Р.А. Шмойлова, В.Г. Минашкин, Н.А. Садовникова, Е.Б. Шувалова

ТЕОРИЯ СТАТИСТИКИ

Под редакцией профессора Р.А. Шмойловой

Пятое издание

Рекомендовано
Министерством образования Российской Федерации
в качестве учебника
для студентов экономических специальностей
высших учебных заведений

МОСКВА
"ФИНАНСЫ И СТАТИСТИКА"
2014

УДК 311(075.8)
ББК 60.6я73
Т33

АВТОРЫ:

Р.А. Шмойлова, В.Г. Минашкин,
Н.А. Садовникова, Е.Б. Шувалова

РЕЦЕНЗЕНТЫ:

Кафедра статистики
Российского государственного торгово-
экономического университета
(заведующая кафедрой – О.Э. Башина,
доктор экономических наук, профессор);

Г.Л. Громыко,
доктор экономических наук,
профессор кафедры статистики
экономического факультета МГУ им. М.В. Ломоносова

Т33 **Теория** статистики: учебник / Р.А. Шмойлова, В.Г. Минашкин, Н.А. Садовникова, Е.Б. Шувалова; под ред. Р.А. Шмойловой. – 5-е изд. – М.: Финансы и статистика, 2014. – 656 с.: ил.

ISBN 978-5-279-03295-2

Излагаются общие вопросы теории статистики. Рассматриваются: метод группировок, расчет относительных и средних величин, показатели вариации, корреляционный и регрессионный анализ, анализ временных рядов, выборочный метод и анализ частотных распределений, экономические индексы. Учебник снабжен приложением, содержащим математико-статистические таблицы, основные формулы и понятия теории статистики, а также предметным указателем.

Для аспирантов и преподавателей вузов, магистров, учащихся спецфакультетов второго высшего образования и программ MBA, специалистов аналитических служб и менеджеров.

УДК 311(075.8)
ББК 60.6я73

ISBN 978-5-279-03295-2

ОГЛАВЛЕНИЕ

ПРЕДИСЛОВИЕ

Коренные экономические преобразования в России, связанные с переходом на рыночные условия хозяйствования, существенно изменили требования к качеству подготовки экономистов. Сегодня, для того чтобы быть конкурентоспособным на рынке труда, необходимо владеть современным статистическим инструментарием анализа экономической информации. Предлагаемый учебник дает представление об основных статистических методах, их возможностях и границах применения.

Независимо от уровня и стадии экономического развития, характера политической системы статистика на протяжении сотен лет своего существования всегда была необходимым и эффективным инструментом государственного управления. Одновременно она была наукой, исследующей количественную сторону массовых явлений, выявляющей конкретные закономерности на основе закона больших чисел. Выполняя самые разнообразные функции сбора, систематизации и анализа сведений, характеризующих экономическое и социальное развитие общества, статистика всегда играла роль главного поставщика фактов для управленческих, научно-исследовательских и прикладных нужд различного рода структур, организаций и населения.

Роль статистики в нашей жизни настолько значительна, что люди, зачастую и не осознавая этого, постоянно используют элементы статистической методологии не только в трудовых процессах, но и в повседневном быту. Работая и отдыхая, общаясь с другими людьми, принимая какие-то решения, человек пользуется определенной системой сведений, сложившихся вкусов и привычек, фактов; он систематизирует, сопоставляет эти факты, анализирует их, делает необходимые для себя выводы и принимает определенные решения и действия. Таким образом, в каждом человеке заложены элементы статистического мышления как способность к анализу и синтезу информации об окружающем нас мире. Это так называемый обыденный компонент статистического мышления.

Цель настоящего учебника – развить заложенный в индивиде научно-исследовательский компонент статистического мышления. Речь пойдет о постижении множества специальных научных правил, методов и приемов количественного анализа разного рода информации. Что же необходимо для этого? Не очень много, но не так уж и мало. Нужна серьезность и вдумчивость, неплохое знание математики, истории, экономической теории, владение основами предпринимательства и информатики. Чем шире кругозор и эрудиция в самых различных областях знаний, тем значительнее успехи в изучении статистики.

Основными объектами приложения статистической теории и методологии являются экономическая деятельность, народонаселение, условия жизни людей и управление экономически-ми общественными процессами. Ядро статистической системы знаний – *теория статистики* – обеспечивает теоретическую и методологическую подготовку профессиональных статистиков, экономистов высшей квалификации, финансистов, менеджеров, коммерсантов и бухгалтеров, демографов и социологов, а также лиц иных профессиональных интересов, самостоятельно изучающих статистику. Настоящий учебник рассчитан на развитие исследовательских и предпринимательских навыков указанных групп специалистов в условиях современной российской действительности.

Основной задачей курса теории статистики является овладение знаниями общих основ статистической науки, искусством организации и проведения статистических исследований, анализа и обобщения их результатов, навыками прогнозирования.

Результатом изучения курса теории статистики должно стать знание принципов современной организации национальных и зарубежных статистических служб, категорий и понятий статистики; владение методами организации сбора, обработки данных (материалов) статистического наблюдения, их анализом с помощью обобщающих показателей, методов статистического моделирования и прогнозирования.

Студенты, изучившие курс, должны уметь организовать сплошное и несплошное наблюдение; строить статистические графики и таблицы; анализировать массивы статистических данных; исчислять и интерпретировать статистические показатели; формулировать выводы, вытекающие из проведенного анализа; осуществлять консалтинговые услуги заказчикам и потребителям обобщенной статистической информации.

Структура учебника несколько отлична от традиционной и отражает два основных направления статистики – описательное (дескриптивное) и аналитическое.

В разделе I – Описательная статистика – рассматриваются, во-первых, сущность статистики как науки и сферы ее практической деятельности; принципы и особенности статистической методологии; отраслевая структура статистики, ее основные понятия и категории. Во-вторых, раскрываются методы сбора статистической информации (формы, виды и способы статистического наблюдения); особенности методологии и практики проведения статистической сводки и группировки; методология построения различных видов статистических таблиц и графиков; методы исчисления абсолютных, относительных и средних величин, их использование в анализе социально-экономических явлений.

Раздел II – Аналитическая статистика – подробно рассматривает методы анализа вариации и частотных распределений; вопросы теории и практики выборочного наблюдения; методы и показатели оценки взаимосвязей признаков; методы и показатели изучения структуры и структурных сдвигов и различий; методологию статистического изучения динамики; основные характеристики, виды и способы исчисления экономических индексов, а также вопросы анализа и обобщения статистических данных.

Настоящее издание учебника, в отличие от предыдущего (3-е издание – 1999 г.), переработано и дополнено новыми материалами. Заново написаны главы: «Показатели вариации и анализ частотных распределений» и «Выборочное наблюдение». Теоретическую часть дополнили примеры, основанные на новых статистических данных.

Учебник подготовлен авторским коллективом кафедры теории статистики и прогнозирования Московского государственного университета экономики, статистики и информатики (МЭСИ) под руководством кандидата экономических наук, профессора *Р.А. Шмойловой.*

Труд авторов распределился следующим образом: *Р. А. Шмойлова,* профессор, кандидат экономических наук – предисловие, гл. 1, гл. 3 (разд. 3.5), гл. 5, гл. 7, гл. 10, приложения 12, 13, 14 и 16, предметный указатель; *В.Г. Минашкин,* профессор, кандидат экономических наук – гл. 6, гл. 8, гл. 11 и приложение 15; *Н.А. Садовникова,* доцент, кандидат экономических наук – гл. 3 (разд. 3.7), гл. 4, гл. 9 и гл. 13, приложения 1–11; *Е.Б. Шувалова,* доцент, кандидат экономических наук – гл. 2, гл. 3 (разд. 3.1, 3.2 (разд. 3.3 – совместно с Р. А. Шмойловой), разд. 3. 4 и 3.6); гл. 12 (совместно с Н.А. Садовниковой).

РАЗДЕЛ I

ОПИСАТЕЛЬНАЯ СТАТИСТИКА

ГЛАВА 1

СТАТИСТИКА КАК НАУКА

1.1 ПОНЯТИЕ СТАТИСТИКИ

Как заметили И. Ильф и Е. Петров в романе «Двенадцать стульев»: «Статистика знает все». Перефразировав афоризм, можно сказать: «О статистике слышали все». Например, мы регулярно слушаем сообщения Госкомстата РФ о выполнении планов федеральных программ развития национального хозяйства, читаем публикуемые сведения о количестве родившихся и умерших в стране, данные о браках и разводах и т.д. К рубрике в газетах «Немного статистики» мы так привыкли, что и не задумываемся, а знаем ли мы, собственно, что это за наука, как не задумываемся над смыслом ежедневно употребляемых слов и выражений.

Вместе с тем едва ли найдется еще одна такая наука, в которой бы так горячо спорили о ее предмете, т.е. о том, что она собой представляет. К тому же само понятие «статистика» употребляется в самых разных значениях.

Итак, что же означает термин «статистика»? В настоящее время насчитывается около тысячи определений статистики. Первое из них относится к 1749 г. Затем на протяжении 250 лет определение уточнялось и дополнялось. Определить статистику как науку пытались философы, математики, экономисты, социологи, государственные деятели и, конечно, сами статистики. Сначала статистику определяли как «Staaten Kunde» – государствоведение (описание достопримечательностей государств). Оставаясь на протяжении многих лет государствоведением, статистика постепенно отходила от описания достопримечательностей (их текстового изложения). Тем более что с развитием знаний вопросами государствоведения стали заниматься многие науки. В настоящее время под термином «статистика» чаще всего понимают следующее.

Статистика – это *самостоятельная общественная наука,* имеющая свой предмет исследования и свои специфические методы.

Статистика – это *эффективное орудие, инструмент познания,* используемый в естественных и общественных науках для установления тех специфических закономерностей, которые действуют в конкретных массовых явлениях, изучаемых данной наукой.

Статистика – это также *одна из форм практической деятельности* людей, цель которой – сбор, обработка и анализ массовых данных о тех или иных явлениях. Когда мы говорим: государственная и ведомственная статистика РФ, организация статистики в России, то как раз имеем в виду особую форму практической деятельности людей.

Статистикой называют также *различного рода числовые,* или, как часто говорят, *цифровые, данные*, характеризующие различные стороны жизни государства: политические отношения, культуру, население, производство и т.д. (табл. 1.1).

Таблица 1.1

Численность постоянного населения РФ за 1959–2001 гг. (на начало года)

Год	Все население, млн чел.	В том числе		В общей численности населения, %	
		городское	сельское	городское	сельское
1959	117,5	61,6	55,9	52,4	47,6
1970	129,9	80,6	49,3	62,0	38,0
1979	137,4	94,9	42,5	69,1	30,9
1989	147,0	107,9	39,1	73,4	26,6
1990	147,7	108,3	38,9	73,7	26,3
1995	147,9	107,9	40,0	73,0	27,0
2001	144,8	105,6	39,2	72,9	27,1

Часто слово «статистика» употребляется в качестве эквивалента для словосочетания *статистические методы*. Статистические методы применяют тогда, когда из большого массива данных требуется выделить полезную для нас информацию.

Иногда термин «статистика» может употребляться одновременно в нескольких значениях. Известный английский статистик У. Дж. Рейхман (р. 1920) заметил: «Мы живем в век статистики. Едва ли не в каждом своем аспекте явления природы, а также человеческая и прочая деятельность поддаются сейчас измерению при помощи статистических показателей» *(Рейхман У.Дж.* Применение статистики. – М.: Статистика, 1969. – С. 11).

Когда употребляют термин «статистика» с тем или иным эпитетом (красноречивая, удручающая и обнадеживающая), имеют в виду те или иные статистические данные, способные вызвать определенные эмоции. В этом смысле употребил термин «статистика» английский государственный деятель и писатель Б. Дизраэли: «Имеются три рода лжи: ложь, наглая ложь и статистика» *(Кимбл Г.* Как правильно пользоваться статистикой. – М.: Финансы и статистика, 1982. – С. 15).

Когда итальянский статистик К. Джини пишет, что статистика – это царица не только полной, но и неполной индукции, он имеет в виду не статистическую информацию, а статистический метод умозаключения. Любопытно и высказывание американского экономиста Митчелла: «Статистика – это солома, которую я, как и всякий другой экономист, должен спрессовать, чтобы получить брикеты» (Занимательная статистика / Под ред. Г. И. Бакланова, Г.С. Кильдишева. – М.: Статистика, 1980. – С. 10).

Многочисленные нелестные эпитеты, адресованные статистике, не затрагивают непосредственно ее как науку, рассматривающую совокупность методов, с помощью которых можно исследовать ту или иную конкретную совокупность социально-экономических явлений. Эти методы и статистические показатели в чистом виде сами по себе безупречны, но каждый из них имеет свои строго определенные условия и границы применения. Даже незначительное нарушение этих условий приводит к сбору недостаточно объективной и удовлетворительной информации.

Распространенное представление о возможности доказать любое явление с помощью статистики, конечно, слишком преувеличено, но и не лишено основания, однако несомненно, что и статистические методы могут ввести людей в заблуждение. Иногда даже квалифицированные специалисты, во всех тонкостях знающие статистику, одно и то же явление могут объяснить по-разному, принять ложное утверждение и отвергнуть правильное. Принятие утверждения как истинное в известной мере зависит от субъективных особенностей исследователя. Российские статистики Г. И. Бакланов и Г. С. Кильдишев отмечают, что «осторожный воздержится от принятия утверждения, а смелый примет его» (Занимательная статистика/Под ред. Г. И. Бакланова, Г. С. Кильдишева. – С. 11).

Следовательно, выводы, которые делаются на основании статистических данных, не всегда однозначны. Думается, основная задача и предназначение статистики в том, чтобы помочь людям лучше понять многие современные явления нашей жизни.

Сила статистики и в том, что она на основе анализа разрозненных, как бы пестрящих случайностями данных помогает исследователю проникнуть в существо изучаемых явлений. Прекрасно выразил эту мысль Г. Успенский в рассказе «Четверть лошади»: «А между тем только ведь в этих-то толстых скучных книгах и сказана цифрами та «сущая» правда нашей жизни, о которой мы совершенно отвыкли говорить человеческим языком, и нужно только раз получить интерес к этим дробям, нулям, нуликам, к этой вообще цифровой крупе, которою усеяны статистические книги и таблицы, так все они, вся эта крупа цифр начнет принимать человеческие образы и облекаться в картины ежедневной жизни, т.е. начнет получать значение не мертвых и скучных знаков, а, напротив, значение самого разностороннейшего изображения жизни» (*Успенский Г.* – ПСС. – Т. 10. – Кн. 2. – М.: Изд-во Академии наук СССР, 1954. – С. 155–156).

1.2
ИСТОРИЯ СТАТИСТИКИ
(краткий обзор)

Развитие статистики похоже на развитие языка и счета. Эта наука имеет древние корни. Она зародилась как результат обобщения уже достаточно развитой статистической практики, вызванной потребностями общества.

Вот лишь некоторые сведения. В *Китае* более чем за две тысячи лет до нашей эры проводились исчисления населения по полу и возрасту, а также собирались сведения о состоянии промышленности и сельского хозяйства. Упоминания о статистических обследованиях встречаются и в библейские времена. В *Древнем Риме* велась статистика численности населения и имущественного положения граждан.

Развитие торговых и международных товарно-денежных отношений явилось стимулом для дальнейшего формирования учета и статистики.

В конце IX в. проводились первые учетные операции: инвентаризация королевских имений, учет населения, пригодного к военной службе. Первыми и основными учетно-статистическими источниками на *Руси* были летописи, в которых уже в IX–XI вв. упоминается о сборе различной информации. Так, приводятся учетные данные о возникновении и развитии городских поселений, расположенных на водных путях, о наличии в них храмов, церквей, монастырей, жилых строений. Однако сбор числовых данных в государствах древнего мира был настолько несовершенным, что говорить о научном подходе к нему не приходится. В тот период статистические операции, как правило, проводились в исключительных случаях, в основном в военных и финансовых целях. Позднее потребность в статистических операциях возникла уже с необходимостью стимулировать рост народонаселения, производительные силы страны, регулировать потребление.

Однако если сбор статистических данных начался в глубокой древности, то обработка и анализ данных, т.е. зарождение статистики как науки, относятся к более позднему периоду.

Во второй половине XVII столетия *в Германии* возникла ***школа государствоведения***. Ее основателем был немецкий ученый Г. Конринг (1606–1681). Дальнейшее развитие это направление получило в работах *Г. Ахенваля и А. Шлецера*. Так, Г. Ахенваль (1719–1772) в 1746 г. впервые в Марбургском, а затем в Геттингенском университете стал читать новую учебную дисциплину, которую он назвал *статистикой*. Школа просуществовала более 150 лет, не меняя свои теоретические основы. Предмет и метод этой науки не были четко определены. В основном собирался описательно-информационный материал, который впоследствии почти не анализировался. Как правило, описывался последний период, иначе, по мнению представителей школы, в статистической работе не было смысла.

В трудах сторонников этого направления содержалось описание государств, в том числе их устройства, быта и нравов населения, естественных условий, климата, финансов, армии. Авторы трудов называли это описание статистическим. С нынешних позиций с этим нельзя согласиться, поскольку информация представляла собой словесное описание «достопримечательностей государства». В основном эти описания носили этнографический характер. Содержание, задачи, предмет изучения статистики в понимании ученых были далеки от современного взгляда на статистику как науку.

Гораздо ближе к современному пониманию статистики была английская ***школа политических арифметиков***. Она возникла на 100 лет раньше немецкой описательной школы. Основоположниками школы политических арифметиков были *Д. Граунт* (1620–1674), *Э. Галлей* (1656–1742) и *В. Петти* (1623–1687). В их трудах наметились два направления: демографическое с уклоном к вопросам страхования жизни у Д. Граунта и Э. Галлея и статистико-экономическое у В. Петти. На основе обработки бюллетеней о естественном движении населения Лондона Д. Граунт впервые открыл некоторые закономерности массовых общественных явлений и показал, как следует обрабатывать и анализировать массовый первичный материал. Он впервые попытался построить таблицу смертности для стационарного населения. Теоретическую разработку проблемы смертности, начатую Д. Граунтом, продолжил Э. Галлей. Знаменитый английский астроном высказал идею закона больших чисел и применил методы устранения случайных отклонений.

В. Петти, друг и современник Д. Граунта, посвятил статистике ряд научных работ. Главным в них было стремление конкретно оценить то или иное явление, несмотря на нехватку числовых данных. *В. Петти является фактическим родоначальником экономической статистики*.

Политические арифметики путем обобщения и анализа фактов стремились цифрами охарактеризовать состояние и развитие общества, вскрыть закономерности развития общественных явлений, проявляющиеся в массовых данных. Цели и задачи, которые ставили перед собой ученые, близки к современному пониманию сущности статистики. Идеи Д. Граунта, Э. Галлея и В. Петти имели своих последователей не только на их родине, но и в других европейских государствах. Наибольшее развитие эта школа получила в XVII и XVIII вв. в *Англии*, *Голландии*, *Франции*. В дальнейшем школа политических арифметиков сосредоточила внимание в основном на демографии, а особенно на вопросах страхования жизни, финансовых расчетах.

В первой половине XIX в. возникло третье направление статистической науки. Оно получило название ***статистико-математическое.*** Особый вклад в развитие этого направления внес бельгийский статистик *А. Кетле* (1796–1874). Он назвал статистику социальной физикой, т.е. наукой, изучающей законы общественной системы с помощью количественных методов. Важнейшей его заслугой стало обоснование идеи использования закономерностей, выявленных из массы случаев, в качестве важнейшего инструмента познания объективного мира. Учение А. Кетле о статистической закономерности оказало значительное влияние на современников. Многие из его последователей пошли дальше своего учителя.

Значительный вклад в развитие статистики внесли английские ученые – *Ф. Гальтон* (1822–1911) и *К. Пирсон* (1857–1936). Ф. Гальтон, родственник Чарлза Дарвина, серьезно заинтересовался проблемой наследственности, к анализу которой он вскоре применил статистические методы. Кроме всего прочего, им было разработано использование понятия перцентиля. Пирсон также провел много плодотворных исследований в статистике. И Гальтон, и Пирсон внесли значительный вклад в развитие теории корреляции.

Наиболее известным ученым XX в. в области статистики на Западе является *Р. Фишер* (1890–1962). Он весьма продуктивно работал в течение полувека. Многие его исследования оказали существенное воздействие на современную статистику.

В российской статистике не было четкого обособления школ и направлений, и тем не менее можно отметить русскую описательную школу, русскую школу политических арифметиков, статистическую мысль революционеров-демократов русской социологической школы, различные направления в русской академической статистике.

Яркими представителями ***русской описательной школы*** являются *И.К. Кириллов* (1689–1737), *В.Н. Татищев* (1686–1750), *М.В. Ломоносов* (1711–1765), *И.И. Голиков* (1735–1801), *С.Н. Плещеев* (1752–1802), *М.И. Чулков* (1740–1793) и др. Собранные ими материалы стали источником сведений по экономической теории России с древних времен до XVIII в.

Превращению статистики из науки описательной в науку теоретическую, формированию статистики как науки способствовали представители школы политических арифметиков, которые изучали общественные явления с использованием меры, веса и числа.Основными представителями этого направления русской статистики были *Д. Бернулли* (1700–1782), *И.Ф. Герман* (1755–1815) и др.

Уже в начале XIX в. статистика нуждалась в уточнении организационных и методологических основ, что было вызвано изменениями в системе государственного управления и распространением прогрессивно-демократических идей. В этот период выходит ряд крупных работ по теории статистики. В книге «Всеобщая теория статистики. Для обучающих сей науке» *К.Ф. Герман* (1767–1838) изложил основные положения, раскрывающие статистику как науку. В истории развития статистики большое значение имеют работы *К.И. Арсеньева* (1789–1856), в которых он утверждал, что статистика в состоянии дать адекватную характеристику жизни государства.

