., Евдокимов В.Т., Хитарова Е.М., Яковлева Н.И.* Общая теория статистики. - М.: Статистика, 1968.

4. *Герчук Я.П.* Графики в математико-статистическом анализе. - М.: Статистика, 1972.

5. *Джессен Р.* Методы статистических обследований. Пер. с англ. Ю.П. Лукашина, Я.Ш. Паппэ; Под ред. Е.М. Четыркина. - М.: Финансы и статистика, 1985.

6. *Дружинин Н.К.* Математическая статистика в экономике. Введ. в мат.-стат. методологию. - М.: Статистика, 1971.

7. *Елисеева И.И.* Моя профессия – статистик. - М.:Финансы и статистика, 1992.

8. *Елисеева И.И., Юзбашев М.М.* Общая теория статистики/ Учебник. - М.: Финансы и статистика, 1995.

9. *Ефимова М.Р.* Статистические методы в управлении производством.- М.: Финансы и статистика. 1988.

10. *Ефимова М.Р., Рябцев В.М.* Общая теория статистики / Учебник.- М.: Финансы и статистика, 1991.

11. *Кендэл М.* Временные ряды/Пер. с англ. и предисл. Ю.П. Лукашина - М.: Финансы и статистика, 1981.

12. *Льюис К.Д.* Методы прогнозирования экономических показателей/Пер. с англ. и предисл. Е.З. Демиденко. - М.: Финансы и статистика, 1986.

13. *Миллс Ф.* Статистические методы/Пер. с англ. - М.: Госстатиздат, 1958.

14. *Пасхавер И.С.* Средние величины в статистике. - М.: Статистика, 1979.

15. *Плошко Б.Г., Елисеева И.И.* История статистики. - М.: Финансы и статистика, 1990.

16. *Рябушкин Т.В., Ефимова М.Р., Ипатова И.М., Яковлева Н.И.* Общая теория статистики/Учебник. - М.: Финансы и статистика, 1981.

17. Статистический анализ в экономике/Под ред. Г.Л. Громыко. - М.: Изд-во МГУ, 1992.

18. Статистический словарь/Под ред. М.А. Королева. 2-е изд. - М.: Финансы и статистика, 1989.

19. Статистическое моделирование и прогнозирование/Учеб. пособие/Под ред. А.Г. Гранберга. - М.: Финансы и статистика, 1990.

20. *Торвей Р.* Индексы потребительских цен. Методологическое руководство/Международная организация труда. Пер. с англ. - М.: Финансы и статистика, 1993.

21. *Четыркин Е.М.* Статистические методы прогнозирования. - М.: Статистика, 1975.

22. *Шураков В.В.* и др. Автоматизированное рабочее место для статистической обработки данных. - М.: Финансы и статистика, 1990.

Оглавление

Глава 1
Задачи статистики и ее организация 3

1.1. Общее представление о статистике и краткие сведения из ее истории 3
1.2. Предмет статистической науки и ее методология 10
1.3. Современная организация статистики в Российской Федерации и ее задачи 18
Контрольные вопросы к главе 1 26

Глава 2
Статистическое наблюдение 27

2.1. Формирование информационной базы статистического исследования 27
2.2. Программно-методологические и организационные вопросы статистического наблюдения 31
2.3. Виды статистического наблюдения и их особенности 38
Контрольные вопросы к главе 2 48

Глава 3
Группировка статистических данных и ее роль в анализе информации 49

3.1. Виды и задачи группировок 49
3.2. Некоторые вопросы техники выполнения группировки 58
3.3. Статистические таблицы 61
3.4. Графическое представление статистических данных 65
Контрольные вопросы к главе 3 73

Глава 4
Абсолютные, относительные и средние величины 75

4.1. Общие принципы построения статистических показателей 75
4.2. Абсолютные величины 77
4.3. Относительные величины 80

4.4. Виды средних величин и их значение в социально-экономических исследованиях 89
4.5. Средняя арифметическая, ее свойства и другие степенные средние 99
Контрольные вопросы к главе 4 106