Наиболее прогрессивные для этого времени теоретические основы статистики как самостоятельной науки были созданы *Д.П. Журавским* (1810–1856). Ему принадлежит системное изложение основ теоретической базы статистики как науки, определение статистической науки. Он уделил большое внимание проблеме достоверности данных, методу группировок, раскрыл принцип единства количественного и качественного анализа.

Большое влияние на развитие русской статистической мысли оказали русские демократы-революционеры: *А.Н. Радищев* (1749–1802), *А.И. Герцен* (1812–1870), *Н.П. Огарев* (1813–1874). Эти выдающиеся деятели внесли определенный вклад в теорию и практику статистики. Ими разработаны программные вопросы экономической и судебной статистики, сделаны попытки определить средние величины, сформулированы проблемы о значении метода группировок.

Свою роль в истории статистики сыграли представители ***академической школы статистики,*** характерной особенностью которой было стремление заменить изучение государства изучением общества. Основоположниками этой школы были *Э.Ю. Янсон* (1835–1893), *А.И. Чупров* (1842–1908), *А.А.Чупров* (1874–1926), *Н.А. Каблуков* (1849–1919) и *А.А. Кауфман* (1864–1919). Академическая статистика и ее представители оказали большое положительное влияние на развитие статистической науки в России и на работу статистических органов. К началу XX в. Россия была одним из признанных центров научной статистической мысли. Большое влияние на развитие математического направления в статистике России произвели работы русских математиков *П.П. Чебышева* (1821–1894), *А.А. Маркова* (1856–1922), *А.М. Ляпунова* (1857–1919).

Исторический опыт советской статистики как науки был обобщен в трудах *В.И. Хотимского* (1892–1937), *В.С. Немчинова* (1894–1964), *В.Н. Старовского* (1905–1975), *А.Я. Боярского* (1906–1985), *Б.С. Ястремского* (1877–1962), *Л.В. Некраша* (1886–1949) и других ученых. В послевоенный период внимание статистической науки было приковано к вопросу о предмете статистики, ее соотношении с математической статистикой. В 1954 г. этот вопрос обсуждался на научном совещании, которое еще раз подтвердило значение статистики как самостоятельной общественной науки. После совещания вышли в свет новые монографии, учебники по общей теории статистики. В это время значительный вклад в теорию индексного метода был внесен учеными *С.М. Югенбергом, Г.И. Баклановым, Л.С. Казинцом, В.Е. Адамовым* и др. Заслуживают серьезного внимания труды по изучению статистической связи *Я.И. Лукомского.* Шагом вперед к развитию статистической науки послужило комплексное применение, наряду со статистическими, экономико-математических методов и широкое использование компьютерной техники в анализе социально-экономических явлений.

В настоящее время ведется работа по совершенствованию статистической методологии и переходу Российской Федерации на принятую в международной практике систему учета и статистики в соответствии с требованиями развития рыночной экономики.

1.3 ОСНОВНЫЕ ЧЕРТЫ ПРЕДМЕТА СТАТИСТИЧЕСКОЙ НАУКИ

Любая наука обладает существенными специфическими особенностями, которые отличаются от других наук и дают ей право на самостоятельное существование как особой отрасли знания. Рассмотрим основные черты и особенности предмета статистической науки.

Первая особенность статистики как науки заключается в исследовании ею не отдельных фактов, *а массовых социально-экономических явлений и процессов,* выступающих как множества отдельных фактов, обладающих как индивидуальными, так и общими признаками.

Задача статистического исследования состоит в получении обобщающих показателей и *выявлении закономерностей общественной жизни в конкретных условиях места и времени,* которые проявляются лишь в большой массе явлений через преодоление случайности, свойственной единичным элементам. Общественная жизнь столь сложна и многообразна, что почти всегда можно подобрать факты, примеры, как подтверждающие, так и опровергающие одно и то же положение. Чтобы охарактеризовать массовое общественное явление или процесс в целом, необходимо рассмотреть всю или очень большую массу относящихся к ним отдельных явлений или процессов.

Объект статистического исследования (в каждом конкретном случае) называют статистической совокупностью. *Статистическая совокупность – это множество единиц, обладающих массовостью, однородностью, определенной целостностью, взаимозависимостью состояний отдельных единиц и наличием вариации.* Например, в качестве особых объектов статистического исследования, т.е. статистических совокупностей, могут выступать множества – сельскохозяйственные предприятия, семьи, браки, студенты, граждане какой-либо страны. Важно помнить, что статистическая совокупность состоит из реально существующих материальных объектов.

Каждый отдельно взятый элемент данного множества называется *единицей статистической совокупности.* Единицы статистической совокупности характеризуются общими свойствами, именуемыми в статистике признаками, т.е. под качественной однородностью совокупности понимается сходство единиц (объектов, явлений, процессов) по каким-либо существенным признакам, но различие по каким-либо другим признакам. Например, из названных совокупностей множество сельскохозяйственных предприятий наряду с качественной определенностью (принадлежность к разряду предприятий, причем определенной отрасли – сельскому хозяйству) обладает различиями по размеру земельных угодий, численности работающих, голов скота, по технологической оснащенности и т.д.

Качественная определенность совокупности хотя и имеет объективную основу, устанавливается в каждом конкретном статистическом исследовании в соответствии с его целями и познавательными задачами. Названная выше совокупность сельскохозяйственных предприятий, качественно однородная с точки зрения одной задачи исследования, может оказаться качественно неоднородной с точки зрения другой задачи, например изучения дифференциации хозяйственных условий и результатов в зависимости от категорий хозяйств по формам собственности.

Таким образом, единицы совокупности наряду с общими для всех единиц признаками, обусловливающими качественную определенность совокупности, обладают индивидуальными особенностями и различиями, отличающими их друг от друга, т.е. существует так называемая *вариация признаков.* Вариация этих признаков обусловлена различным сочетанием условий, составляющих развитие элементов множества. Например, уровень производительности труда отдельного рабочего определяется его возрастом, квалификацией, отношением к труду и т.д. Именно наличие вариации предопределяет необходимость статистики.

Вариация признака, к примеру, может быть отражена следующим статистическим распределением (табл. 1.2). Данные показывают, что 50,1% безработных приходится на самый активный трудоспособный возраст 30–49 лет, т.е. средний возраст безработного – 35,1 года, и на людей, имеющих профессиональное и среднее общее образование, – 72,1%.

Таблица 1.2

Распределение численности безработных по полу, возрасту и образованию в РФ в 2000 г., %

Группы безработных по возрасту, лет	Всего	В том числе распределение по полу	
		мужчины	женщины
Всего безработных	100,0	100,0	100,0
в том числе в возрасте:			
До 20	7,0	7,0	6,9
20–24	17,3	17,7	16,8
25–29	13,4	14,1	12,6
30–49	50,1	49,6	50,6
50–54	6,1	5,1	7,4
55–59	3,4	3,7	3,1
60–72	2,7	2,8	2,6
Средний возраст, лет	35,1	34,7	35,6
в том числе имеют образование:			
высшее профессиональное	11,1	8,6	14,1
неполное высшее профессиональное	4,1	3,8	4,4
среднее профессиональное	22,5	19,3	26,2
среднее (полное) общее	45,8	47,6	43,7
основное общее	14,1	17,2	10,4
начальное общее, не имеют (начального) общего образования	2,4	3,5	1,2

Статистические распределения имеют большое практическое и научное значение. Казалось бы, например, что распределение жителей по росту, окружности головы, длине стопы и другим физическим признакам интересует лишь антропологию. Однако без знания этих распределений невозможна успешная работа предприятий, производящих одежду, обувь, головные уборы и т.п.

Заметим, что характер распределения и его параметры зависят от конкретных условий места и времени.

Наряду с распределением массовых общественных явлений статистика устанавливает их размещение в пространстве. Размещение массовых явлений в пространстве требует тщательного анализа. Примером такого размещения является распределение беженцев и вынужденных переселенцев, получивших официальный статус в органах Федеральной миграционной службы России (табл. 1.3).

Таблица 1.3

Число вынужденных переселенцев и беженцев на 1 января 2001 г.

Регион проживания	Человек
Всего в Российской Федерации	808280
из них ранее проживали на территории:	
Азербайджана	37478
Армении	2696
Беларуси	174
Грузии	60067
Казахстана	292445
Киргизии	35819
Латвии	11927
Литвы	1922
Молдавии	9609
России	140657
Таджикистана	85101
Туркмении	12837
Узбекистана	102125
Украины	6633
Эстонии	9064
Территория не указана	726

Вторая особенность статистики как науки в том, что она изучает прежде всего *количественную сторону общественных явлений и процессов в конкретных условиях места и времени, т.е. предметом статистики являются размеры и количественные соотношения социально-экономических явлений, закономерности их связи и развития.* Например, статистика изучает экономические характеристики производства, распределения и потребления, уровень материального благосостояния населения, явления культурной жизни, численность населения земного шара, его распределение по континентам и странам и т.д.

Другим выражением количественной стороны общественной жизни являются ***числовые соотношения*** размеров общественных явлений. Например, в 2000 г. на 1000 человек населения в РФ приходилось 6,2 брака и 4,3 развода. В 2000 г. в нашей стране в расчете на душу населения за год потреблено мяса и мясопродуктов 41 кг, яиц 229 шт., рыбы и рыбопродуктов 10,4 кг. Валовой внутренний продукт РФ в 2000 г. по сравнению с 1999 г. составил 108,3%.

Количественная определенность – это объективное свойство предмета познания статистикой. Количественные характеристики, устанавливаемые статистикой, не являются зафиксированными раз и навсегда, одинаковыми для всех единиц совокупности. В каждый момент общественной жизни явления обладают определенными уровнями независимо от того, находят они свое выражение в данных статистики или нет. Они меняются (варьируют) от одной единицы совокупности к другой в пространстве и времени. Итак, как было уже сказано, статистику к жизни вызывает вариация явлений.

Количественную характеристику статистика выражает через определенного рода числа, которые называются статистическими показателями. *Статистический показатель отражает результат измерения единиц совокупности и совокупности в целом.*

Однако чем же тогда статистика отличается от математики?

Основная особенность статистики, по нашему мнению, состоит в том, что *она изучает количественную сторону качественно определенных массовых общественных явлений в данных условиях места и времени. При этом качественную определенность единичных явлений обычно определяют сопряженные науки.*

Так, например, статистика изучает смертность как массовое явление, а факт и причину смерти устанавливает медицина.

Если для арифметики конкретное содержание количества не имеет значения (всегда $4 \cdot 4 = 16$), то для статистики цифры без качественного содержания, а также без обстоятельства места и времени лишены всякого смысла. Назовем цифру 10,5 млн чел. – это численность населения Москвы (качественная определенность). Однако число жителей в городе постоянно меняется, следовательно, необходимо назвать и обстоятельство времени – 9 октября 2002 г.

Как статистика отражает в показателях качественную сторону процесса, видно также из данных табл. 1.4.

Таблица 1.4

Распределение численности занятых в экономике по формам собственности за 1990–2000 гг.

Распределение занятого населения по формам собственности	Млн чел.		% к итогу	
	1990	2000	1990	2000
Всего в экономике	75,3	64,3	100,0	100,0
в том числе по формам собственности:				
государственная, муниципальная	62,2	24,4	82,6	37,9
частная	9,4	29,7	12,5	46,1
собственность общественных и религиозных организаций (объединений)	0,6	0,5	0,8	0,8
смешанная российская	3,0	8,0	4,0	12,5
иностранная, совместная российская и иностранная	0,1	1,7	0,1	2,7

Данные табл. 1.4 характеризуют качественные изменения, которые произошли за рассматриваемый период в сфере экономики в России. За 1990–2000 гг. появились предприятия негосударственных форм собственности, увеличился удельный вес работников, занятых на этих предприятиях.

Таким образом, в статистике нет просто цифр 200, 300 и так далее, в этой науке числа всегда *именованные,* относящиеся к определенному месту и времени.

Статистический показатель имеет *три обязательных атрибута: количественную определенность, место и время* (момент или период времени). Измеряя объем статистической совокупности или ее частей, статистик получает объемные показатели.

Количественная характеристика уровня, степени распространения явления, его динамика измеряются качественными показателями, например уровнем производительности труда, себестоимостью продукции, количеством браков на 1000 жителей, темпом роста населения и т.д.

Такое свойство совокупности, как *типичность,* измеряется средними показателями.

Совокупность показателей всесторонне отражает сложное явление и составляет систему показателей. Однако статистику нельзя назвать только наукой о размерах социально-экономических явлений.

Третья особенность статистики как науки заключается в том, что она характеризует структуру общественных явлений. *Структура – это внутреннее строение массовых явлений, т.е. внутреннее строение статистического множества.* Статистика должна эту структуру обнаружить, выразить и отразить с помощью статистических показателей.

При анализе структуры выявляются составные части социально-экономических явлений. Эти составные части сопоставляются с явлением в целом и между собой. Производится сравнение данной структуры с другими однотипными структурами, а также с заданной (плановой, нормативной и т.п.) и выявление причин отклонений. Подготавливаются предложения по оптимизации структуры. В процессе ее анализа используется метод группировок.

Признаки структуры многообразны, и задача статистики – выбрать наиболее существенные и важные признаки, отражающие структуру социально-экономических явлений. Выбор системы признаков предопределяется задачами, стоящими на данный момент, и зависит от условий места и времени.

Например, применительно к статистике населения *первое направление исследований* составляет изучение численности населения и структуры по полу, возрасту, национальности, образованию, роду занятий, источнику средств существования и т.д.; *второе направление* включает изучение зависимости одних структурных характеристик (например, числа детей в семье) от других (уровень дохода и образования, жилищные условия и т.п.); *третье* – это изучение прироста населения и изменение его структуры.

Таким образом, каждому общественному явлению свойственны изменения в пространстве и времени. Изменения в пространстве, т.е. *в статике,* выявляются анализом структуры общественного явления, а изменения во времени, т.е. *в динамике,* – исследованием уровня и структуры явления. Такова ***четвертая особенность статистики как науки***.

Анализ динамики включает:

- установление уровня общественного явления на определенный момент или промежуток времени и определение среднего уровня;
- выявление характера изменений за каждый промежуток времени и в целом;
- определение величины и темпов изменения;
- установление основной тенденции изменений, их закономерностей и составление статистического прогноза.

Рассмотрим для примера данные о составе осужденных на территории Российской Федерации (табл. 1.5). Только статистика может так наглядно отразить изменение социально-экономического явления во времени.

Таблица 1.5

Состав осужденных по возрасту на территории России за 1990–2000 гг., %

Группы осужденных по возрасту, лет	1990	1995	2000
Всего осуждено	100,0	100,0	100,0
из них в возрасте:			
14–17	14,7	11,2	12,6
18–24	22,9	24,0	31,3
25–29	20,3	15,6	16,3
30–49	36,4	43,3	34,9
50 и старше	5,7	5,9	4,9

Явления общественной жизни взаимосвязаны и взаимообусловлены: изменение одних явлений предопределяет другие; например, снижение затрат на сырье и материалы приводит к снижению себестоимости, и наоборот. Поэтому выявление связей является ***пятой особенностью*** статистики как науки, так как познание действительности невозможно без познания всех или по крайней мере основных взаимосвязей общественных явлений. Наибольшее значение имеет выявление причинно-следственных связей, чтобы воздействовать на общественные явления с целью их изменения в интересах общества.

С помощью специальной методологии статистика определяет количественные связи между общественными явлениями.

Учитывая вышесказанное, сформулируем определение статистики как науки.

Статистика – общественная наука, которая изучает количественную сторону качественно определенных массовых социально-экономических явлений и процессов, их структуру и распределение, размещение в пространстве, движение во времени, выявляет действующие количественные зависимости, тенденции и закономерности в конкретных условиях места и времени.

Исходя из характера и основных черт предмета определим следующие *познавательные задачи статистики как науки.* Это изучение следующих характеристик:

- уровня и структуры массовых социально-экономических явлений;
- взаимосвязи массовых социально-экономических явлений и процессов;
- динамики массовых социально-экономических явлений.

Таким образом, цель статистического исследования, как и любого научного исследования, – раскрыть сущность массовых явлений и процессов, а также присущие им закономерности. Отличительная особенность этих закономерностей в том, что они относятся не к каждой отдельной единице совокупности, а ко всей массе единиц в целом. Общим принципом, лежащим в основе исследования статистических закономерностей, выступает так называемый *закон больших чисел* (ЗБЧ).

1.4 ТЕОРЕТИЧЕСКИЕ ОСНОВЫ СТАТИСТИКИ КАК НАУКИ

Теоретическую основу любой науки, в том числе и статистики, составляют понятия и категории, в совокупности которых выражаются основные принципы данной науки. В статистике к важнейшим категориям и понятиям относятся: совокупность, вариация, признак, закономерность.

В разделе 1.3 было дано общее представление об основных категориях статистической науки. Рассмотрим подробнее эти категории.

Статистическая совокупность – это множество (масса) однокачественных (однородных) хотя бы по одному какому-либо признаку явлений, существование которых ограничено в пространстве и времени. Статистической совокупностью можно считать, к примеру, совокупность жителей России по состоянию на 1 января 2002 г., совокупность фермерских хозяйств Ростовской области в 2002 г., совокупность студентов III курса МГУ в 2001/02 учебном году и т.п. Однако статистическая совокупность (множество) совсем необязательно представляет большую численность единиц, в принципе совокупность может быть и очень маленькой; например, объем совокупности малой выборки может составлять иногда 8–10 единиц.

От реально существующих статистических совокупностей следует отличать стохастические совокупности, или гипотетические множества, т.е. совокупности, предполагаемые мысленно нереальными, например совокупность бесконечно большого числа бросаний монеты, падающей либо орлом, либо решкой.

Самостоятельное значение имеют *совокупности социально-экономических явлений*. Они представляют собой отдельные грани общественных процессов, которые более сложны и разнородны, чем природные явления, и менее многочисленны, т.е. объединяют значительно меньшее число единиц.

Важнейшим свойством статистической совокупности является ее *неразложимость*. Это означает, что дальнейшее дробление индивидуальных явлений не изменяет их качественной основы. Исчезновение или ликвидация одного или ряда явлений не разрушает качественной основы статистической совокупности, так как все характеристики относятся к совокупности в целом. Так, население страны или города останется населением, несмотря на постоянно происходящие процессы механического и естественного движения населения.

Существует *понятие однородности статистической совокупности*. Оно относительно и вовсе не означает полного соответствия друг другу всех единиц совокупности, а лишь подразумевает наличие для всех единиц совокупности основного свойства, качества, типичности. Одна и та же совокупность единиц, к примеру, может быть однородна по одному признаку и неоднородна по другому. Однородность единиц статистической совокупности формируется под воздействием определенных внутренних причин и условий. Одинаковые для всех единиц данной совокупности причины и условия существования создают то общее, что объединяет единицы совокупности, но эти же причины и условия формируют то, что отличает одну единицу совокупности от другой. В статистической совокупности эти отличия чаще имеют количественную природу. Количественные изменения значений признака при переходе от одной единицы совокупности к другой называются *вариацией*. Вариация возникает под воздействием случайных, прежде всего социально-экономических, причин. Социально-экономические явления обладают большой вариацией. Например, вариация городов страны по численности населения складывается под влиянием большого числа факторов – исторических, этнографических, экономических, социальных и множества других.

Статистические совокупности имеют определенные свойства, носителями которых выступают *единицы* (*отдельные элементы*) *совокупности,* обладающие определенными *признаками*. По *форме внешнего выражения* признаки делятся на:

- атрибутивные (описательные, качественные);
- количественные.

Атрибутивные (качественные) *признаки* не поддаются прямому количественному (числовому) выражению. Отличие количественных признаков от качественных состоит в том, что первые можно выразить итоговыми значениями (табл.1.2), вторые – только числом единиц в совокупности (табл. 1.6). *Количественные признаки* делятся на дискретные (прерывные) и непрерывные. Все разнообразие признаков можно представить в табл. 1.7.

Таблица 1.6

Распределение театров в РФ по их направлению за 1980–2000 гг. (на конец года)

Группы театров по направлению	1980	2000
Всего профессиональных театров	324	547
в том числе:		
оперы и балета	22	65
драмы, комедии и музыкальных	199	318
детских и юного зрителя	103	151

Таблица 1.7

Классификация признаков по их видам

Код классификации	Признаки
По содержательности	Существенные, несущественные, первичные, вторичные
По принадлежности	Индивидуальные, общие
По направлению	Прямые, косвенные

Продолжение

Код классификации	Признаки
По причинности	Причины, следствия, факторные, результативные
По управляемости	Управляемые, неуправляемые
По степени детерминированности	Детерминированные, статистические, стохастические
По наблюдаемости	Наблюдаемые, ненаблюдаемые
По измеряемости	Непосредственно измеряемые, косвенно измеряемые, условно измеряемые
По времени	Статические, динамические, периодические

Важнейшей категорией статистики является статистическая закономерность. Под *закономерностью* вообще принято называть *повторяемость, последовательность и порядок изменений в явлениях.* Статистическая же закономерность в статистике рассматривается как количественная закономерность изменения в пространстве и времени массовых явлений и процессов общественной жизни, состоящих из множества элементов (единиц совокупности). Закономерность свойственна не отдельным единицам совокупности, а всей их массе, или совокупности в целом. В силу этого закономерность, присущая данному явлению (процессу), проявляется только при достаточно большом числе наблюдений и только в среднем. Таким образом, статистическая закономерность – это закономерность усредненных параметров некоторого основного свойства (качества или типичности).

Первые предположения, что статистика познает закономерности общественной жизни, были высказаны в середине XVII в Д. Граунтом и В. Петти при исследовании бюллетеней о естественном движении населения Лондона.

Статистическая закономерность – это форма проявления причинной связи, выражающаяся в последовательности, регулярности, повторяемости событий с достаточно высокой степенью вероятности, если причины (условия), порождающие события, не изменяются или изменяются незначительно. Статистические закономерности устанавливаются на основе анализа массовых данных. Они неприложимы к отдельным явлениям, как это возможно в естественных науках (биологии, механике, физике). Данные закономерности возникают как результат воздействия большого числа постоянно действующих причин и причин случайных, действующих временами. Постоянно действующие причины придают изменениям в явлениях регулярность, повторяемость; случайные – вызывают отклонения в этой регулярности.