Глава 5
Статистические распределения и их основные характеристики 107

5.1. Вариация признака в совокупности и значение ее изучения 107
5.2. Основные характеристики и графическое изображение вариационного ряда 115
5.3. Показатели центра распределения 119
5.4. Показатели вариации (колеблемости) признака 122
5.5. Моменты распределения 136
5.6. Изучение формы распределения 138
5.7. Критерии согласия 147
5.8. Вариационный ряд и группировка 152
Контрольные вопросы к главе 5 156

Глава 6
Выборочное наблюдение 157

6.1. Понятие о выборочном наблюдении и его теоретические основы 157
6.2. Простая случайная выборка 160
6.3. Определение необходимой численности выборки 172
6.4. Малые выборки 178
6.5. Статистическая проверка гипотез (Общие понятия. Выбор типа критической области) 182
6.6. Проверка гипотезы о принадлежности «выделяющихся» наблюдений исследуемой генеральной совокупности 186
6.7. Проверка гипотезы о величине средней арифметической и доли 192
6.8. Элементы дисперсионного анализа 198

6.9. Различные формы организации выборочного наблюдения 203
6.10. Практика применения выборочного метода наблюдения 212
Контрольные вопросы к главе 6 219

Глава 7
Корреляционная связь и ее статистическое изучение 221

7.1. Понятие о корреляционной связи и предпосылки ее использования 221
7.2. Статистические методы выявления наличия корреляционной связи между двумя признаками 225
7.3. Измерение степени тесноты корреляционной связи в случае парной зависимости 231
7.4. Уравнение регрессии 256
7.5. Множественная корреляция 269
Контрольные вопросы к главе 7 279

Глава 8
Ряды динамики 281

8.1. Виды рядов динамики 281
8.2. Показатели ряда динамики и методы их исчисления 287
8.3. Средние характеристики ряда динамики 293
8.4. Выявление и характеристика основной тенденции развития 298
8.5. Понятие сезонной неравномерности и ее характеристика 318
8.6. Корреляционная зависимость между уровнями различных рядов динамики 335
Контрольные вопросы к главе 8 338

Глава 9
Индексы и их использование в экономико-статистических исследованиях 339

9.1. Общее понятие об индексах и значение индексного метода анализа 339

9.2. Индексы количественных показателей 342
9.3. Индексы качественных показателей 353
9.4. Использование индексов в экономическом анализе 366
9.5. Использование индексов в макроэкономических исследованиях 377
Контрольные вопросы к главе 9 391

Приложения 393

Приложение I 393
Приложение II 394
Приложение III 397
Приложение IV 398
Приложение V 399
Приложение VI 401
Приложение VII 402
Приложение VIII 404
Приложение IX 405
Список рекомендуемой литературы 408

Марина Романовна Ефимова
Екатерина Валериановна Петрова
Владимир Николаевич Румянцев

ОБЩАЯ ТЕОРИЯ СТАТИСТИКИ

Учебник

Редактор Л.Г. Соловьева
Компьютерная верстка О.Н. Емельяновой
Художественное оформление «Ин-Арт»

ЛР № 070824 от 21.01.93 г.

Подписано в печать 23.04.99.
Формат 60x90/16. Усл. печ. л. 26,0.
Печать офсетная. Гарнитура «Петербург».
Доп. тираж 6000 экз.
Заказ 4204093.

Издательский Дом «ИНФРА-М»
127214, Москва, Дмитровское шоссе, 107.
Тел.: (095) 485-74-00, 485-70-63.
Факс: (095) 485-53-18. Робофакс: (095) 485-54-44.
E-mail: books@infra-m.ru
http://www.infra-m.ru

Отпечатано с готовых диапозитивов
на ФГУИПП «Нижполиграф».
603006, Нижний Новгород, ул. Варварская, 32.

ISBN 5-16-000012-7

9 785160 000121