Статистические закономерности представляют собой не что иное, как статистические факты, и выраженные в виде обобщающих статистических показателей дают исследователю неоценимые типизированные величины, которые чаще всего лишены конкретности. Но известно, что любое общее понятие является абстрактным и поэтому лишено конкретности: оно содержит в себе существенные признаки класса предметов и не включает их несущественные, единичные, индивидуальные свойства.

Таким образом, статистическая закономерность предопределяет типичное распределение единиц статистического множества на определенный момент времени под воздействием всей совокупности факторов. Статистическая закономерность, не определяя положение каждого случая, устанавливает общее распределение в данных условиях времени и места. Сила статистики в том, что она дает нам общую картину, тенденцию развития, исключая, «нивелируя» случайные, индивидуальные отклонения и колебания. Без статистики мы бы «утонули» в море единичных случайных колебаний и отклонений, в «неразберихе» отдельных процессов.

Статистическая закономерность – объективная количественная закономерность массового процесса. Она возникает в результате действия объективных законов и выражает каузальные отношения.

Любое заметное изменение условий существования данного множества воздействует на статистическую закономерность. В этом смысле закономерность является своего рода лакмусовой бумажкой при проверке на постоянство факторов.

Статистическая закономерность практически гарантирует сравнительно малую вероятность больших отклонений фактических частот от теоретических. Например, в магазинах имеется ассортимент продуктов или товаров, соответствующий среднему спросу, с резервным запасом, обеспечивающим его возможные колебания в нормальных условиях. Относительный размер резервного запаса уменьшается с ростом числа покупателей. Статистическая закономерность гарантирует устойчивость средних величин при сохранении постоянного комплекса условий, порождающих данное явление.

Так как статистическая закономерность обнаруживается в итоге массового статистического наблюдения, это обусловливает ее взаимосвязь с законом больших чисел. *Закон больших чисел* в наиболее простой формулировке гласит, что количественные закономерности массовых явлений отчетливо проявляются лишь в достаточно большом их числе. Например, 104–106 мальчиков рождаются на 100 девочек, однако в отдельной семье, и даже в небольшом населенном пункте, это соотношение может быть совершенно иным.

Характеризуя роль закона больших чисел и его значение для статистики, А. Боярский отмечает, что в средней величине момент случайности оказывается снятым. Следовательно, закон больших чисел по форме говорит о том же, о чем и статистика: результат складывается из многих элементов. Закон больших чисел позволяет определить число необходимых наблюдений за массовым процессом, а также и некоторые условные критерии для оценки однородности явлений. Закон характеризует достоверность полученных результатов и необходим для понимания ошибок массового статистического наблюдения. Однако закон больших чисел не определяет собой постановку наблюдения и исследования в целом. Последняя диктуется необходимостью точного и всестороннего учета существенных факторов, влияющих на результат исследования. Закон больших чисел выражает диалектику случайного и необходимого, единичного, особенного и всеобщего. В результате взаимопогашения случайных отклонений средние величины, исчисленные для величин одного и того же вида, становятся типичными, отражающими действие постоянных и существенных факторов в данных условиях места и времени.

В соответствии с природой массовой закономерности тенденции, вскрытые с помощью закона больших чисел, *имеют силу лишь как массовые тенденции, но не как тенденции, действительные для каждого отдельного факта без исключения.*

Однако анализ социально-экономических явлений нужно проводить с учетом существенных сторон единичного, т.е. отдельного факта. О единичном в статистическом анализе нельзя забывать потому, что единичное хотя и теряет свое значение в общей массе, но продолжает существовать.

В ряде случаев статистика может и должна изучать явления, которые еще не стали массовыми, но которым принадлежит будущее, и они станут массовыми. Так, например, российская статистика изучала кооперативное движение на самой заре его развития, и в настоящее время малый бизнес принял массовый характер.

1.5
МЕТОД СТАТИСТИЧЕСКОЙ НАУКИ
(статистическая методология)

Методологической основой статистики являются положения социально-экономической теории и принципы диалектического метода познания. Они составляют теоретическую базу статистики.

Опираясь на знание положений экономической теории, статистика анализирует конкретные формы проявления категорий, оценивает размеры явлений, осуществляет разработку адекватных методов их изучения и анализа.

Одновременно, руководствуясь принципами диалектического метода познания, статистика изучает все явления в их взаимосвязи, в движении и изменении, выявляет их различные типы и формы, устанавливает то новое, прогрессивное, что зарождается в существующем и определяет направления развития. В процессе развития в изучаемом явлении происходят наряду с количественными изменениями и коренные качественные изменения.

Таким образом, знание диалектических законов и категорий позволяет статистике правильно понять и истолковать явления, подлежащие статистическому исследованию, выбрать надлежащий инструмент и методологически правильный подход к их изучению. Вместе с тем статистика, опираясь на теоретическую базу, имеет свои специфические приемы изучения, зависящие от особенностей ее предмета. Совокупность приемов, с помощью которых статистика изучает свой предмет, образует статистическую методологию. Таким образом, под ***статистической методологией*** понимается система приемов, способов и методов, направленных на изучение количественных закономерностей, проявляющихся в структуре, динамике и взаимосвязях социально-экономических явлений.

Статистическое исследование состоит из трех основных стадий:

- статистического наблюдения;
- первичной обработки, сводки и группировки результатов наблюдения;
- анализа полученных сводных материалов.

Все эти этапы связаны между собой; отсутствие одного из них ведет к разрыву целостности статистического исследования. Так, проведение статистического наблюдения бессмысленно без дальнейшего анализа, а анализ невозможен без информации, полученной на стадии первичной обработки данных.

Прохождение каждой стадии исследования связано с использованием специальных методов, объясняемых содержанием выполняемой работы.

Метод массовых наблюдений. *Первой стадией* статистических исследований является *статистическое наблюдение* – научно организованный сбор сведений об изучаемых социально-экономических процессах или явлениях.

Полученные в результате статистического наблюдения данные являются исходным материалом для выполнения последующих этапов статистического исследования. Характерным для этой стадии является *метод массовых наблюдений*. Это объясняется тем, что статистика изучает закономерности, которые выделяются через исследование многочисленных массовых явлений под действием закона больших чисел. Кроме того, на этом этапе формируются цели и задачи наблюдения; разрабатываются программы исследования в целом и по вышеуказанным стадиям; определяются конкретные способы и методы, используемые на каждом этапе исследования; составляется организационный план его проведения; определяются объект (совокупность общественных явлений или процессов) и единица наблюдения.

Результатом статистического наблюдения является получение данных, характеризующих каждую единицу наблюдения. Цель же исследования – получение характеристики объекта наблюдения в целом. Поэтому результаты статистического наблюдения представляют собой лишь исходный статистический материал, который, по выражению известного русского статистика А. Кауфмана, «относится к статистике приблизительно так же, как собранные на месте постройки запасы кирпича, балок, труб и иных строительных материалов относятся к будущему зданию, которое еще предстоит строить из этих материалов» (*Кауфман А.А.* Теория и методы статистики. – М.–Л.: Госиздат, 1928. – 174 с.).

Эти результаты необходимо определенным образом обработать с тем, чтобы из статистического «сырья» выявить статистические данные. Такая обработка является следующей после наблюдения стадией статистического исследования и представляет собой сводку исходных данных для получения обобщающих характеристик исследуемого процесса или явления, проводимую с помощью использования метода группировок и таблиц.

Метод статистических группировок и таблиц. *Вторая стадия* статистического исследования представляет собой комплекс последовательных действий по обобщению конкретных единичных фактов, образующих совокупность в целях выявления типичных черт и закономерностей, присущих изучаемому явлению в целом. Важнейшим специфическим методом на этой стадии является *метод группировок.* Статистическая сводка включает в себя распределение исходных данных по группам, качественно однородным по одному или нескольким признакам, и получение групповых итогов. На правильность выводов, получаемых в результате исследования, оказывает существенное влияние обоснованный выбор группировочных признаков.

Для правильного выделения качественно однородных групп следует выбирать основные, наиболее существенные для данного явления или процесса признаки. В зависимости от числа и вида признаков, решаемых задач и исходных данных группировки подразделяются на следующие виды: простые и комбинационные; по количественным и качественным признакам: типологические, структурные и аналитические; многомерные; первичные и вторичные.

Одним из этапов группировки является построение *рядов распределения,* т.е. группировка единиц наблюдения по величине или значению признака.

Результаты статистической группировки и сводки излагаются в виде *статистических таблиц,* являющихся наиболее рациональной, систематизированной, компактной и наглядной формой представления массовых данных.

Разновидностью табличных построений можно считать и различного рода матрицы абсолютных и относительных статистических показателей, построение которых связано с процессом компьютерной обработки информации.

Методы анализа с помощью обобщающих показателей. *Статистический анализ* является *заключительной стадией статистического исследования.*

В соответствии с ранее сформулированными познавательными задачами статистики как науки в процессе статистического анализа исследуются структура, динамика и взаимосвязи общественных явлений или процессов.

Характерным для статистических методов на этой стадии является применение *обобщающих показателей:* абсолютных, относительных, средних величин и индексных систем. Некоторые общие черты формирования обобщающих показателей устанавливаются посредством измерения их вариации. Изучение вариации наряду с применением средних и относительных величин имеет большое практическое и научное значение. Показатели вариации дополняют средние величины, за которыми скрываются индивидуальные различия. Они характеризуют степень однородности статистической совокупности по данному признаку. Показатели вариации определяют степень и границы вариации признака. Соотношение показателей вариации может выражать взаимосвязь признаков.

Изучение структуры сложных явлений – исходный пункт статистического исследования. Здесь вопросы изменения и развития явления возникают в отраженной форме с различным уровнем развития элементов структуры. Конечная задача статистического исследования структуры – анализ внутренних связей в объекте исследования. Характер этих связей более наглядно проявляется в динамике структурных изменений. Исследование динамики обычно носит дифференциальный или интегральный характер. Фиксация состояний процесса образует *интегральный динамический ряд,* который исследуется на основе обобщающих аналитических показателей, специальных приемов обработки и моделирования рядов динамики. Прогнозирование дальнейшего хода развития общественных явлений осуществляется с помощью экстраполяции.

Закономерности причинно-следственных связей общественных процессов и явлений устанавливаются с помощью методов математической статистики, в частности, *корреляционно-регрессионного* анализа, а также *методов многомерного статистического анализа*. Взаимосвязи явлений также изучаются и с помощью статистических методов: статистических группировок, сопоставления параллельных рядов, построения систем взаимосвязанных индексов и т.д.

Широкое применение в статистике находят графические методы, позволяющие в наглядной форме представлять результаты статистических исследований.

Большое значение для развития статистической методологии имеет компьютеризация статистических исследований, позволяющая создавать базы статистических данных и программы их обработки, в значительной мере сокращать сроки обработки информации, широко использовать многомерные методы, улучшать качество и наглядность проводимого анализа.

1.6 ОБЩАЯ ТЕОРИЯ СТАТИСТИКИ КАК ОТРАСЛЬ СТАТИСТИЧЕСКОЙ НАУКИ

Общество в процессе своего развития ставит перед статистикой все новые и новые задачи, что способствует выделению отдельных отраслей из единой статистической науки. Каждая из этих отраслей имеет свой объект исследования, выясняет сущность определенной системы показателей, разрабатывает правила и методы их получения и использования в научной и практической деятельности. Однако во всех отраслевых статистиках применяются принципы и методы общей теории статистики. На рис. 1.1 представлены три уровня статистики.

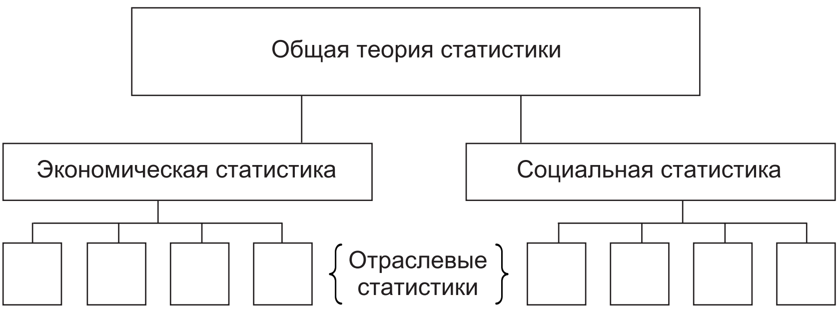

Рис. 1.1. Структура отраслей статистической науки

Первый уровень определяет общая теория статистики. *Общая теория статистики является наукой о наиболее общих принципах, правилах и законах цифрового освещения социально-экономических явлений.* Она разрабатывает наиболее общие понятия, категории статистической науки, которые имеют общестатистический смысл (например, закономерность, показатель, средняя величина, признак и т.д.), и методы изучения социально-экономических явлений.

Важнейшими разделами общей теории статистики являются учения о статистическом наблюдении, статистических группировках и обобщающих показателях в форме абсолютных, относительных и средних величин. Общая теория статистики выясняет сущность этих показателей и разрабатывает научную методологию их построения (вычисления), а также общие принципы и методы статистического исследования. Ее категориями, показателями и методами пользуются все отраслевые статистики, т.е. общая теория статистики является *методологической основой, ядром всей системы отраслевых статистик.*

Например, теория статистических группировок, средних, индексов разрабатывается общей теорией статистики, а отраслевые статистики применяют эти методы для исследования системы показателей своей отрасли и конкретного их исчисления.

Общая теория статистики изучает также исторический аспект изменения воззрений на статистическую науку, разрабатывает организацию статистических служб, изучает опыт организации статистики в других странах.

Таким образом, роль и место общей теории статистики особые. Она формирует общестатистическое мировоззрение у экономистов, менеджеров, руководителей предприятий независимо от места их работы.

Второй уровень занимают две большие обобщающие отрасли – экономическая статистика и социальная статистика.

Экономическая статистика разрабатывает и анализирует синтетические показатели, отражающие состояние национальной экономики, взаимосвязи отраслей, особенности размещения производительных сил, наличие материальных, трудовых и финансовых ресурсов, достигнутый уровень их использования.

Социальная статистика формирует систему показателей для характеристики образа жизни населения и различных аспектов социальных отношений.

Третий уровень определяют *отрасли экономической и социальной статистики.* В структуре экономической статистики выделяется макроэкономическая статистика, разрабатывающая методы комплексного изучения экономики страны, межотраслевые связи и др. Помимо этого, в совокупность отраслей экономической статистики входят статистика промышленности, сельского хозяйства, строительства и других отраслей производственной сферы. В состав социальной статистики входят статистика народонаселения, уровня жизни, культуры, общественного мнения, политическая, моральная и другие отрасли. Каждая отраслевая статистика формируется на базе показателей экономической или социальной статистики, а те и другие основываются на категориях (показателях) и методах анализа, разработанных общей теорией статистики, т.е. статистика развивается как единая наука, и развитие каждой отрасли содействует ее совершенствованию в целом.

Исходя из вышеизложенного отметим следующее: с точки зрения преподавания статистики в высшей школе в нее включили целый ряд дисциплин: теорию статистики, экономическую статистику, социальную статистику и целую серию отраслевых статистик: фирмы, ценообразования, рынка товаров и услуг, населения, демографическая, труда, здравоохранения, культуры и т.д.

Курс теории статистики открывает первый этап изучения цикла статистических дисциплин. Он является для отраслевых статистик основополагающей дисциплиной, которая закладывает фундамент для усвоения и конкретного применения статистических методов анализа социально-экономических явлений.

1.7 ОРГАНИЗАЦИЯ СОВРЕМЕННОЙ СИСТЕМЫ ГОСУДАРСТВЕННОЙ СТАТИСТИКИ В РОССИЙСКОЙ ФЕДЕРАЦИИ, ЕЕ ЗАДАЧИ И ФУНКЦИИ

Общие принципы организации государственной статистики в Российской Федерации. Преобразования в российской экономике, связанные с рыночными реформами, в значительной степени увеличивают потребности управляющих структур в своевременной и достоверной информации по различным аспектам жизнедеятельности общества. Необходимость наблюдения и анализа происходящих изменений экономической и социальной ситуации для принятия своевременных управленческих решений требует создания развитой современной инфраструктуры. Важная роль в этом управлении отводится федеральной системе государственной статистики, главными для которой являются следующие *принципы*: централизованное руководство статистикой, единое организационное строение и методология, неразрывная связь с органами государственного управления.

Основополагающим документом для принятия решения о дате образования государственной российской статистики является архивная справка директора Российского государственного исторического архива В.В. Лапина[1]. В этой справке сказано, что «…19 сентября 1802 г. …

[1] Архивная справка о дате и истории организации российской государственной статистики от 03.11.95 № 1232.

специальным циркуляром (предписано) губернаторам представлять в МВД сведения о численности населения, податях, произрастании хлебов, о сельских магазинах и народном продовольствии, о фабриках и заводах и о городских доходах и публичных зданиях». Высочайший манифест по этому поводу был подписан 8(20) сентября 1802 г. (Санкт-Петербургский журнал, 1804, № 1). Таким образом, указанную дату можно считать *началом статистической деятельности в России,* т.е. организацией систематического сбора статистической информации по министерствам, губерниям и ее обобщения по России в целом.

Обработка статистических данных была поручена «Обществу дворян» из 10 человек, но оно еще не было «упорядоченной государственной структурой» и «не смогло стать регулярным статистическим учреждением»[1].

О специальном учреждении, занимавшемся государственной статистикой в России, можно было официально говорить с 25 июня 1811 г., когда «государственная статистика, понимаемая тогда как статистика полицейская, была поручена… Министерству полиции… и было образовано *Статистическое отделение*»[2] .

В настоящее время роль государственной статистики трудно переоценить. Согласно статьи 71 Конституции Российской Федерации официальный статистический учет находится в ведении Российской Федерации. Государственный комитет Российской Федерации по статистике (Госкомстат России) является федеральным органом исполнительной власти, осуществляющим межотраслевую координацию и функциональное регулирование в сфере государственной статистики.

Госкомстат России в соответствии с законодательством Российской Федерации издает постановления, указы, распоряжения, инструкции и иные нормативные акты, в том числе совместно с другими федеральными органами исполнительной власти.

На Госкомстат России возложено как методологическое, так и практическое руководство по сбору, обработке и анализу статистических данных на государственном уровне. Методология статистических показателей, формы и методы сбора и обработки статистических данных, установленные Госкомстатом России, являются *официальными статистическими стандартами Российской Федерации.*

Статус Госкомстата России, осуществляющего межотраслевую координацию и функциональное регулирование в сфере государственной

---

[1] Архивная справка о дате и истории организации российской государственной статистики от 03.11.95 № 1232.

[2] Там же.

статистики, позволяет формировать официальную статистическую информацию на основе данных, получаемых как от хозяйствующих субъектов и граждан, так и от федеральных и территориальных органов государственной власти и местного самоуправления, и обеспечивать эффективное общение с ними как потребителями статистики.

В своей деятельности Госкомстат России руководствуется Конституцией Российской Федерации, федеральными конституционными законами, федеральными законами, указами и распоряжениями Президента Российской Федерации, постановлениями и распоряжениями Правительства Российской Федерации, международными договорами Российской Федерации, а также осуществляет свою деятельность во взаимодействии с другими федеральными органами исполнительной власти, органами исполнительной власти субъектов Российской Федерации, органами местного самоуправления, общественными объединениями и организациями.

Госкомстат России ежегодно разрабатывает и вносит в Правительство Российской Федерации программу статистических работ, которая формируется на основе предложений федеральных органов исполнительной власти, органов исполнительной власти субъектов Российской Федерации и других пользователей статистической информации. Работы, предусмотренные федеральной программой статистических работ и выполняемые дополнительно по поручениям Президента Российской Федерации, Федерального Собрания Российской Федерации и Правительства Российской Федерации, финансируются за счет средств федерального бюджета. Работы, не входящие в федеральную программу статистических работ, выполняются территориальными органами Комитета и находящимися в его ведении организациями за счет средств заказчика.

Органы государственной власти, органы местного самоуправления, организации, общественные объединения и граждане, осуществляющие предпринимательскую деятельность без образования юридического лица, ответственные за выполнение федеральной программы статистических работ, обязаны представлять документированную информацию в Госкомстат России и другие федеральные органы исполнительной власти.

Задачи и функции органов государственной статистики. Госкомстат России: в соответствии с Положением о Государственном комитете Российской Федерации по статистике, утвержденным постановлением Правительства Российской Федерации от 2 февраля 2001 г. № 85, осуществляет возложенные на него задачи и функции непосредственно, а также через свои территориальные органы, созданные в каждом субъекте Российской Федерации (в том числе и в муниципальных образованиях).

В соответствии с этим Положением *основными задачами* Госкомстата России являются:

- реализация государственной политики в области государственной статистики;
- разработка и совершенствование системы статистических показателей, характеризующих состояние экономики и социальной сферы;
- разработка научно обоснованной официальной статистической методологии, соответствующей международным стандартам, учитывающей особенности деятельности различных министерств и ведомств, отражающей новые явления, процессы и технологии в экономике, а также предназначенной для проведения государственных статистических наблюдений и формирования статистических показателей;
- координация деятельности федеральных органов исполнительной власти по формированию государственных информационных ресурсов в области статистики;
- разработка информационной системы государственной статистики, которая обеспечивала бы ее совместимость и взаимодействие с другими государственными информационными системами в едином информационном пространстве Российской Федерации;
- обеспечение хранения и защиты статистической информации;
- реализация обязательств России, вытекающих из членства в международных организациях и участия в международных договорах;
- осуществление международного сотрудничества в области статистики.

Госкомстат России в целях выполнения возложенных на него задач и в соответствии с предоставленным правом реализует следующие основные функции:

- утверждает формы государственных статистических наблюдений (государственной статистической отчетности), порядок их заполнения и представления;
- согласовывает формы ведомственных статистических наблюдений, проводимых федеральными органами исполнительной власти;
- запрашивает в установленном порядке информацию по формам государственной статистической и бухгалтерской отчетности у всех юридических лиц, их филиалов и представительств и граждан, занимающихся предпринимательской деятельностью без образования юридического лица;
- запрашивает у физических лиц информацию, необходимую для проведения государственных статистических наблюдений (указанная информация предоставляется на безвозмездной основе);

- запрашивает копии учредительных документов и различные сведения о юридических лицах и гражданах, занимающихся предпринимательской деятельностью без образования юридического лица, необходимые для ведения Единого государственного регистра предприятий и организаций, у органов, осуществляющих государственную регистрацию юридических лиц и указанных граждан, органов, осуществляющих лицензирование отдельных видов деятельности, иных органов исполнительной власти, а также государственных внебюджетных фондов;
- принимает нормативные правовые акты по вопросам государственной статистики, обязательные для выполнения федеральными органами исполнительной власти, органами исполнительной власти субъектов Российской Федерации и органами местного самоуправления, юридическими лицами, их филиалами и представительствами, гражданами, занимающимися предпринимательской деятельностью без образования юридического лица;
- осуществляет контроль за выполнением законодательства Российской Федерации в области государственной статистики всеми юридическими лицами и гражданами, осуществляющими предпринимательскую деятельность без образования юридического лица;
- применяет в соответствии с законодательством Российской Федерации меры административного взыскания за нарушение порядка представления статистической информации;
- создает территориальные органы комитета, штат численности работников и размер бюджетных ассоциаций на их содержание в порядке, установленном законодательством Российской Федерации и Правительством России;
- подготавливает и вносит предложения о создании, реорганизации и ликвидации организаций, находящихся в ведении комитета в порядке, установленном Российской Федерацией;
- заключает в установленном порядке со статистическими ведомствами других государств и международными статистическими и иными организациями соглашения о сотрудничестве по вопросам статистической деятельности;
- осуществляет приоритетные направления научных исследований в области статистики, привлекает в установленном порядке для этих целей научные учреждения Российской академии наук, другие российские и иностранные организации, а также ученых и специалистов;

- созывает совещания по вопросам, входящим в его компетенцию, с участием представителей органов государственной власти Российской Федерации, российских, иностранных и международных организаций.

Госкомстат России осуществляет возложенные на него задачи и функции непосредственно, а также через свои территориальные органы и находящиеся в ведении комитета организации. Комитет, его территориальные органы и находящиеся в ведении комитета организации составляют *федеральную систему государственной статистики.*

Структура Госкомстата Российской Федерации. Во главе Государственного комитета Российской Федерации по статистике (Госкомстат России) стоит председатель, назначенный на должность Правительством Российской Федерации. Председатель несет персональную ответственность за выполнение возложенных на комитет задач и осуществление им своих функций. Председатель комитета имеет заместителей, которые также назначаются Правительством Российской Федерации.

В целях обеспечения эффективной работы системы органов государственной статистики в Госкомстате России образована *коллегия* в составе председателя, заместителей и руководящих сотрудников комитета. Состав коллегии утверждается Правительством Российской Федерации, а секретарь коллегии назначается председателем комитета. Коллегия рассматривает важнейшие вопросы организации и методологии статистических работ, итоги основных статистических работ, а также вопросы, связанные с выполнением поручений Президента Российской Федерации, Правительства Российской Федерации, Администрации Президента Российской Федерации, Аппарата Правительства Российской Федерации, задач федеральных программ по развитию и реформированию системы государственной статистики Российской Федерации, кадровые и другие вопросы, требующие коллегиального решения.

На заседаниях коллегии присутствуют руководители структурных подразделений центрального аппарата комитета, подведомственных ему предприятий и организаций, руководитель юридической службы, а также приглашаются (по решению председателя коллегии) ответственные работники министерств и ведомств, представители СМИ.

Расширенные заседания коллегии проводятся с участием представителей территориальных органов Госкомстата России, заинтересованных министерств, ведомств, организаций, СМИ и международных организаций.

Для разработки и обсуждения методологических вопросов статистики, проблем межотраслевой координации деятельности в области статистики, учета потребностей пользователей в официальной статистике при Госкомстате работает *Статистический совет,* который является коллегиальным совещательным органом, утверждается Правительством Российской Федерации и состоит из числа ученых и высококвалифицированных специалистов системы Госкомстата России, представителей федеральных органов исполнительной власти, деловых кругов, научной общественности, средств массовой информации.

Статистический совет осуществляет свою работу по утвержденному коллегией Госкомстата России годовому плану, который разрабатывается с учетом предложений структурных подразделений Госкомстата России, членов Статистического совета.

В своей деятельности Статистический совет руководствуется Конституцией Российской Федерации, федеральными законами, указами и распоряжениями Правительства Российской Федерации.

Для содействия в выработке решений по совершенствованию деятельности территориальных органов государственной статистики при председателе Госкомстата России создан *Совет руководителей территориальных органов государственной статистики,* который является совещательным органом и состоит из председателей комитетов – постоянных членов Совета руководителей – и председателей ассоциаций территориальных органов государственной статистики. Решения Совета руководителей носят рекомендательный характер.

Совет руководителей в своей работе руководствуется законодательством Российской Федерации, Положением о Госкомстате России, приказами, постановлениями, распоряжениями Госкомстата России, решениями его коллегии, указаниями председателя Госкомстата России.

Задачи и функции Госкомстата России реализуются функционально-отраслевыми подразделениями, управлениями статистического планирования и организации статистического наблюдения; национальных счетов; статистики предприятий и структурных обследований; статистики основных фондов и строительства; статистики окружающей среды и сельского хозяйства; статистики цен и финансов; статистики населения и ряда других статистик по основным отраслям экономики и социальной среды.

В ведении Госкомстата России находится и действует Главный межрегиональный центр обработки и распространения статистической информации – *ГМЦ Госкомстата России*, который обладает необходимыми мощными вычислительными ресурсами для обработки поступающих из регионов статистических данных.

При Госкомстате России работают Научно-исследовательский институт проблем социально-экономической статистики – НИИ статистики Госкомстата России; Межотраслевой институт повышения квалификации руководящих работников и специалистов в области учета и статистики – МИПК Госкомстата России; Научно-исследовательский и проектно-технологический институт статистической информации – НИПИстатинформ; Информационно-издательский центр «Статистика России» – АНО ИИЦ «Статистика России»; Межведомственный центр социально-экономических измерений РАН и Госкомстата России, а также действует Управление по эксплуатации зданий.

Система органов государственной статистики. Система органов государственной статистики представляет собой сеть иерархически и функционально зависимых организаций, которые занимаются сбором, разработкой и распространением статистических данных, характеризующих темпы и пропорции социально-экономического развития страны, обеспечивающих сравнение с другими странами и оценивающих положение в современном мире.

Система сформирована в соответствии с административно-территориальным делением страны в целях обеспечения органов государственной власти и управления всех уровней, средств массовой информации, научной общественности, коммерческих структур, населения и международных организаций полной и объективной статистической информацией по вопросам социально-экономического развития Российской Федерации, ее регионов, отраслей и секторов экономики.

К началу 2002 г. в системе органов государственной статистики насчитывалось 31,9 тыс. работающих, эксплуатировалось свыше 14,5 тыс. ед. компьютерной техники, в том числе компьютеров типа Pentium – 7,8 тыс. Из общей численности работающих 2,59% приходилось на центральный аппарат Госкомстата России; 3,64% – на организации, подведомственные Госкомстату России; 93,77% – на территориальные органы субъектов Российской Федерации, в том числе 27,96% на уровень районного звена.

По имеющимся приблизительным оценкам в органах *негосударственной статистики,* учебных заведениях, общественных и научных организациях, в средствах массовой информации в настоящее время в России занято более 70 тыс. чел.

Система включает два уровня организаций (рис. 1.2): федеральный уровень, который представляет федеральные органы государственной статистики, и территориальный уровень, представленный органами государственной статистики субъектов Российской Федерации и статистическими органами районного звена.

Федеральный уровень органов государственной статистики представляет Государственный комитет Российской Федерации по стати-

Рис. 1.2. Структура централизованной системы органов государственной статистики Российской Федерации

стике (Госкомстат России), его центральный аппарат и его подведомственные организации.

Территориальный уровень органов государственной статистики представляют комитеты государственной статистики в субъектах Российской Федерации (республиках, краях, областях, а также в Москве и Санкт-Петербурге, всего в 89 субъектах) и районные (городские) отделы статистики, а также специалисты в районах (городах) области (края) без образования отдельного структурного подразделения.

Комитет государственной статистики в субъектах Российской Федерации обеспечивает сбор, предварительную обработку статистической информации на подведомственной территории, представление информации для республиканских, краевых и областных администраций и других потребителей, ведение регистров и классификаторов по соответствующему региону в пределах своей компетенции.

С 2000 г. на территориальные органы Госкомстата России, в связи с образованием федеральных округов, возложены дополнительные функции, которые закреплены в новых положениях о соответствующих территориальных органах Госкомстата России. Территориальные органы Госкомстата России, расположенные в центре федеральных округов, обеспечивают в настоящее время статистической информацией полномочных представителей Президента Российской Федерации в федеральных округах и их аппараты. Статистическая информация, разрабатываемая в соответствии с Федеральной программой статистических работ, отражает социально-экономическое положение субъектов Российской Федерации и представляется в пределах федеральных округов и федеральных округов в целом, в порядке, установленном Госкомстатом России. Территориальные органы также организуют сбор и разработку статистической информации по запросам полномочных представителей Президента Российской Федерации в соответствующих федеральных округах сверх информации, разрабатываемой в рамках Федеральной программы статистических работ.

Сбор, обработка и анализ данных *по экономическим районам России* в настоящее время не регламентируется отдельным положением, поэтому данную работу выполняют отдельные подразделения Госкомстата России и ряд территориальных комитетов по статистике. На базе этой работы регулярно в стране издается один из наиболее объемных статистических сборников «Регионы России».

Органы государственной статистики *районного звена* осуществляют первичный сбор, контроль и обработку данных, передачу их по принадлежности в вышестоящие звенья, а также проводят информационную поддержку местным органам управления по обеспечению статистической информацией.

В некоторых субъектах Российской Федерации созданы межрайонные (кустовые) отделы (узлы) статистики, которые занимаются сводной и предварительной обработкой информации, а также и подготовкой отчетов по заказам потребителей, в то время как районные отделы статистики ограничиваются выполнением функций сбора и контроля первичной информации. Районные отделы статистики, как правило, входят в состав комитетов государственной статистики соответствующих субъектов Российской Федерации по принадлежности. Районные отделы (узлы) статистики, созданные на правах самостоятельных образований в отдельных субъектах Российской Федерации, не входят в состав соответствующих территориальных комитетов по статистике, а функционируют с ними наравне, подчиняясь напрямую Госкомстату России.

В *муниципальных образованиях* для органов местного самоуправления разработку статистической информации осуществляют соответствующие подразделения территориальных органов Госкомстата России.

Следовательно, 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, организации и их подразделения. Управлять сложными социальными и экономическими системами нельзя, не располагая оперативной, достоверной и полной статистической информацией, которую дает система органов государственной статистики России.

Распространение статистической информации. Основной формой распространения статистической информации Госкомстата России являются печатные издания: периодика и статистические сборники.

К основным периодическим изданиям относятся: ежемесячный доклад «Социально-экономическое положение России»; ежемесячный журнал «Вопросы статистики»; ежеквартальный журнал «Статистическое обозрение».

Статистические сборники представляют: «Российский статистический ежегодник», а также справочники «Регионы России» (в двух томах), «Россия в цифрах», «Малое предпринимательство в России», «Женщины и мужчины России», «Демографический ежегодник России» и ряд других изданий.

Для оперативного обеспечения статистической информацией федеральных и региональных органов власти на базе современных информационных технологий ведется Банк готовых документов (БГД)

«Статистика России». БГД является электронной версией официальных публикаций Госкомстата России и его территориальных органов. С банком готовых документов «Статистика России» можно ознакомиться через сеть Интернета на сайте Госкомстата России – http://www.gks.ru.

ОСНОВНЫЕ ПОНЯТИЯ

Аналитическая статистика – процедура оценки характеристик совокупности по данным выборок.

Вариация – колеблемость, многообразие, изменяемость значения признака у отдельных единиц совокупности.

Единица статистической совокупности – каждый отдельно взятый элемент данного множества, обладающий определенными признаками.

Задача статистического исследования – получение обобщающих показателей и выявление закономерностей социально-экономических явлений и процессов в конкретных условиях места и времени.

Закономерность – повторяемость, последовательность и порядок изменений в явлениях.

Общая теория статистики – отрасль статистической науки о наиболее общих принципах, правилах и законах цифрового освещения социально-экономических явлений.

Описательная (дескриптивная) статистика – получение статистических показателей, с помощью которых обобщаются характеристики только наблюдаемой совокупности. Задача ее заключается в том, чтобы дать сжатую и концентрированную характеристику изучаемого явления.

Предмет статистики – количественная сторона качественно определенных массовых социально-экономических явлений и процессов, отображаемых посредством статистических показателей.

Признак – общее свойство, характерная черта или иная особенность единиц совокупности, которые могут быть наблюдаемы или измерены.

Статистика – общественная наука, имеющая целью сбор, упорядочение, анализ и сопоставление данных, относящихся к самым разнообразным массовым явлениям.

Система показателей – совокупность взаимосвязанных показателей, которые отражают состояние и развитие массовых социально-экономических явлений с разных сторон.

Статистическая закономерность – форма проявления причинной связи, выражающаяся в последовательности, регулярности, повторяемости событий с достаточно высокой степенью вероятности, если причины, порождающие события, не изменяются или изменяются незначительно. Статистические закономерности устанавливаются на основе анализа массовых данных.

Статистическая методология – система приемов, способов и методов, направленных на изучение количественных закономерностей, проявляющихся в структуре, динамике и взаимосвязи социально-экономических явлений.

Статистическая совокупность – множество единиц, обладающих массовостью, однородностью, определенной целостностью, взаимозависимостью состояний и наличием вариации.

Статистический показатель – обобщающая количественная характеристика социально-экономических явлений в конкретных условиях места и времени.

ТЕСТЫ

1. Укажите правильное научное значение термина «статистика»:

а) сбор сведений о различных общественных явлениях;

б) различные статистические сборники;

в) особая отрасль науки;

г) различного рода цифры и числовые данные.

2. Предметом статистики как науки являются:

а) метод статистики;

б) статистические показатели;

в) группировки и классификации;

г) количественные закономерности массовых варьирующих общественных явлений.

3. Укажите правильный ответ. Статистическая наука зародилась:

а) до начала современной эры летоисчисления;

б) в VII в.;

в) в XVII в.;

г) в XIX в.

4. Работник, для которого сбор статистических данных является профессиональной деятельностью, именуется:

а) статистом;

б) статистиком;

в) переписчиком;

г) сборщиком данных.

5. Основным разделом статистической науки является:

а) математическая статистика;

б) теория вероятностей;

в) промышленная статистика;

г) общая теория статистики.

6. Совокупность – это:

а) любое предметное множество явлений природы и общества;

б) множество элементов, обладающих общими признаками;

в) реально существующее множество однородных элементов, обладающих общими признаками и внутренней связью;

г) математическое множество.

7. Элемент совокупности – это:

а) признак совокупности;

б) элемент математического множества;

в) носитель информации;

г) элемент таблицы Менделеева.

8. Какой из перечисленных признаков является варьирующим:

а) цена одного килограмма товара;

б) температура кипения воды;

г) курс доллара;

г) скорость падения тела в пустоте.

9. Признаки элементов статистической совокупности бывают только:

а) количественные;

б) количественные и качественные;

в) качественные;

г) безразмерные.

10. Вариация – это:

а) изменение массовых явлений во времени;

б) изменение структуры статистической совокупности в пространстве;

в) изменение значений признака;

г) изменение состава совокупности.

ЛИТЕРАТУРА

Боярский А.Я. Теоретические исследования по статистике. – М.: Статистика, 1974. – 320 с.

Володарский Л.М. Статистика рассказывает. – М.: Молодая гвардия, 1982.

Едронова В.Н., Едронова М.В. Общая теория статистики. – М.: Юрист, 2001 (глава 8). – 511 с.

Елисеева И.И. Моя профессия – статистик. – М.: Финансы и статистика, 1992.

Ефимова М.Р., Петрова Е.В., Румянцев В.Н. Общая теория статистики. – М.: Инфра-М, 1999 (глава 1). – 411 с.

Пасхавер И.С. Закон больших чисел и статистическая закономерность. – М.: Статистика, 1972. – 149 с.

Ряузов Н.Н. Общая теория статистики. – М.: Финансы и статистика, 1984. – 342 с.

Статистический словарь. – М.: Финансы и статистика, 1989. – 621 с.

Толстик Н.В., Матегорина Н.М. Статистика. – Ростов-на-Дону: Феникс, 2001 (глава 1). – 479 с.

Успенский Г.И. Живые цифры. Избранные произведения. – М.: Московский рабочий, 1949. – 720 с.

ГЛАВА 2

СБОР СТАТИСТИЧЕСКОЙ ИНФОРМАЦИИ (теория статистического наблюдения)

2.1 ПОНЯТИЕ О СТАТИСТИЧЕСКОМ НАБЛЮДЕНИИ, ЭТАПЫ ЕГО ПРОВЕДЕНИЯ

Статистическое наблюдение — это массовое, планомерное, научно организованное наблюдение за явлениями социальной и экономической жизни, которое заключается в регистрации отобранных признаков у каждой единицы совокупности.

Статистическое наблюдение может проводиться органами государственной статистики, научно-исследовательскими институтами, экономическими службами банков, бирж, фирм.

Проведение статистического наблюдения включает следующие этапы:

- подготовку наблюдения;
- проведение массового сбора данных;
- подготовку данных к автоматизированной обработке;
- разработку предложений по совершенствованию статистического наблюдения.

Любое статистическое наблюдение требует тщательной, продуманной подготовки. От нее во многом будут зависеть надежность и достоверность информации, своевременность ее получения.

Подготовка статистического наблюдения включает разные виды работ. Сначала решаются методологические вопросы, важнейшие из которых – определение цели и объекта наблюдения, состава признаков, подлежащих регистрации; разработка документов для сбора данных; выбор отчетной единицы и единицы, относительно которой будет проводиться наблюдение, а также определение методов и средств получения данных.

Кроме методологических необходимо решить проблемы организационного характера, например, определить состав служб, проводящих наблюдение; подобрать и подготовить кадры для проведения наблюдения; составить календарный план работ по подготовке, проведению и обработке материалов наблюдения; провести тиражирование документов для сбора данных.

Проведение массового сбора данных включает работы, связанные непосредственно с заполнением статистических формуляров. Сбор данных начинается с рассылки переписных листов, анкет, бланков, форм статистической отчетности и после заполнения заканчивается их сдачей в органы, проводящие наблюдение.

Собранные данные на этапе их *подготовки к автоматизированной обработке* подвергаются *арифметическому и логическому контролю*. Оба эти контроля основываются на знании взаимосвязей между показателями и качественными признаками.

На заключительном этапе проведения наблюдения анализируются причины, которые привели к неверному заполнению статистических бланков, и *разрабатываются предложения по совершенствованию наблюдения*. Это очень важно для организации будущих обследований.

Получение сведений в ходе статистического наблюдения требует немалых затрат финансовых и трудовых ресурсов, а также времени.

2.2 ПРОГРАММНО-МЕТОДОЛОГИЧЕСКИЕ ВОПРОСЫ СТАТИСТИЧЕСКОГО НАБЛЮДЕНИЯ

Цель наблюдения. Статистические наблюдения чаще всего преследуют практическую цель – получение достоверной информации для выявления закономерностей развития явлений и процессов. Например, целью обследования малых предприятий по итогам работы за 2000 г. было получение полного и достоверного перечня малых предприятий; данных о видах осуществляемой ими деятельности, объемах производства и продаж.

Задача наблюдения предопределяет его программу и формы организации. Неясно поставленная цель может привести к тому, что в процессе наблюдения будут собраны ненужные данные или, наоборот, не будут получены сведения, необходимые для анализа.

Объект и единица наблюдения. Отчетная единица. При подготовке наблюдения кроме цели следует точно определить, что именно подлежит обследованию, т.е. установить объект наблюдения.

Под *объектом наблюдения* понимается некоторая статистическая совокупность, в которой протекают исследуемые социально-экономические явления и процессы. Объектом наблюдения может быть совокупность физических лиц (население отдельного региона, страны; лица, занятые на предприятиях отрасли), физические единицы (станки, машины, жилые дома), юридические лица (предприятия, фермерские хозяйства, коммерческие банки, учебные заведения).

Чтобы определить объект статистического наблюдения, необходимо установить границы изучаемой совокупности. Для этого следует указать важнейшие признаки, отличающие данный объект от других похожих объектов. Например, прежде чем проводить обследование рентабельности промышленных предприятий, следует определить формы собственности, организационно-правовые формы предприятий, отрасли промышленности и регионы, подлежащие наблюдению.

Всякий объект статистического наблюдения состоит из отдельных элементов – единиц наблюдения.

В статистике *единицей наблюдения* (в зарубежной литературе используется термин «элементарная единица») называют составной элемент объекта, являющийся носителем признаков, подлежащих регистрации. Например, при демографических обследованиях единицей наблюдения может быть человек, но может быть и семья; при бюджетных обследованиях – семья или домашнее хозяйство.

Единицу наблюдения следует отличать от отчетной единицы. *Отчетной единицей* является субъект, от которого поступают данные о единице наблюдения. Так, при организации статистического наблюдения в капитальном строительстве информация может быть получена от проектных или подрядных организаций или от предприятий-застройщиков.

Единица наблюдения и отчетная единица могут совпадать. Например, если надо определить объем освоенных за год капитальных вложений, то предприятие-застройщик будет одновременно и единицей наблюдения, и отчитывающейся организацией. Однако при изучении процесса концентрации капитальных вложений отчетной единицей по-прежнему будет застройщик, а единицей наблюдения станут стройки и объекты, строительство которых ведет данный застройщик.

Программа статистического наблюдения. Всякое явление обладает множеством различных признаков. Собирать информацию по всем признакам нецелесообразно, а часто и невозможно. Поэтому необходимо отобрать те *признаки*, которые являются *существенными*, основными для характеристики объекта исходя из цели исследования. Для определения состава регистрируемых признаков разрабатывают программу наблюдения.

Программа наблюдения – это перечень признаков (или вопросов), подлежащих регистрации в процессе наблюдения. От того, насколько хорошо разработана программа статистического наблюдения, во многом зависит качество собранной информации.

Чтобы составить правильно программу наблюдения, исследователь должен ясно представить задачи обследования конкретного явления или процесса, определить состав используемых в анализе методов, необходимые группировки и уже на основе этого выявить те признаки, которые нужно определить при проведении работы. Обычно программа выражается в форме вопросов переписного (опросного) листа.

Требования к программе статистического наблюдения. Программа должна содержать существенные признаки, непосредственно характеризующие изучаемое явление, его тип, основные черты, свойства. Не следует включать в программу признаки, имеющие второстепенное значение по отношению к цели обследования, или признаки, значения которых будут заведомо недостоверны или отсутствовать, например представление информации, которая является предметом коммерческой тайны.

Вопросы программы должны быть точными и недвусмысленными (иначе полученный ответ может содержать неверную информацию), а также легкими для понимания во избежание лишних трудностей при получении ответа.

При разработке программы следует не только определить состав вопросов, но и их последовательность. Логичный порядок исследования вопросов (признаков) поможет получить достоверные сведения о явлениях и процессах.

В программу целесообразно включать вопросы контрольного характера для проверки и уточнения собираемых данных.

Вопросы в программе задаются в различной форме. Они могут быть закрытые и открытые. Закрытый вопрос – это вопрос альтернативный, т.е. предполагающий выбор одного из двух ответов: «да» или «нет» или же вопрос с выборочным ответом, где предлагаются три и более варианта ответа на выбор. Например, ответ на вопрос «состояние в браке» может быть одним из следующих: а) состоит в браке; б) никогда не состоял в браке; в) в браке; г) вдовец (вдова); д) разведен(а), разошелся(лась).

На открытые вопросы можно ответить практически бесчисленным количеством способов, если вопрос поставлен без заданной структуры ответа. Например, «какие ценности являются для вас главными?»

Для обеспечения единообразия получаемых сведений от каждой отчетной единицы (это важно при последующей обработке информации) программа оформляется в виде документа, называемого статистическим формуляром.

Статистический формуляр. Это документ единого образца, содержащий программу и результаты наблюдения.

Обязательными элементами статистического формуляра являются титульная и адресная части. *Титульная часть* содержит наименование статистического наблюдения и органа, проводящего наблюдение, информацию о том, кто и когда утвердил этот формуляр, иногда его номер. *Адресная часть* включает адрес отчетной единицы, ее подчиненность.

Формуляр может иметь разные названия: отчет, карточка, переписной лист, опросный бланк, анкета и т.д.

Различают две системы статистического формуляра: индивидуальную (карточную) и списочную.

Индивидуальный формуляр предусматривает запись на нем ответов на вопросы программы только об одной единице наблюдения, *списочный* – о нескольких единицах. Так, все формы статистической отчетности заполняются каждым предприятием в отдельности, а при проведении переписи населения члены каждой семьи записываются в один переписной лист.

Кроме формуляра разрабатывается *инструкция,* определяющая порядок проведения наблюдения и заполнения формы отчетности, переписного листа, анкеты. В зависимости от сложности программы наблюдения инструкция публикуется в виде отдельной брошюры или помещается на оборотной стороне формуляра. Формуляр и инструкция по его заполнению составляют *инструментарий статистического наблюдения.*

Место и время наблюдения. Выбор места проведения обследования зависит главным образом от цели наблюдения. Если необходимо получить данные для изучения состава населения по стране, то в этом случае наблюдение охватит территорию всей страны. При сборе сведений о стоимости потребительской корзины в Москве и Санкт-Петербурге местом проведения обследования будут территории этих двух крупнейших городов страны.

Выбор времени наблюдения заключается в решении двух вопросов:

- установлении критического момента (даты) или интервала времени;
- определении срока (периода) наблюдения.

Под *критическим моментом* (датой) понимаются конкретный день года, час дня, по состоянию на который должна быть проведена регистрация признаков по каждой единице исследуемой совокупности. Так, критическим моментом микропереписи населения Российской Федерации в 1994 г. был 0 часов в ночь с 13 на 14 февраля 1994 г. Критический момент устанавливается с целью получения сопоставимых статистических данных. В случае исследования варьирования биржевых котировок на торгах валютных бирж в различных городах России необходимы данные о курсах доллара США, японской иены, немецкой марки и других валют, зарегистрированных в один и тот же день. Если же надо проанализировать изменение объема продаж какой-либо валюты на биржевом рынке в отчетном месяце по сравнению с предыдущим месяцем, то устанавливается не критический момент, а *интервал времени*, за который следует получить статистические данные.

Выбор критического момента или интервала времени определяется прежде всего целью исследования.

Срок (период) наблюдения – это время, в течение которого происходит заполнение статистических формуляров, т.е. время, необходимое для проведения массового сбора данных. Этот срок определяется исходя из объема работы (числа регистрируемых признаков и единиц в обследуемой совокупности), численности персонала, занятого сбором информации. Следует учитывать, что отдаление периода наблюдения от критического момента или интервала времени может привести к снижению достоверности получаемых сведений. Например, микроперепись населения, упомянутая ранее, проводилась в течение десяти дней – с 14 по 23 февраля 1994 г.

2.3 ВАЖНЕЙШИЕ ОРГАНИЗАЦИОННЫЕ ВОПРОСЫ СТАТИСТИЧЕСКОГО НАБЛЮДЕНИЯ

Успех любого статистического наблюдения зависит не только от тщательной методологической подготовки, но и от правильного и своевременного решения широкого спектра организационных вопросов.

Важнейшее место в организационной работе занимает подготовка кадров, в процессе которой проводятся различного рода инструктажи с сотрудниками статистических органов, с организациями, предоставляющими данные, по вопросам заполнения статистических документов, подготовки материалов наблюдения к автоматизированной обработке и т.д.

Если проведение наблюдения связано с большими затратами трудовых ресурсов, то для регистрации сведений в период проведения обследований привлекаются лица из числа неработающих (в том числе безработные) и некоторых категорий учащихся (студенты высших учебных заведений, учащиеся старших курсов техникумов). При проведении переписи населения таких лиц называют *счетчиками*. Обычно организуется обучение персонала. Оно проводится для выработки у счетчиков навыков правильного заполнения статистических формуляров.

Размножение документации как для самого обследования, так и для проведения инструктажей и рассылка ее по территориальным органам Госкомстата России также относятся к организационным вопросам наблюдения.

В период подготовки большая роль отводится массовой работе: проведению лекций, бесед, организации выступлений в печати, по радио и телевидению, разъясняющих населению значение, цели и задачи предстоящего обследования.

Для согласования деятельности всех служб, занятых подготовкой и проведением наблюдения, составляется календарный план, представляющий собой перечень (наименование) работ и сроки их исполнения отдельно для каждой организации, занятой в проведении обследования.

2.4 ОСНОВНЫЕ ОРГАНИЗАЦИОННЫЕ ФОРМЫ, ВИДЫ И СПОСОБЫ СТАТИСТИЧЕСКОГО НАБЛЮДЕНИЯ

На этапе подготовки обследования нужно выяснить, как часто оно будет проводиться, будут ли обследоваться все единицы совокупности или только часть их, как получать информацию об объекте (путем интервью по телефону, по почте, простым наблюдением и т.п.). Другими словами, необходимо определить формы, способы и виды статистического наблюдения.

Формы статистического наблюдения (рис. 2.1). В отечественной статистике используются три организационные формы (типы) статистического наблюдения:

• отчетность (предприятий, организаций, учреждений и т.п.);

• специально организованное статистическое наблюдение (переписи, единовременные учеты, обследования сплошного и несплошного характера);

• регистры.

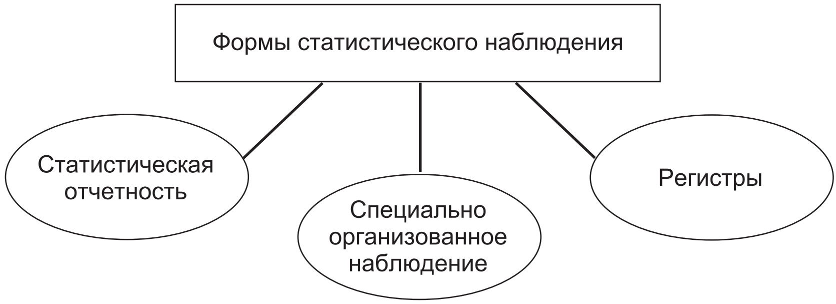

Рис. 2.1. Классификация форм статистического наблюдения

Статистическая отчетность. *Отчетность* – это основная форма статистического наблюдения, с помощью которой статистические органы в определенные сроки получают от предприятий, учреждений и организаций необходимые данные в виде установленных в законном порядке отчетных документов, скрепленных подписями лиц, ответственных за их предоставление и достоверность собираемых сведений. Таким образом, отчетность – это официальный документ, содержащий статистические сведения о работе предприятия, учреждения, организации и т.п.

Отчетность как форма статистического наблюдения основана на первичном учете и является его обобщением. Первичный учет – это регистрация различных фактов, событий по мере их совершения, как правило, на особом документе, называемом *первичным учетным документом.*

Отчетность утверждается *органами государственной статистики.* Представление информации по неутвержденным формам является нарушением отчетной дисциплины.

Отчетность *имеет обязательный характер* (т.е. все предприятия, учреждения, организации должны представлять ее в указанные сроки), а также *юридическую силу,* потому что подписывается руководителем предприятия (учреждения, организации).

Отчетность обладает *документальной обоснованностью,* так как все данные базируются на документах первичного учета.

Действующую статистическую отчетность делят на типовую и специализированную. Состав показателей в *типовой отчетности* является единым для предприятий всех отраслей народного хозяйства. В *специализированной отчетности* состав показателей изменяется в зависимости от особенностей отдельных отраслей экономики.

По *срокам представления* отчетность бывает ежедневная, недельная, двухнедельная, месячная, квартальная и годовая. Кроме годовой отчетности все перечисленные виды представляют собой текущую отчетность.

По *способу представления* сведений отчетность делится на электронную, телеграфную, телетайпную, почтовую.

Специально организованное статистическое наблюдение. Перепись. Специально организованное наблюдение проводится с целью получения сведений, отсутствующих в отчетности, или для проверки ее данных. Наиболее простым примером такого наблюдения является перепись. Российская практическая статистика проводит переписи населения, материальных ресурсов, многолетних насаждений, неустановленного оборудования, строек незавершенного строительства, оборудования и др.

Перепись – это специально организованное наблюдение, повторяющееся, как правило, через равные промежутки времени с целью получения данных о численности, составе и состоянии объекта статистического наблюдения по ряду признаков.

Характерными особенностями переписи являются:

- одновременность проведения ее на всей территории, которая должна быть охвачена обследованием;
- единство программы наблюдения;
- регистрация всех единиц наблюдения по состоянию на один и тот же критический момент времени.

Программа наблюдения, приемы и способы получения данных по возможности должны оставаться неизменными. Это позволяет обеспечить сопоставимость собираемой информации и получаемых в ходе разработки материалов переписи обобщающих показателей. Тогда можно не только определить численность и состав исследуемой совокупности, но и проанализировать ее количественное изменение в период между двумя обследованиями.

Из всех переписей наиболее известны *переписи населения*. Цель последних состоит в установлении численности и размещения населения по территории страны, в получении характеристики состава населения по полу, возрасту, занятиям и другим показателям. Первая всеобщая перепись населения России была проведена в 1897 г., а последняя – в 2002 г.

В период подготовки всеобщей переписи для уточнения и апробации программно-методических и организационных вопросов наблюдения проводят пробную перепись.

Пробная перепись населения является неотъемлемым этапом подготовки всеобщей переписи населения. Поэтому в рамках подготовки Всероссийской переписи населения 2002 г. в г. Москве с 11 по 18 октября 2000 г. была проведена пробная перепись населения. Этой переписью было охвачено около 100 тыс. чел.

В ходе пробной переписи был опробован проект программы Всероссийской переписи населения 2002 г. Эта программа позволит получить данные, отражающие современное состояние населения России. Программа имеет преемственность с прошлыми переписями населения (неизменными остаются вопросы, касающиеся структуры населения, а именно: пол, возраст, семейное положение, национальность) и содержит вопросы, отражающие изменения в социально-экономической сфере, которые произошли в стране с начала 90-х годов двадцатого века.

Переписи получили большое распространение и в зарубежной статистике. Среди них наиболее интересными являются систематически проводимые в США переписи отраслей национального хозяйства, в частности переписи обрабатывающей промышленности, называемые цензами. (Следует иметь в виду, что слово «ценз» имеет несколько значений. Это не только синоним слова «перепись». Под ним еще понимается ряд признаков, наличие которых при организации наблюдения служит основанием для отнесения той или иной единицы к исследуемой совокупности.) Американские переписи охватывают все предприятия и проводятся один раз в пять лет (в годы, оканчивающиеся на цифру 2 или 7). В промежутке между переписями проводятся ежегодные выборочные обследования для заполнения пробелов в данных.

Программа таких переписей предусматривает получение данных о численности занятого населения, заработной плате, отработанных человеко-часах, затратах по снабжению; уточнение сведений о потреблении электроэнергии, капитальных вложениях, стоимости и количестве отгруженной продукции, запасах готовой продукции, стоимости незавершенного производства, материалов и топлива на конец года, а также содержит специальные вопросы о типе предприятия, его оборудовании и т.д.

Опросные формы рассылаются предприятиям для заполнения по почте за 4–7 месяцев до начала переписи. Это позволяет отчетным единицам своевременно и правильно заполнить переписные листы.

Кроме переписей статистика проводит и другие специально организованные наблюдения, в частности, к специально организованному статистическому наблюдению относится *сплошное статистическое обследование малых предприятий по итогам работы* за 2000 г. Оно проводилось в два этапа. На первом этапе собирались сведения о фактических видах деятельности малых предприятий на основе упрощенного вопросника. На втором этапе каждому предприятию высылался основной вопросник, содержащий общую часть и специальные вопросы, связанные с видом деятельности, указанной предприятием на первом этапе. Данное обследование было направлено на решение следующих задач: получение полного и достоверного перечня малых предприятий с содержанием сведений о фактически осуществимых ими видах деятельности; получение сведений о производстве малыми предприятиями конкретных видов товаров и услуг, которые невозможно получить при выборочном наблюдении; уточнение генеральной совокупности малых предприятий как основы для проведения регулярных выборочных обследований; получение полной и объективной информации для анализа и прогнозирования развития малого предпринимательства.

Регистровая форма наблюдения. *Регистровое наблюдение* – это форма непрерывного статистического наблюдения за долговременными процессами, имеющими фиксированное начало, стадию развития и фиксированный конец. Оно основано на ведении статистического регистра. *Регистр* представляет собой систему, постоянно следящую за состоянием единицы наблюдения и оценивающую силу воздействия различных факторов на изучаемые показатели. В регистре каждая единица наблюдения характеризуется совокупностью показателей. Одни из них остаются неизменными в течение всего времени наблюдения и регистрируются один раз; другие показатели, периодичность изменения которых неизвестна, обновляются по мере изменения; третьи – представляют собой динамические ряды показателей с заранее известным периодом обновления. Все показатели хранятся до полного завершения наблюдения за единицей обследуемой совокупности.

Организация и ведение регистра невозможны без решения следующих вопросов:

- Когда заносить в регистр и исключать из него единицы совокупности?
- Какая информация должна храниться?

• Из каких источников следует брать данные?
• Как часто обновлять и дополнять информацию?

В практике статистики различают регистры населения и регистры предприятий.

Регистр населения — поименованный и регулярно актуализируемый перечень жителей страны. Программа наблюдения ограничена общими признаками, такими, как пол, дата и место рождения, дата вступления в брак (эти данные остаются неизменными в течение всего периода наблюдения) и брачное состояние (переменный признак). Как правило, регистры хранят информацию только по тем переменным признакам, изменение значений которых документально оформлено.

Информация в регистр заносится на каждого родившегося и прибывшего из-за границы. Если человек умер или выехал на постоянное место жительства из страны, то сведения о нем изымаются из регистра. Регистры населения ведутся по отдельным регионам страны. При перемене места жительства сведения по единице наблюдения передаются в регистр соответствующей территории. В связи с тем что правила регистрации довольно сложны и ведение регистра требует больших затрат, эта форма наблюдения практикуется в государствах с небольшой численностью и высокой культурой населения (в основном это европейские страны).

Необходимо отметить, что регистр населения, как любой регистр, охватывающий наблюдением значительную совокупность единиц, содержит данные по ограниченному числу признаков. Поэтому ведение регистра предполагает проведение специально организованных обследований, в том числе и переписей населения.

Регистр предприятий включает в себя все виды экономической деятельности и содержит значения основных признаков по каждой единице наблюдаемого объекта за определенный период или момент времени. Регистры предприятий содержат данные о времени создания (регистрации) предприятия, его название и адрес, телефон, об организационно-правовой форме, структуре, виде экономической деятельности, количестве занятых (этот показатель отражает размер предприятия) и др.

В нашей стране были разработаны три регистра: промышленных предприятий, строек и подрядных организаций. Внедрение их в статистическую практику существенно повысило информационный и аналитический уровень статистики, позволило решить ряд экономико-статистических задач, для которых непригодны другие формы статистического наблюдения.

В настоящее время завершены работы по созданию единого регистра для всех хозяйственных единиц. Ему отводится важное значение во внедрении системы национальных счетов в статистическую практику.

Единый государственный регистр предприятий и организаций всех форм собственности (ЕГРПО) дает возможность организовать сплошное наблюдение по ограниченному кругу статистических показателей предприятий, зарегистрированных на территории России, позволяет получить непрерывные ряды показателей в случае изменения территориальной, отраслевой и других структур совокупности.

В регистр заносятся данные по всем предприятиям, организациям, учреждениям и объединениям независимо от форм собственности, включая предприятия с иностранными инвестициями, банковские учреждения, общественные объединения и другие юридические лица.

Информационный фонд регистра содержит:

- регистровый код субъекта;
- сведения об отраслевой, территориальной принадлежности субъекта, его подчиненности, виде собственности, организационной форме;
- справочные сведения (фамилии руководителей, адреса, номера телефонов, факсов и т.д., сведения об учредителях);
- экономические показатели.

Значения последних будут заноситься в регистр на основе бухгалтерской и статистической отчетностей, представляемых в региональные органы статистики.

Регистр содержит данные о следующих показателях: среднесписочная численность работников; средства, направляемые на потребление; остаточная стоимость основных средств; балансовая прибыль (убыток); уставный фонд. Так как регистр ведется по отдельным территориям, региональные статистические службы могут расширять состав экономических показателей в случае необходимости.

ЕГРПО позволяет проводить отбор и группировку любой совокупности единиц по одному или нескольким признакам.

Сбор данных о единицах наблюдения осуществляется в процессе их государственной регистрации и последующего учета.

При закрытии предприятия ликвидационная комиссия в десятидневный срок уведомляет об этом службу ведения регистра.

С 1 июля 1996 г. ведется учет индивидуальных предпринимателей в составе ЕГРПО. В ЕГРПО учтено свыше 1 млн индивидуальных предпринимателей, а также более 3 млн организаций и их обособленных подразделений. На базе регистра сформирована генеральная совокупность объектов статистического наблюдения, которая обеспечивает основу организации статистических наблюдений, применение интегрированного принципа сбора и обработки статистической отчетности, проведение анализа по сопоставимому кругу объектов и типу предприятий.

Пользователями регистра могут быть любые юридические или физические лица, заинтересованные в получении информации.

Способы статистического наблюдения. Статистическая информация может быть получена различными способами, важнейшими из которых являются непосредственное наблюдение, документальный учет фактов и опрос (рис. 2.2).

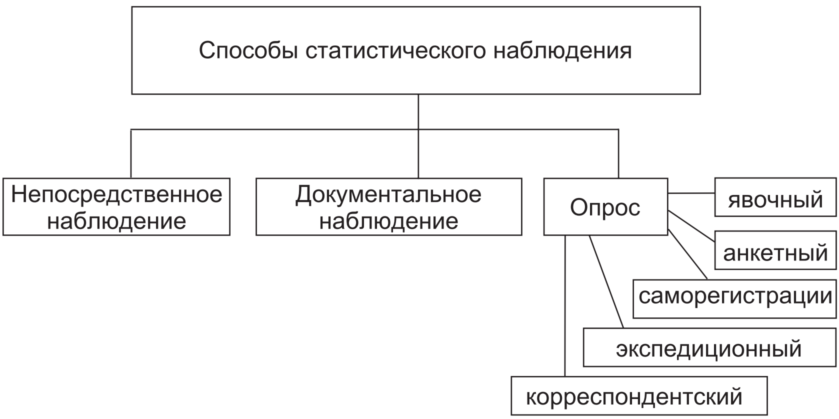

Рис. 2.2. Способы статистического наблюдения

Непосредственным называют наблюдение, при котором сами регистраторы путем непосредственного замера, взвешивания, подсчета или проверки работы и т.д. устанавливают факт, подлежащий регистрации, и на этой основе производят запись в формуляр наблюдения. Этот способ применяют при наблюдении за вводом в действие жилых домов.

Документальное наблюдение основано на использовании в качестве источника статистической информации различного рода документов, как правило, учетного характера. При надлежащем контроле за постановкой первичного учета и правильном заполнении статистических формуляров документальное наблюдение дает наиболее точные результаты.

Опрос – это способ наблюдения, при котором необходимые сведения получают со слов респондента. Опрос предполагает обращение к непосредственному носителю признаков, подлежащих регистрации во время наблюдения, и используется для получения информации о явлениях и процессах, не поддающихся непосредственному прямому наблюдению.

В статистике применяются следующие виды опросов: устный (экспедиционный), саморегистрация, корреспондентский, анкетный и явочный.

При *устном* (экспедиционном) опросе специально подготовленные работники (счетчики, регистраторы) получают необходимую информацию на основе опроса соответствующих лиц и сами фиксируют ответы в формуляре наблюдения. По форме проведения устный опрос может быть прямым (например, при переписи населения), когда счетчик «лицом к лицу» встречается с каждым респондентом, и опосредованным, например по телефону.

При *саморегистрации* формуляры заполняются самими респондентами, а счетчики раздают респондентам бланки опросного листа, разъясняют правила их заполнения, а затем бланки собирают.

Корреспондентский способ заключается в том, что сведения в органы, ведущие наблюдения, сообщает штат добровольных корреспондентов. Этот вид опроса требует наименьших затрат, но не дает уверенности в том, что полученный материал высококачественный, так как не всегда возможно проверить на месте правильность полученных ответов.

Анкетный способ предполагает сбор информации в виде анкет. Определенному кругу респондентов вручаются специальные вопросники (анкеты) либо лично, либо путем публикации в периодической печати. Заполнение этих вопросников носит добровольный характер и осуществляется, как правило, анонимно. Обычно обратно получают меньше анкет, чем рассылают. Этот способ сбора информации используется при несплошном наблюдении. Анкетный опрос применяется в обследованиях, где не требуется высокая точность, а нужны приближенные, ориентировочные результаты, например при изучении общественного мнения о работе городского транспорта, торговых предприятий и т.д.

Явочный способ предусматривает представление сведений в органы, ведущие наблюдение, в явочном порядке, например при регистрации браков, рождений, разводов и т.д.

При выборе вида того или иного опроса необходимо учитывать следующее: с какой точностью надо провести наблюдения; есть ли возможность практического применения того или иного способа; каковы финансовые возможности.

Виды статистического наблюдения (рис.2.3). Статистические наблюдения можно разбить на группы по следующим признакам:

- времени регистрации фактов;
- охвату единиц совокупности.

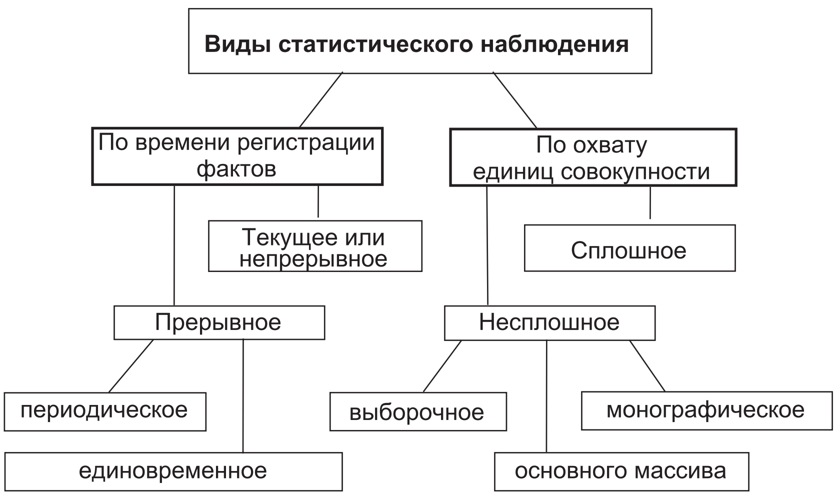

Рис. 2.3. Классификация видов статистического наблюдения

По времени регистрации фактов наблюдение бывает непрерывным (текущим), периодическим и единовременным. При *текущем наблюдении* изменения в отношении изучаемых явлений фиксируются по мере их наступления, например при регистрации рождения, смерти, состояния в браке. Такое наблюдение проводится с целью изучения динамики какого-либо явления.

Данные, отражающие изменение объекта, могут быть собраны в ходе нескольких обследований. Они обычно проводятся по схожей программе и инструментарию и называются *периодическими.* К такому виду наблюдений относятся переписи населения, которые проводятся через каждые 10 лет; регистрация цен производителей по отдельным товарам, которая в настоящее время проводится ежемесячно.

Единовременное обследование дает сведения о количественных характеристиках какого-либо явления или процесса в момент его исследования. Повторная регистрация проводится спустя какое-то время (не определенное заранее) или может не проводиться вообще. Единовременным обследованием была инвентаризация незавершенного производственного строительства 1990 г.

По охвату единиц совокупности статистическое наблюдение бывает сплошное и несплошное.

Задачей *сплошного наблюдения* является получение информации о всех единицах исследуемой совокупности. Поэтому при проведении сплошного наблюдения важной задачей является формирование перечня признаков, подлежащих обследованию. От этого в конечном итоге зависит качество и достоверность результатов обследования.

До последнего времени российская система государственной статистики опиралась прежде всего на сплошное наблюдение. Однако такой вид наблюдения имеет серьезные недостатки: высокую стоимость получения и обработки всего объема информации; большие затраты трудовых ресурсов; недостаточную оперативность информации, так как для ее сбора и обработки необходимо много времени. И, наконец, ни одно сплошное наблюдение, как правило, не обеспечивает полного охвата всех без исключения единиц совокупности. Большее или меньшее число единиц обязательно остается вне наблюдения как при проведении единовременных обследований, так и при получении сведений по такой форме наблюдения, как отчетность.

Например, при проведении сплошного статистического обследования малых предприятий по итогам работы за 2000 г. бланки форм (вопросники) были направлены 911 840 малым предприятиям. Отчеты были получены от 61% предприятий, которым были направлены вопросники.

Причины неответов предприятий следующие: %

- не были найдены по адресу, указанному в ЕГРПО, ... 15
- не осуществляли деятельность в 2000 г. 14
- были ликвидированы на момент проведения обследования 4
- не ответили на вопросник по другим причинам 6

Количество и доля неохваченных единиц зависят от многих факторов: вида обследования (по почте, с помощью устного опроса); типа отчетной единицы; квалификации регистратора; содержания вопросов, предусмотренных программой наблюдения; времени дня или года, когда проводится обследование, и др.

Несплошное наблюдение изначально предполагает, что обследованию подлежит лишь часть единиц изучаемой совокупности. При его проведении следует заранее определить, какая часть совокупности должна быть подвергнута наблюдению и каким образом следует отобрать те единицы, которые должны быть обследованы.

Одним из преимуществ несплошных наблюдений является возможность получения информации в более короткие сроки и с меньшими затратами ресурсов, чем при сплошном наблюдении. Это связано с меньшим объемом собираемой информации, а следовательно, с более низкими затратами на ее получение, проверку достоверности, обработку, анализ.

Существует несколько видов несплошного наблюдения. Одно из них – *выборочное наблюдение*. Это довольно распространенный вид, основанный на принципе случайного отбора тех единиц изучаемой совокупности, которые должны быть подвергнуты наблюдению. При правильной организации выборочное наблюдение дает достаточно точные результаты, вполне пригодные для характеристики всей исследуемой совокупности. В этом состоит достоинство выборочного наблюдения по сравнению с другими видами несплошного наблюдения.

Численность выборочной совокупности зависит от природы (характера) исследуемого социально-экономического явления. В выборочной совокупности должны быть представлены все типы единиц, имеющихся в исследуемой совокупности. В противном случае выборочная совокупность не будет точно воспроизводить пропорции и зависимости, характерные для совокупности во всем ее объеме.

Разновидностью выборочного наблюдения является *метод моментных наблюдений*. Суть его состоит в том, что информация собирается путем регистрации значений признаков у единиц выборочной совокупности в некоторые заранее определенные моменты времени. Поэтому метод моментных наблюдений предполагает отбор не только единиц исследуемой совокупности (выборку в пространстве), но и моментов времени, в которые проводится регистрация состояния исследуемого объекта (выборка во времени).

Этот вид наблюдения применяется при проведении обследований доходов населения.

Следующий вид несплошного наблюдения – это *метод основного массива.* При нем обследованию подвергаются самые существенные, обычно наиболее крупные единицы изучаемой совокупности, которые по основному (для конкретного исследования) признаку имеют наибольший удельный вес в совокупности. Именно этот вид используется для организации наблюдения за работой городских рынков.

Монографическое обследование представляет собой вид несплошного наблюдения, при котором тщательному обследованию подвергаются отдельные единицы изучаемой совокупности, обычно представители каких-либо новых типов явлений. Оно проводится с целью выявления имеющихся или намечающихся тенденций в развитии данного явления.

Монографическое обследование, ограничиваясь отдельными единицами наблюдения, изучает их с высокой степенью детализации, которой нельзя достигнуть при сплошном, или даже выборочном обследовании. Детальное статистико-монографическое изучение одного завода, фермы, бюджета семьи и т.д. позволяет уловить те пропорции и связи, которые ускользают из поля зрения при массовых наблюдениях.

Таким образом, при монографическом обследовании статистическому наблюдению подвергаются отдельные единицы совокупности, причем они могут представлять собой как действительно единичные случаи, так и совокупности малого размера. Монографическое обследование часто проводится для составления программы нового массового наблюдения. Можно сказать, что существует тесная связь между сплошным (или выборочным) и монографическим наблюдениями. С одной стороны, для отбора единиц наблюдения, которые должны быть подвергнуты монографическому изучению, используют данные массовых обследований. С другой – результаты монографических обследований дают возможность уточнить структуру исследуемой совокупности и, что очень важно, установить связь между отдельными признаками, характеризующими изучаемое явление. Это позволяет уточнить программу массового наблюдения, характерные черты и основные признаки объекта исследования.

2.5
ТОЧНОСТЬ НАБЛЮДЕНИЯ

Точностью статистического наблюдения называют степень соответствия величины какого-либо показателя (значение какого-либо признака), определенной по материалам статистического наблюдения, действительной его величине.

Расхождение между расчетным и действительным значениями изучаемых величин называется *ошибкой наблюдения*.

Точность данных – это основное требование, предъявляемое к статистическому наблюдению. Чтобы избежать ошибок наблюдения, предупредить, выявить и исправить их, необходимо:

- обеспечить качественное обучение персонала, который будет проводить наблюдение;
- организовать специальную частичную или сплошную контрольную проверку правильности заполнения статистических формуляров;
- провести логический и арифметический контроль полученных данных после окончания сбора информации.

В зависимости от причин возникновения различают ошибки регистрации и ошибки репрезентативности.

Ошибки регистрации – это отклонения между значением показателя, полученного в ходе статистического наблюдения, и фактическим, действительным его значением. Этот вид ошибок может быть и при сплошном, и при несплошном наблюдениях.

Ошибки регистрации бывают случайные и систематические.

Случайные ошибки – это результат действия различных случайных факторов (например, цифры переставлены местами, перепутаны соседние строки или графы при заполнении статистического формуляра). Такие ошибки имеют разную направленность: они могут и повышать, и понижать значения показателей. При достаточно большой обследуемой совокупности в результате действия закона больших чисел эти ошибки взаимно погашаются.

Систематические ошибки регистрации всегда имеют одинаковую тенденцию либо к увеличению, либо к уменьшению значения показателей по каждой единице наблюдения, и поэтому величина показателя по совокупности в целом будет включать в себя накопленную ошибку. Примером статистической ошибки регистрации при проведении социологических опросов населения может служить округление возраста населения, как правило, на цифрах, оканчивающихся на 5 и 0. Многие опрашиваемые, например, вместо 48–49 лет и 51–52 года говорят, что им 50 лет.

Ошибки репрезентативности характерны только для несплошного наблюдения. Они возникают потому, что отобранная и обследованная совокупность недостаточно точно воспроизводит (репрезентирует) всю исходную совокупность в целом.

Отклонение значения показателя обследованной совокупности от его величины по исходной совокупности называется *ошибкой репрезентативности.*

Ошибки репрезентативности также бывают случайные и систематические.

Случайные ошибки репрезентативности возникают, если отобранная совокупность неполно воспроизводит всю совокупность в целом. Ее величина может быть оценена.

Систематические ошибки репрезентативности появляются вследствие нарушения принципов отбора единиц из исходной совокупности, которые должны быть подвергнуты наблюдению.

После получения статистических формуляров следует прежде всего *провести проверку полноты собранных данных*, т.е. определить, все ли отчетные единицы заполнили статистические формуляры и значения всех ли показателей отражены в них.

Следующим этапом контроля точности информации является *арифметический контроль*. Он основывается на использовании количественных связей между значениями различных показателей. Например, если среди собранных данных имеются сведения о численности промышленно-производственного персонала, выработке товарной продукции в среднем на одного работающего и стоимости товарной продукции, то произведение первых двух показателей должно дать значение третьего показателя. Если арифметический контроль покажет, что данная зависимость не выполняется, это будет свидетельствовать о недостоверности собранных данных. Поэтому в программу статистического наблюдения целесообразно включать показатели, которые дают возможность провести арифметический контроль.

Логический контроль так же, как и арифметический, основывается на знании взаимосвязей между показателями, но не количественных, а логических. Например, человек в возрасте 6 лет не может иметь высшего образования. Поэтому если в бланке переписи имеются одновременно обе записи, то это показывает, что одна из них не соответствует действительности.

Обычно для исправления ошибок, выявленных в ходе логического контроля, требуется повторно обратиться к источнику сведений.

ОСНОВНЫЕ ПОНЯТИЯ

Документальный способ наблюдения – использование в качестве источника статистической информации различного рода документов, как правило, учетного характера.

Единица наблюдения – составной элемент объекта наблюдения, являющийся носителем признаков, подлежащих регистрации.

Единовременное обследование – сведения собираются о количественных характеристиках какого-либо явления или процесса в момент его исследования.

Критический момент (дата) – день года, час дня, по состоянию на который должна быть проведена регистрация признаков по каждой единице исследуемой совокупности.

Непосредственное наблюдение – регистраторы путем непосредственного замера, взвешивания, подсчета или проверки работы и так далее устанавливают факт, подлежащий регистрации, и на этой основе производят запись в формуляре наблюдения.

Несплошное наблюдение – обследованию подлежит лишь часть единиц изучаемой совокупности.

Объект наблюдения – статистическая совокупность, в которой протекают исследуемые социально-экономические явления и процессы.

Опрос – способ наблюдения, при котором наблюдаемые сведения получают со слов респондента.

Отчетная единица – субъект, от которого поступают данные о единице наблюдения.

Отчетность – основная форма статистического наблюдения, с помощью которой статистические органы в определенные сроки получают от предприятий, учреждений и организаций необходимые данные в виде установленных в законном порядке отчетных документов, скрепляемых подписями лиц, ответственных за их представление и достоверность собираемых сведений.

Ошибка наблюдения – расхождение между расчетным и действительным значениями изучаемых величин.

Перепись – специально организованное наблюдение, повторяющееся, как правило, через равные промежутки времени, с целью получения данных о численности, составе и состоянии объекта статистического наблюдения по ряду признаков.

Программа наблюдения – перечень признаков (или вопросов), подлежащих регистрации в процессе наблюдения.

Регистровое наблюдение – форма непрерывного статистического наблюдения за долговременными процессами, имеющими фиксированное начало, стадию развития и фиксированный конец.

Сплошное наблюдение – получение информации о всех единицах исследуемой совокупности.

Срок (период) наблюдения – время, в течение которого происходит заполнение статистических формуляров.

Статистический формуляр – документ единого образца, содержащий программу и результаты наблюдения.

Статистическое наблюдение – массовое, планомерное, научно организованное наблюдение за явлениями социальной и экономической жизни, которое заключается в регистрации признаков, отобранных у каждой единицы совокупности.

Текущее наблюдение – регистрация изменений в отношении изучаемых явлений по мере их наступления.

Точность статистического наблюдения – степень соответствия величин какого-либо показателя, определяемого по материалам статистического наблюдения, действительной его величине.

Цель наблюдения – получение достоверной информации для выявления закономерностей развития явлений и процессов.

ТЕСТЫ

1. Чем отличается статистическое наблюдение от наблюдения писателя, художника:

а) различием объекта наблюдения;
б) различием времени наблюдения;
в) научной организованностью и планомерностью;
г) различной целью наблюдения.

2. Объект статистического наблюдения – это:

а) единица наблюдения;
б) статистическая совокупность;
в) единица статистической совокупности;
г) отчетная единица.

3. Субъект, от которого поступают данные в ходе статистического наблюдения, называется:

а) единицей наблюдения;
б) единицей статистической совокупности;
в) отчетной единицей.

4. Перечень признаков (или вопросов), подлежащих регистрации в процессе наблюдения, называется:

а) статистическим формуляром;
б) программой наблюдения;
в) инструментарием наблюдения.

5. Срок наблюдения – это:

а) время, в течение которого происходит заполнение статистических формуляров;
б) конкретный день года, час дня, по состоянию на который должна быть проведена регистрация признаков по каждой единице исследуемой совокупности.

6. Статистическая отчетность – это:

а) вид статистического наблюдения;
б) способ статистического наблюдения;
в) форма статистического наблюдения.

7. Метод основного массива – это:

а) вид статистического наблюдения;
б) способ статистического наблюдения;
в) форма статистического наблюдения.

8. Перепись населения России (2002 г.) – это:

а) единовременное, специально организованное, сплошное наблюдение;
б) периодическое, специально организованное, сплошное наблюдение;
в) периодическое, регистровое, сплошное наблюдение;
г) периодическое, специально организованное, несплошное наблюдение;
д) единовременное, специально организованное, выборочное наблюдение;
е) периодическое, регистровое, выборочное наблюдение.

9. Обследование малых предприятий по итогам работы за 2000 г. – это:

а) текущее наблюдение;
б) периодическое наблюдение;
в) единовременное обследование.

10. Расхождение между расчетными значениями и действительным значением изучаемых величин называется:

а) ошибкой наблюдения;
б) ошибкой регистрации;
в) ошибкой репрезентативности.

ЛИТЕРАТУРА

Альбом наглядных пособий по общей теории статистики. – М.: Финансы и статистика, 1991. – 75 с.

Васильев К.И. К вопросу о методике и технике массовых наблюдений. – Л., 1929. – 74 с.

Джессен Р. Методы статистических обследований. – М.: Финансы и статистика, 1985. – 478 с.

Овсиенко В.Е. Статистическое наблюдение. – М., 1961. – 24 с.

Основные организационные положения и календарный план работ по подготовке, проведению и автоматизированной обработке материалов микропереписи населения 1994 г. – М., 1994. – 24 с.

Пономаренко А.Н., Башкатов Б.И. Система национального счетоводства: принципы построения. – М.: Экономика, 1992. – 47 с.

Рябушкин Б.Т., Хоменко Т.А. Система национальных счетов. – М.: Финансы и статистика, 1993. – 51 с.

ГЛАВА 3
СТАТИСТИЧЕСКАЯ СВОДКА И ГРУППИРОВКА

3.1 ЗАДАЧИ СВОДКИ И ЕЕ СОДЕРЖАНИЕ

На основе информации, собранной в ходе статистического наблюдения, как правило, нельзя непосредственно выявить и охарактеризовать закономерности социально-экономических явлений. Это связано с тем, что наблюдение дает сведения по каждой единице исследуемого объекта. Полученные данные не являются обобщающими показателями. С их помощью нельзя сделать выводы в целом об объекте без предварительной обработки данных.

Поэтому *цель следующего этапа статистического исследования* – систематизация первичных данных и получение на этой основе сводной характеристики всего объекта при помощи обобщающих статистических показателей.

Сводка представляет собой комплекс последовательных операций по обобщению конкретных единичных фактов, образующих совокупность, для выявления типичных черт и закономерностей, присущих изучаемому явлению в целом.

Таким образом, если при статистическом наблюдении собирают данные о каждой единице объекта, то результатом сводки являются подробные данные, отражающие в целом всю совокупность.

Статистическая сводка должна вестись на основе предварительного теоретического анализа явлений и процессов. Это необходимо для того, чтобы во время сводки не потерять информацию об исследуемом явлении; кроме того, все статистические итоги должны отражать характерные черты объекта.

Сводка может быть различной, в зависимости от ряда характеризующих ее признаков (рис.3.1).

По глубине обработки данных сводка бывает простая и сложная. *Простой сводкой* называется операция по подсчету общих итогов по совокупности единиц наблюдения или общего объема изучаемого

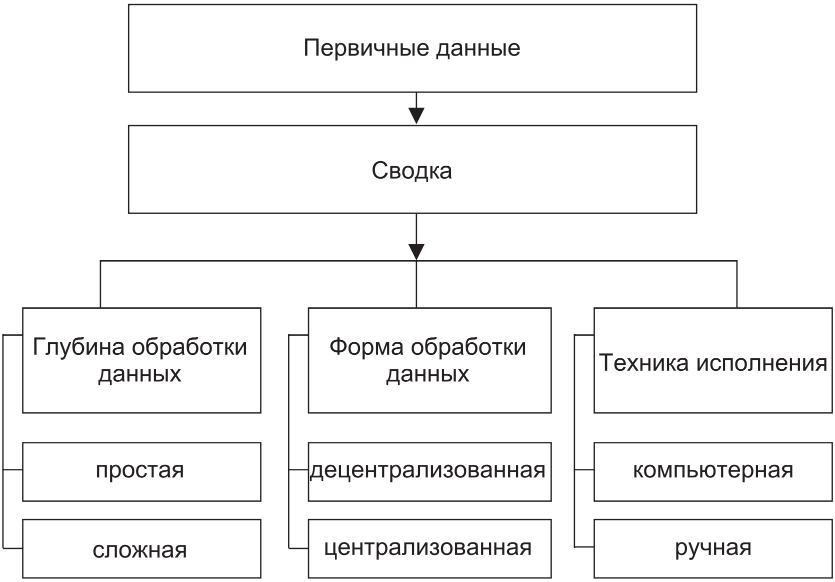

Рис. 3.1. Классификация видов статистической сводки

показателя. Например, чтобы получить общую численность студентов вузов в России, достаточно сложить данные по всем высшим учебным заведениям страны.

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

Проведению сводки предшествует разработка ее программы, которая состоит из следующих этапов:

- выбор группировочных признаков;
- определение порядка формирования групп;
- разработка системы статистических показателей для характеристики групп и объекта в целом;
- разработка системы макетов статистических таблиц, в которых должны быть представлены результаты сводки.

По *форме обработки данных* сводка бывает децентрализованная и централизованная.

При *децентрализованной сводке* (именно она используется, как правило, при обработке статистической отчетности) разработка материала производится последовательными этапами. Так, отчеты предприятий сводятся территориальными органами Госкомстата России, а уже итоги по региону поступают в Госкомстат России, и там определяются показатели в целом по народному хозяйству страны.

При *централизованной сводке* весь первичный материал поступает в одну организацию, где и подвергается обработке от начала и до конца. Централизованная сводка обычно используется для обработки материалов единовременных статистических обследований.

По *технике исполнения* сводка может быть компьютерной и ручной.

Компьютерная сводка – это способ выполнения сводки статистических данных, при котором все операции осуществляются с использованием компьютеров и программных продуктов, позволяющих обработать любые объемы информации с различной степенью детализации. При *ручной сводке* все основные операции (подсчет групповых и общих итогов) осуществляются вручную. В настоящее время в связи с широким применением персональных компьютеров, созданием автоматизированных рабочих мест, разработкой и внедрением новых информационных технологий ручная сводка в обработке информации используется крайне редко.

Для проведения сводки составляется план, в котором излагаются организационные вопросы: кем и когда будут осуществляться все операции, порядок ее проведения, состав сведений, подлежащих опубликованию в периодической печати.

3.2
МЕТОД ГРУППИРОВКИ И ЕГО МЕСТО В СИСТЕМЕ СТАТИСТИЧЕСКИХ МЕТОДОВ

Отдельные единицы статистической совокупности объединяются в группы при помощи метода группировки. Это позволяет «сжать» информацию, полученную в ходе наблюдения, и на этой основе выявить закономерности, присущие изучаемому явлению.

Группировкой называется разделение множества единиц изучаемой совокупности на группы по определенным существенным для них признакам. Группировка является одним из самых сложных в методологическом плане этапов статистического исследования.

Причины, обусловливающие необходимость проведения группировки и определяющие ее место в системе статистических методов, кроются в своеобразии объекта статистического исследования. Он представляет собой комплекс частных совокупностей, которые могут быть качественно и глубоко различны, обладать различными свойствами, степенью сложности, характером развития.

Невозможность статистической характеристики объекта исследования без выделения групп легко показать на примере совокупности промышленных предприятий. Каждое промышленное предприятие имеет индивидуальные особенности: год образования, место положения, состав установленного оборудования и т.д. Без преодоления этих индивидуальных черт исследовать закономерности развития промышленности, которые теряются в многочисленных характеристиках, отличающих одно предприятие от другого, нельзя. Поэтому предприятия следует объединить в группы по отрасли промышленности, назначению выпускаемой продукции, численности занятых и форме собственности и т.д. Таким образом, в показателях, исчисленных по достаточно большим группам, произойдет погашение случайного и выяление общего, существенного для развития исследуемого явления.

Итак, группировки являются важнейшим статистическим методом обобщения данных, основой для правильного исчисления статистических показателей.

С помощью метода группировок решаются следующие задачи:

- выделение социально-экономических типов явлений;
- изучение структуры явления и структурных сдвигов, происходящих в нем;
- выявление связи и зависимости между явлениями.

3.3. ВИДЫ СТАТИСТИЧЕСКИХ ГРУППИРОВОК

Статистические группировки можно классифицировать по следующим признакам: целям и задачам, числу группировочных признаков, упорядоченности исходных статистических данных (рис.3.2).

В зависимости от цели и задач исследования статистические группировки бывают типологическими, структурными и аналитическими.

Типологическая группировка – это разделение исследуемой качественно разнородной совокупности на классы, социально-экономи-

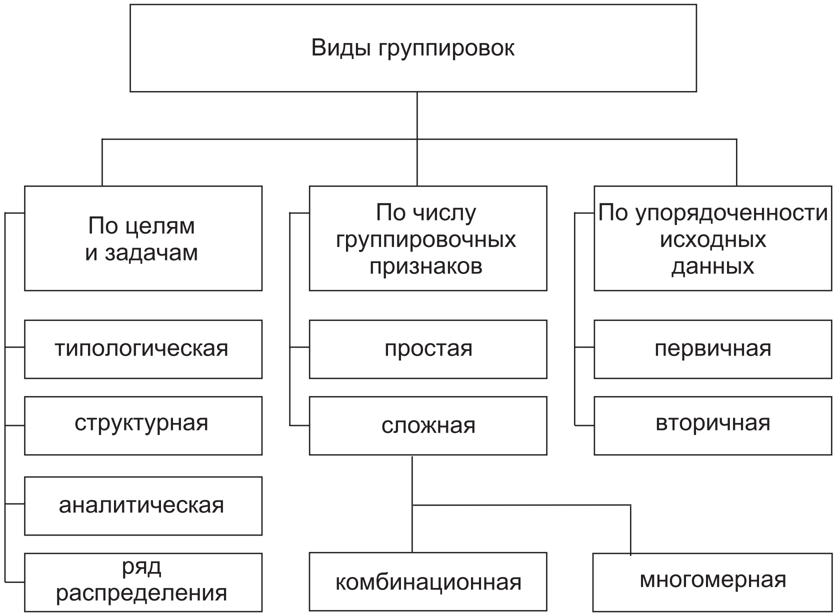

Рис. 3.2. Классификация видов статистических группировок

ческие типы, однородные группы единиц в соответствии с правилами научной группировки. При этом под однородностью понимается подчинение всех единиц совокупности одному закону развития в отношении рассматриваемого свойства. Схема типологической группировки изображена на рис.3.3.

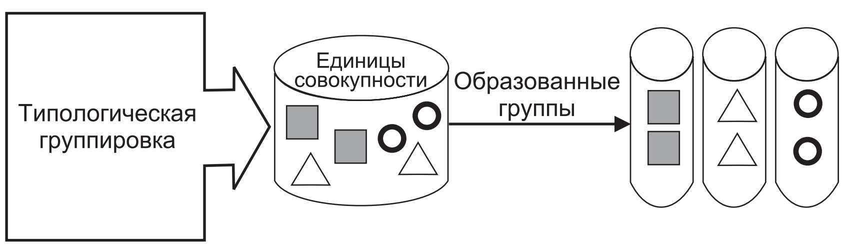

Рис. 3.3. Разделение разнородной совокупности на однородные группы

Типологические группировки позволяют проследить зарождение, развитие и отмирание различных типов явлений (развитие различных форм собственности, формирование новых слоев населения). Данный вид группировки также дает возможность выделить в составе массового явления те его части, которые однородны по качеству и условиям развития, в которых действуют одни и те же закономерности, на которые влияют одни и те же факторы.

При построении типологической группировки особое внимание уделяется идентификации типов и выбору группировочного признака. Вопрос об основании группировки должен решаться исходя из сущности изучаемого явления.

Однако формирование типов явлений связано с конкретными условиями места и времени. Сущность того или иного типа явления или процесса может проявляться и раскрываться как в одном, так и во множестве признаков. Одни и те же значения группировочных признаков в различных сочетаниях с другими признаками могут определять принадлежность единиц наблюдения то к одному, то к другому типу. Поэтому для выделения социально-экономических типов целесообразнее всего рассматривать не отдельные, изолированные признаки, а их совокупность, характеризующую изучаемое явление с различных сторон.

При построении типологической группировки в качестве группировочного признака могут выступать как количественные, так и атрибутивные (качественные) признаки.

Примером типологической группировки по атрибутивному признаку является группировка предприятий и организаций по формам собственности (табл.3.1).

Анализ данных табл.3.1 показал, что подавляющее большинство предприятий находится в частной собственности, т.е. 75%; предприятия государственной собственности составили 4,5%; предприятия муниципальной собственности – 6,5%; в собственности общественных объединений и прочих форм собственности находится 14% предприятий.

Информационная ценность типологической группировки повышается, если для оценки типов явлений используется не один, а несколько показателей, характеризующих каждую выделенную группу (табл.3.2).

Анализ данных табл.3.2 показал, что за три года число приватизированных предприятий всех форм собственности увеличилось в 1,1 раза, или на 6,8%, в том числе число приватизированных предприятий, находящихся в федеральной собственности и собственности

Таблица 3.1

Группировка предприятий и организаций по формам собственности в России (на январь 2001 г.)

№ п/п	Группы предприятий по формам собственности	Число предприятий	
		всего, тыс.	% к итогу
	А	1	2
1	Государственная	151	4,5
2	Муниципальная	217	6,5
3	Частная	2510	75,0
4	Собственность общественных и религиозных организаций (объединений)	223	6,7
5	Прочие формы собственности, включая смешанную российскую, иностранную, совместную российскую и иностранную	247	7,3
	Всего	3346	100

субъектов РФ, снизилось соответственно на 35,6 и 14,6%, а число приватизированных муниципальных предприятий возросло на 18,5%. Сумма средств, полученных от приватизации по всем формам собственности, возросла в 2,4 раза, в том числе от приватизации федеральной собственности – в 2,5 раза, собственности субъектов РФ – в 1,8 раза, а муниципальной собственности – в 1,7 раза.

Структурная группировка разделяет однородную в качественном отношении совокупность единиц по определенным, существенным признакам на группы, характеризующие ее состав и структуру. Схема структурной группировки изображена на рис. 3.4.

Структурные группировки применяются в изучении практически всех социально-экономических явлений и процессов. С их помощью исследуется состав населения по полу, возрасту, месту проживания; состав коммерческих банков по уставному фонду, капиталу, численности работающих и т.д. В качестве группировочных признаков, так же как и при построении типологической группировки, могут рассматриваться количественные и атрибутивные признаки. Примером структурной группировки по количественному признаку является группировка крестьянских хозяйств по размеру земельного участка (табл. 3.3).

Таблица 3.2

Группировка приватизированных предприятий (объектов) по формам собственности в Российской Федерации

(на 1 января 2001 г.)

№ п/п	Группы предприятий (объектов) по формам собственности	Число приватизированных предприятий (объектов)		Темп роста в 2000 г. по сравнению с 1998 г., %	Получено средств от приватизации, млн руб.		Темп роста в 2000 г. по сравнению с 1998 г., %
		1998	2000		1998	2000	
	А	1	2	3	4	5	6
1	Федеральная	264	170	64,4	15439,5	37927,5	245,7
2	Субъектов Российской Федерации	321	274	85,4	1250,8	2198,8	175,8
3	Муниципальная	1544	1830	118,5	847,8	1482,6	174,9
	Всего	2129	2274	106,8	17538,1	41608,9	237,2

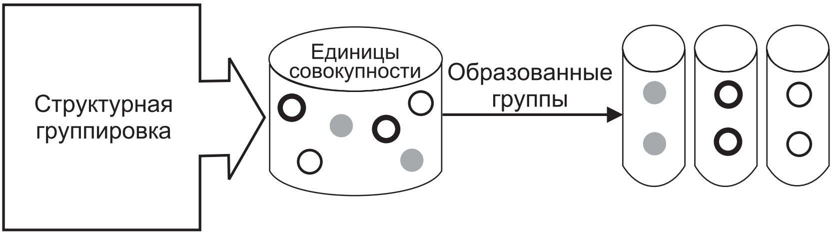

Рис. 3.4. Разделение однородной совокупности на однородные группы

Таблица 3.3

Группировка крестьянских хозяйств по размеру земельного участка в Ярославской области (по состоянию на 1 января 2000 г.)

Группы хозяйств с площадью земли, га	Число хозяйств		Площадь земли	
	единиц	% к итогу	га	% к итогу
А	1	2	3	4
Земля не предоставлялась	16	0,8	–	–
До 3	179	8,6	405	0,9
4–10	805	38,6	5744	12,9
11–20	484	23,2	8018	18,0
21–70	486	23,3	17936	40,3
71–100	72	3,4	5587	12,5
101–200	38	1,8	5071	11,4
201 и выше	6	0,3	1768	4,0
Итого	2 086	100,0	44529	100,0

Данные табл. 3.3 показывают, что 38,6% хозяйств имеют земельный участок 4–10 га, и им принадлежит всего около 13% земельной площади. Доли следующих двух групп составляют соответственно 23,2 и 23,3%. Именно этим крестьянским хозяйствам принадлежит около 40,3% земли. Число хозяйств в группах 101–200 га и 201 га и выше составляет всего лишь 2,1%, им принадлежит 15,4% земли (11,4 + 4,0).

При группировке по атрибутивному признаку группы отличаются друг от друга по характеру признака. Число групп, на которые делится изучаемая совокупность, как правило, определяется числом градаций атрибутивного признака. Это наглядно представлено в табл. 3.4.

Таблица 3.4

Распределение численности безработных по семейному положению в Российской Федерации в 2000 г.* (на конец ноября), %

№ п/п	Группы безработных по полу	Всего	В том числе			
			состоят в браке	холосты, не замужем	вдовы, вдовцы	разведены
	А	1	2	3	4	5
1	Мужчины	100	50,2	35,4	1,0	13,4
2	Женщины	100	56,6	22,6	4,7	16,1
	Итого	100	53,1	29,6	2,7	14,6

* По данным выборочного обследования населения по проблемам занятости.

Практическое применение структурных группировок позволяет на локальном уровне раскрыть структуру совокупности, проанализировать изучаемые явления и процессы, изменение их во времени и закономерности изменения состава совокупности во времени, если совокупности прослеживаются за ряд последовательных периодов времени.

Аналитическая группировка выявляет взаимосвязи и взаимозависимости между изучаемыми социально-экономическими явлениями и признаками, их характеризующими.

В статистике признаки делятся на факторные и результативные. *Факторными* называются признаки, под воздействием которых изменяются другие, *результативные,* признаки.

Особенностью аналитической группировки является то, что в основание группировки кладется факторный признак, затем подсчитывается количество единиц совокупности и общее суммарное значение результативного признака по каждой выделенной группе и даже производится расчет среднего значения результативного признака по выделенным группам. Взаимосвязь проявляется в том, что с возрастанием (убыванием) значения факторного признака систематически возрастает (убывает) среднее значение результативного признака. Результаты группировки излагаются в статистической таблице. Пример аналитической группировки представлен в табл. 3.5.

Данные табл. 3.5 показывают, что с ростом процентной ставки, под которую выдается кредит, средняя сумма кредита, выдаваемая одним банком, уменьшается! Это говорит о том, что между исследуемыми признаками существует обратная связь.

Таблица 3.5

Группировка зависимости суммы выданных кредитов коммерческими банками от размера процентной ставки

№ п/п	Группы банков по величине процентной ставки	Число банков	Сумма выданных кредитов, млн руб.	
			всего	в среднем на один банк
	А	1	2	3
1	11–15	7	168,1	24,0
2	15–19	13	200,5	15,4
3	19–23	7	54,4	7,8
4	23–27	3	6,8	2,3
	Итого	30	429,8	14,3

Деление рассмотренных группировок, в зависимости от цели и решаемых задач, на три вида носит условный характер, так как группировка может быть *универсальной*, т.е. одновременно выделяя типы, показывать структуру совокупности и отражать закономерности изменения значений признака в зависимости от другого.

По числу группировочных признаков различаются *простые* группировки (один признак) и *сложные* (два и более признаков). Пример простых группировок приведен в табл. 3.1, 3.2, 3.3 и 3.4.

Сложные группировки, в свою очередь, делятся на комбинационные (два – четыре признака) и многомерные (любое число признаков свыше четырех).

Принцип построения *комбинационной* группировки заключается в том, что сначала группы формируются по одному признаку, затем они делятся на подгруппы по другому признаку, а эти, в свою очередь, делятся по третьему и т.д. Данная группировка позволяет изучить единицы совокупности одновременно по нескольким признакам. При построении комбинационной группировки возникает вопрос о последовательности разбивки единиц объекта по признакам. Как правило, рекомендуется сначала производить группировку по атрибутивным признакам, значения которых имеют ярко выраженные качественные различия, а затем дополнять ее группировкой по количественным признакам.

Комбинационные группировки применяются, как правило, при изучении сложных социально-экономических явлений и процессов. Необходимым и обязательным условием построения данного вида группировок является наличие достаточно большого числа наблюдений. Дело в том, что комбинация группировочных признаков приводит к резкому увеличению числа групп. Численность же единиц в каждой из них может оказаться недостаточной.

В результате исследователь может прийти к малообоснованным выводам. Поэтому на практике строят комбинационные группировки не более чем по трем признакам.

Примером комбинационной группировки являются данные, приведенные в табл.3.6.

Сохранить сложность описания групп и преодолеть недостатки комбинационной группировки позволяют *методы многомерных группировок,* или иначе *многомерных классификаций*. Эти методы получили распространение благодаря использованию компьютеров и программных продуктов, позволяющих разрабатывать любые объемы информации с различной степенью детализации. Цель этих методов – классификация данных, т.е. группировка на основе множества признаков. Многомерные группировки позволяют решать целый ряд важных задач экономико-статистического исследования, таких, как формирование однородных совокупностей, выбор существенных признаков, выделение типичных групп объектов по множеству существенных признаков и др.

Таблица 3.6

Группировка коммерческих банков по величине кредитных вложений и объему вложений в ценные бумаги в одном из регионов*

№ п/п	Группы коммерческих банков по величине кредитных вложений, млн руб.	Подгруппы по объему вложений в ценные бумаги, млн руб.	Число банков	Капитал, млн руб.	Чистые активы, млн руб.	Ценные бумаги, млн руб.
	А	Б	1	2	3	4
1	До 52	2–55	10	881	1 298	12
		55–108	10	869	982	77
		108–161	–	–	–	–
		161–214	–	–	–	–
	Итого по группе		20	1 750	2 280	890
2	52–104	2–55	41	3 665	8 167	568
		55–108	11	978	1 923	788
		108–161	10	879	1971	1258
		161–214	–	–	–	–
	Итого по группе		62	5 522	10 261	2 614
3	104–156	2–55	40	3 802	10 049	1 546
		55–108	15	839	2 786	623
		108–161	–	–	–	–
		161–214	11	809	3 392	1 712
	Итого по группе		66	5 450	16 227	3 881

Продолжение

№ п/п	Группы коммерческих банков по величине кредитных вложений, млн руб.	Подгруппы по объему вложений в ценные бумаги, млн руб.	Число банков	Капитал, млн руб.	Чистые активы, млн руб.	Ценные бумаги, млн руб.
	А	Б	1	2	3	4
4	156–208	2–55	23	1 721	5 696	729
		55–108	12	958	4649	796
		108–161	–	–	–	–
		161–214	–	–	–	–
	Итого по группе		35	2 679	10 345	1 525
5	208 и выше	2–55	–	–	–	–
		55–108	–	–	–	–
		108–161	22	1 776	11 396	2 922
		161–214	11	977	11 598	2 112
	Итого по группе		33	2 753	22 994	5 034
	Итого по подгруппам	2–55	114	10 069	25 210	2 963
		55–108	48	3 644	10 340	2 977
		108–161	32	2 655	13 367	4 180
		161–214	22	1 786	14 990	3 824
	Всего		216	18 154	63 907	13 944

* Цифры условные.

3.4 ПРИНЦИПЫ ПОСТРОЕНИЯ СТАТИСТИЧЕСКИХ ГРУППИРОВОК И КЛАССИФИКАЦИЙ

Построение статистических группировок предполагает решение ряда основных задач. Прежде всего, необходимо выбрать группировочный признак, затем определить число групп, на которые нужно разбить изучаемую совокупность, и зафиксировать границы интервалов группировки. На завершающей стадии необходимо для каждой группировки найти конкретные показатели или их систему, которые должны характеризовать выделенные группы.

Выбор группировочного признака является одним из самых важных и сложных вопросов теории статистической группировки.

Группировочным признаком называется признак, по которому единицы совокупности разбиваются на отдельные группы. Его часто называют *основанием* группировки. От правильного выбора группировочного признака зависят выводы, которые получают в результате статистического исследования.

В качестве основания группировки необходимо использовать наиболее существенные признаки, которые теоретически обоснованы и отражают сущность изучаемых явлений в условиях поставленных целей и задач. В основание группировки могут быть положены как количественные, так и качественные признаки. Первые имеют числовое выражение (объем торгов, курс доллара в рублях, возраст человека, денежный доход семьи и т.д.), а вторые отражают состояние единицы совокупности: пол человека, его национальность, семейное положение, отраслевую принадлежность предприятия, его форму собственности и организационно-правовую форму и т.д.

После определения основания группировки следует решить вопрос о *количестве групп*, на которые надо разбить исследуемую совокупность. Число групп зависит от задач исследования и вида признака, положенного в основание группировки, численности совокупности, степени вариации признака.

При построении группировки по качественному (атрибутивному) признаку групп, как правило, будет столько, сколько имеется градаций, видов, состояний у этого признака. Например, в случае проведения группировки населения по полу можно образовать только две группы: мужчины и женщины.

Другой пример. В табл. 3.1 приведена группировка предприятий и организаций по формам собственности. Из нее видно, что группировка проведена по всем градациям форм собственности, возможным в условиях современной России (федеральная, муниципальная, частная, совместная).

От группировок следует отличать классификацию. *Классификацией* называется систематизированное распределение явлений и объектов на определенные группы, классы, разряды на основании их сходства и различия.

Отличительные черты классификации:

- в основе классификации лежит качественный признак;
- классификации стандартны: они устанавливаются органами государственной и международной статистики. Если в каждом конкретном исследовании строится своя группировка, то классификация едина для любого исследования независимо от того, проводят его органы государственной статистики или другие учреждения и ведомства (министерства, налоговые органы и т.п.);
- классификации устойчивы. Они остаются неизменными в течение длительного времени. Однако если появляются новые группы единиц, их классы, разряды, то в классификации вносятся соответствующие изменения и дополнения.

Классификация, предопределяя важнейшие признаки группировки единиц совокупности, является основой группировок. В классификации точно определены все возможные группы и имеются подробные указатели, которые помогают отнести любую единицу объекта в ту или иную группу в каждом конкретном случае.

Таким образом, *цель классификации* – однозначно идентифицировать единицы совокупности, обеспечить эффективный поиск информации и ее систематизацию, достичь сопоставимости с международными стандартами.

Если группировка проводится по количественному признаку, то необходимо тщательно изучить экономическую (социальную) сущность изучаемого явления. Лишь после этого в соответствии с задачами исследования можно решать вопрос о числе групп, близких по значению к варьирующему признаку единиц совокупности.

При небольшом объеме совокупности не следует образовывать большое число групп, так как группы будут малочисленными. Поэтому показатели, рассчитанные для таких групп, не будут представительными и не позволят получить адекватную характеристику исследуемого явления.

Часто группировка по количественному признаку имеет задачу отразить распределение единиц совокупности по этому признаку. В данном случае количество групп зависит в первую очередь от степени колеблемости группировочного признака: чем больше его колеблемость, тем больше следует образовать групп. (Степень колеблемости признака измеряется показателями вариации, которые подробно рассмотрены в гл. 7.) Чем больше групп, тем точнее будет воспроизведен характер исследуемого объекта. Однако слишком большое число групп затрудняет выявление закономерностей при исследовании социально-экономических явлений и процессов. Поэтому в каждом конкретном случае при определении числа групп следует исходить не только из степени колеблемости признака, но еще учитывать и особенности объекта, и цель исследования.

При использовании персональных компьютеров для обработки статистических данных группировка единиц объекта проводится с помощью стандартных процедур.

Одна из таких процедур основана на использовании следующей формулы Стерджесса для определения оптимального числа групп:

$$n = 1 + 3{,}322 \cdot \lg N, \qquad (3.1)$$

где n – число групп;
N – число единиц совокупности.

Согласно формуле (3.1) выбор числа групп зависит от объема совокупности.

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

Другой способ определения числа групп основан на применении показателя среднего квадратического отклонения (σ). При этом весь диапазон изменения показателя предполагается равным. Если величина интервала равна $0{,}5\sigma$, то совокупность разбивается на 12 групп, а когда величина интервала равна $2/3\sigma$ и σ, то совокупность делится соответственно на 9 и 6 групп.

Однако при определении числа групп данными методами существует большая вероятность получения «пустых» или малочисленных групп. «Пустыми» считаются группы, в которые не попала ни одна единица совокупности. Поэтому данными формулами нельзя пользоваться механически. Их показания требуют корректировки.

После определения числа групп решается задача определения интервалов группировки.

Интервал группировки – это интервал значений варьирующего признака, лежащих в пределах определенной группы. Каждый интервал имеет свою ширину, верхнюю и нижнюю границы или хотя бы одну из них. *Нижней границей* интервала называется наименьшее значение признака в интервале, а *верхней границей* – наибольшее значение признака в нем. *Ширина интервала* (ее еще часто называют интервальной разностью) представляет собой разность между верхней и нижней границами интервала.

Интервалы группировки, в зависимости от их величины, бывают равные и неравные. Последние делятся на прогрессивно возрастающие, прогрессивно убывающие, произвольные и специализированные.

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

Величина равного интервала определяется по следующей формуле:

$$h = \frac{R}{n}, \qquad (3.2)$$

где $R = X_{max} - X_{min}$ – размах вариации;
X_{max}, X_{min} – максимальное и минимальное значения признака в совокупности;
n – число групп.

Прежде чем определять размах вариации, из совокупности рекомендуется исключить аномальные наблюдения; это значит, если максимальные или минимальные значения сильно отличаются от смежных с ними значений вариантов в упорядоченном ряду значений группировочного признака, то для определения величины интервала следует использовать не максимальное и минимальное значения, а значения, несколько превышающие минимум и несколько меньшие, чем максимум.

Полученную по формуле (3.2) величину называют *шагом интервала*.

Существуют следующие правила его определения.

Если шаг интервала, рассчитанный по формуле (3.2), представляет собой величину, имеющую один знак до запятой (например, 0,66; 1,372; 5,8), то в этом случае полученные значения целесообразно округлить до десятых долей. В приведенном выше примере шагом интервала будут соответственно значения 0,7; 1,4; 5,8.

Когда рассчитанный шаг интервала имеет две значащие цифры до запятой и несколько знаков после запятой, то это значение надо округлить до целого числа. Пусть величина интервала, исчисленная по формуле (3.2), равна 12,785. Тогда это значение следует округлить до целого числа, т.е. до 13.

В случае, когда рассчитанный шаг интервала представляет собой трехзначное, четырехзначное и так далее число, эту величину необходимо округлить до ближайшего числа, кратного 100 или 50. Например, 248 следует округлить до 250.

Пример. Пусть требуется произвести группировку с равными интервалами предприятий по стоимости основных фондов; при этом максимальное значение признака равно 2040 млн руб., а минимальное его значение – 290 млн руб. Совокупность включает 80 единиц. Согласно формуле (3.1) она должна быть разбита на 7 групп. Сначала найдем размах вариации (R):

$$R = 2040 - 290 = 1750 \text{ млн руб.}$$

Затем определим величину интервала:

$$h = 1750 : 7 = 250 \text{ млн руб.}$$

После этого построим интервалы групп (табл. 3.7).

Чтобы не писать каждый раз от ... до, границы групп обозначают следующим образом: 290–540, 540–790 и т.д.

Особенностью 1-го варианта построения групп является то, что у всех групп имеются закрытые интервалы. Во 2-м варианте первая и последняя группы — это группы с открытыми интервалами.

Открытые – это те интервалы, у которых указана только одна граница: верхняя – у первого, нижняя – у последнего. Ширина открытого интервала принимается равной ширине смежного с ним интервала.

Закрытыми называются интервалы, у которых обозначены обе границы.

Таблица 3.7

Варианты построения интервалов групп

Группа	Интервал	
	1-й вариант	2-й вариант
I	от 290 до 540	До 540
II	от 540 до 790	540 – 790
III	от 790 до 1040	790 – 1040
IV	от 1040 до 1290	1040 – 1290
V	от 1290 до 1540	1290 – 1540
VI	от 1540 до 1790	1540 – 1790
VII	от 1790 до 2040	1790 и более

При группировке по количественному признаку границы интервалов могут быть обозначены по-разному. Если основанием группировки служит непрерывный признак, то одно и то же значение признака выступает и верхней, и нижней границами у двух смежных интервалов. Таким образом, верхняя граница i-го интервала равна нижней границе i+1-го интервала. Примером такой группировки служат интервалы, приведенные в табл.3.7.

При таком обозначении границ может возникнуть вопрос, в какую группу включать единицы объекта, значения признака у которых совпадают с границами интервалов. Например, во вторую или третью группу должно войти предприятие со стоимостью фондов 790 млн руб. Если нижняя граница формируется по принципу «включительно», а верхняя – по принципу «исключительно», то предприятие должно быть отнесено к третьей группе, в противном случае – ко второй. Для того чтобы правильно отнести к той или иной группе единицу объекта, у которой значение признака совпадает с границами интервалов, можно использовать открытые интервалы.

Если в основании группировки лежит дискретный признак, то нижняя граница i-го интервала равна верхней границе i–l-го интервала, увеличенной на 1.

Пример. Пусть совокупность состоит из 80 предприятий и ее надо разделить на группы по численности занятых. Минимальное и максимальное значения группировочного признака соответственно равны 290 и 2040 человек. В этом случае возможны следующие варианты построения групп (табл. 3.8).

Таблица 3.8

Варианты построения интервалов групп

Группа	Интервал	
	1-й вариант	2-й вариант
I	290 – 540	До 541
II	541 – 790	541 – 790
III	791 – 1040	791 – 1040
IV	1041 – 1290	1041 – 1290
V	1291 – 1540	1291 – 1540
VI	1541 – 1790	1541 – 1790
VII	1791 – 2040	1791 и более

Неравные интервалы применяются в статистике, когда значения признака варьируют неравномерно и в значительных размерах, что характерно для большинства социально-экономических явлений, особенно при анализе макроэкономических показателей.

Неравные интервалы могут быть прогрессивно возрастающие или убывающие в арифметической или геометрической прогрессии. Величина интервалов, изменяющихся в арифметической прогрессии, определяется следующим образом:

$$h_{i+1} = h_i + a; \qquad (3.3)$$

в геометрической прогрессии:

$$h_{i+1} = h_i \cdot q, \qquad (3.4)$$

где a–константа – число, которое будет положительным при прогрессивно возрастающих интервалах и отрицательным при прогрессивно убывающих интервалах;

q–константа – положительное число, которое при прогрессивно возрастающих интервалах будет больше 1, а при прогрессивно убывающих – меньше 1.

Пример. Если необходимо построить группировку предприятий отрасли по показателю выручки от реализации продукции, который варьирует от 500 млн руб. до 4 000 млн руб., то строить группировку с равными интервалами нецелесообразно, потому что, как правило, совокупность предприятий любой отрасли промышленности, торговли включает большое число малых предприятий, имеющих небольшую выручку. С ростом выручки от реализации продукции значительно снижается число предприятий. Таким образом, распределение числа предприятий по величине выручки является неравномерным. Поэтому следует построить группировку с неравными интервалами (табл. 3.9).

Таблица 3.9

Построение групп с неравными интервалами

Группа	Интервал
I	500–800
II	800–1300
III	1300–2000
IV	2000–2900
V	2900–4000

Величина каждого последующего интервала у этой группировки больше предыдущего на 200 млн руб., т.е. увеличивается в арифметической прогрессии.

При определении границ интервалов статистических группировок исходят из того, что изменение количественного признака приводит к появлению нового качества. В этом случае граница интервала устанавливается там, где происходит переход от одного качества к другому. Рамки границ зависят от условий места и времени. Например, группировка предприятий по числу занятых показывает, что в промышленности и строительстве предприятия со среднесписочной численностью работающих 75–100 чел. относятся к группе малых предприятий, а в отраслях непроизводственной сферы и в розничной торговле – к крупным.

Поэтому, строя такую группировку, следует дифференцированно устанавливать границы интервалов для разных отраслей народного хозяйства. Это достигается путем использования группировок со специализированными интервалами. *Специализированными* называются интервалы, применяющиеся для выделения из совокупности одних и тех же типов по одному и тому же признаку для явлений, находящихся в различных условиях.

При изучении социально-экономических явлений на макроуровне часто применяют группировки, интервалы которых не будут ни прогрессивно возрастающими, ни прогрессивно убывающими. Такие интервалы называются *произвольными*. Произвольные интервалы часто используются при группировке рабочих по выработке продукции, предприятий по уровню рентабельности, прибыльности и др.

После определения группировочного признака и границ групп строится ряд распределения.

3.5 РЯДЫ РАСПРЕДЕЛЕНИЯ И ГРУППИРОВКИ

Рядом распределения в статистике называется ряд цифровых показателей, представляющих распределение единиц совокупности по одному существенному признаку, разновидности которого расположены в определенной последовательности.

Ряды распределения обычно входят в качестве составной части сводной обработки данных, при которой группы единиц совокупности характеризуются многими показателями, например, группы промышленных объединений характеризуются численностью промышленно-производственного персонала, величиной продукции, средней выработкой на одного работающего и т.п.

Однако в ряде случаев ряды распределения приобретают самостоятельное значение. Они строятся с целью изучения состава исследуемой совокупности, ее однородности, колеблемости значений признаков и границ их изменения. На основе рядов распределения рассчитываются относительные величины структуры, средние показатели, устанавливается типичность обобщающих показателей с позиций наблюдаемых единиц совокупности.

По своей конструкции ряд распределения состоит из двух элементов: *вариантов* (групп по выделенному признаку) и *частот* (численности групп). Частоты, выраженные в виде относительных величин (доли единиц, процентов), называются *частостями*. Сумма всех частот называется *объемом распределения,* или его *численностью*.

Сумма частостей равна 1, если они выражены в долях единицы, и 100%, если они выражены в процентах. Он оформляется в виде статистической таблицы. Общая схема ряда распределения такова: в совокупности, состоящей из N единиц, некоторая переменная величина x (т.е. какой-либо варьирующий признак) принимает различные значения $x_1, x_2, x_3, \dots, x_n$. Каждое из этих значений имеет частоту $f_1, f_2, f_3, \dots, f_n$. Исходя из этого вариационный ряд распределения можно представить в следующем виде:

Вариант x_i	Частота f_i
x_1 x_2 x_3 . . . x_n	f_1 f_2 f_3 . . . f_n
Итого	$\sum_i f_i$ (или N)

Ряды распределения, являясь группировкой, могут быть образованы по качественному (атрибутивному) и количественному (прерывному или непрерывному) признакам. В первом случае они называются *атрибутивными*, во втором – *вариационными*.

Примером атрибутивного ряда могут служить данные табл. 3.10.

Вариационные ряды распределения по способу построения бывают дискретные и интервальные.

Дискретный вариационный ряд распределения. В этом ряду группы составлены по признаку, изменяющемуся дискретно и принимающему только целые значения. Примером данного ряда является распределение семей по числу детей в одном из районов города (табл. 3.11).

В табл. 3.11 представлены варианты дискретного вариационного ряда, в графе 1 – частоты, в графе 2 – частости. В случае непрерывной вариации величина признака у единиц совокупности может принимать в определенных пределах любые значения, отличающиеся друг от друга на сколько угодно малую величину.

Таблица 3.10

Распределение оборота розничной торговли по формам собственности в России
(в фактически действующих ценах)

№ п/п	Группы оборота по формам собственности	Млрд руб.		% к итогу	
		1998	2000	1998	2000
	А	1	2	3	4
1	Негосударственная	990,8	2149	93,0	95,5
	из нее частная	852,3	1845	80,0	82,0
2	Государственная	74,4	102	7,0	4,5
	Всего оборота розничной торговли	1065,2	2251	100,0	100,0

Таблица 3.11

Распределение семей по числу детей в одном из районов города*

№ п/п	Группы семей по числу детей x	Число семей		Накопленные частоты S
		тыс. f	% к итогу w	
	А	1	2	3
1	0	6	5,9	6
2	1	28	27,5	34
3	2	22	21,6	56
4	3	20	19,6	76
5	4	13	12,7	89
6	5	8	7,8	96
7	6 и более	5	4,9	102
	Итого	102	100,0	–

*Цифры условные.

Интервальный вариационный ряд распределения. В этом ряду в группировочный признак, составляющий основание группировки, может принимать в определенном интервале любые значения. Данный ряд распределения целесообразно строить прежде всего при непрерывной вариации признака, а также если дискретная вариация проявляется в широких пределах, т.е. число вариантов дискретного признака достаточно велико.

Правила и принципы построения интервальных рядов распределения аналогичны правилам и принципам построения статистических группировок. В случае, если интервальный вариационный ряд распределения построен с равными интервалами, частоты позволяют судить о степени заполнения интервала единицами совокупности. При построении неравных интервалов нельзя получить информацию о степени заполнения каждого интервала.

Примером интервального вариационного ряда распределения с равными интервалами могут служить данные табл. 3.12.

Таблица 3.12

Распределение семей по размеру жилой площади, приходящейся на одного человека

№ п/п	Группы семей по размеру жилой площади, приходящейся на одного человека, м^2 x	Число семей с данными размерами жилой площади f	Накопленное число семей S
	А	1	2
1	3–5	10	10
2	5–7	20	30
3	7–9	30	60
4	9–11	40	100
5	11–13	15	115
	Итого	115	–

Если вариационный ряд распределения имеет группы с неравными интервалами (табл. 3.13), то частоты в отдельных интервалах непосредственно несопоставимы, так как зависят от ширины интервала. Для того чтобы частоты можно было бы сравнивать, исчисляют *плотность распределения*. Можно рассчитать как абсолютную, так и относительную плотность распределения.

Таблица 3.13

Распределение действующих кредитных организаций по величине зарегистрированного уставного капитала в России
(по состоянию на 01.07.2001 г.)

№ п/п	Группы кредитных организаций по уставному капиталу, млн руб.	Ширина интервала, млн руб. h_i	Число кредитных организаций f_i	Плотность распределения $f_o = \frac{f_i}{h_i}$,
	А	1	2	3(гр.2:гр.1)
1	До 3	2	150	75
2	3–10	7	254	36,3
3	10–30	20	316	15,8
4	30–60	30	256	8,5
5	60–150	90	144	1,6
6	150–300	150	90	0,6
7	300 и выше	300	112	0,37
	Итого	–	1322	–

Абсолютная плотность распределения – это частота, приходящаяся на единицу длины интервала, т.е. f_i/h_i, а *относительная плотность распределения* – частость, приходящаяся на единицу длины интервала, т.е. w_i/h_i.

Плотность распределения используется в рядах распределения с неравными интервалами для расчета моды (см. раздел 7.2), или для графического изображения вариационного ряда в виде гистограммы.

Для расчета плотности распределения обратимся к данным табл. 3.13. В ней дана группировка кредитных учреждений по уставному капиталу с неравными интервалами. Эта группировка, как видим, хорошо выявляет характер распределения. Однако неравные интервалы не позволяют сравнивать частоты в разных интервалах. В самом деле, в 1-й группе (графа 2) 150 кредитных организаций попадают на интервал 2 млн руб. (примем, что нижняя граница первого интервала равна 1 млн руб., а верхняя последнего – 600 млн руб.), а в 4-й группе 256 кредитных организаций попадают в интервал 30 млн руб. Число организаций в 4-й группе больше, чем в 1-й, а плотность (в расчете на ширину интервала 30 млн руб.) будет равна 8,5 кредитных организаций (256:30), то есть значительно меньше. Показатели плотности распределения приведены в табл. 3.13, графа 3.

Для различных целей бывает уместным осуществлять еще одно преобразование ряда распределения, заключающееся в построении *ряда накопленных частот* (кумулятивного ряда). Этот ряд показывает число случаев ниже или выше определенного уровня. Отсюда и возникают два варианта в построении ряда накопленных частот: один показывает число случаев, менее определенного значения варьирующего признака, а другой – число случаев, превышающее определенное значение варьирующего признака. Например, по приведенному в табл. 3.11 распределению семей по числу детей, меньше 1 ребенка имеют 6 тыс. семей; меньше 2 детей (т.е. 0 и 1-го) – 6 + 28 = 34 тыс. семей; меньше 3 детей (т.е. 0, 1-го и 2-х) – 6 + 28 + 22 = 56 тыс. семей; менее 4 детей (т.е. 0, 1-го, 2-х, 3-х) – 6 + 28 + 22 + 20 = 76 тыс. семей и т.д. Таким путем получаем числа: 6, 34, 56, 76, 89, 96, 102 (последнее совпадает с объемом совокупности). Накопленные частоты приведены в табл. 3.11, графа 3.

Аналогично определяются накопленные частости (последовательным суммированием частостей всех вариантов или интервалов). Накопленная частость показывает долю членов совокупности, у которых интересующий нас признак меньше данного значения.

Графическое изображение рядов распределения. Графическое изображение облегчает анализ ряда распределения и позволяет судить о форме распределений единиц совокупности по значениям группировочного признака.

Для изображения вариационных рядов применяются *линейные* и *плоскостные диаграммы* (см. глава 5, раздел 5.2), построенные в прямоугольной системе координат: полигон, гистограмма, огива, кумулята и кривая Лоренца.

Полигон используется при изображении дискретных вариационных рядов. Он представляет собой замкнутый многоугольник, абсциссами вершин которого являются значения варьирующего признака, а ординатами – соответствующие им частоты или частости.

Для примера рассмотрим построение полигона распределения по данным табл. 3.11. Отложив в выбранном масштабе число детей по оси абсцисс (x_i), а соответствующее им число семей по оси ординат (f_i), получаем совокупность точек. Соединим эти точки последовательно отрезками прямой, а из первой и последней точки опустим перпендикуляры на ось x, в результате получим замкнутую фигуру в виде многоугольника (рис. 3.5).

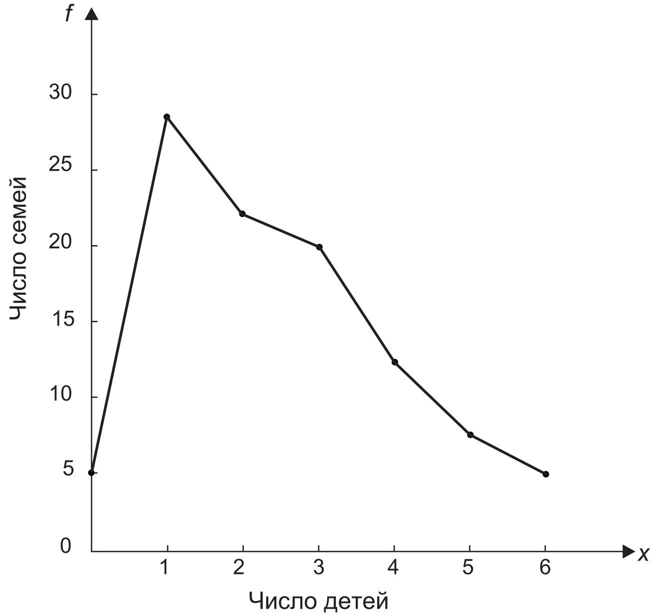

Рис. 3.5. Распределение семей по числу детей в одном из районов города

Гистограмма (гр. histos – ткань, строение) применяется для изображения интервального вариационного ряда, который представляют столбики с основаниями, равными ширине интервалов, и высотой, соответствующей частоте. По сути гистограмма является одной из разновидностей столбиковых диаграмм.

Изобразим графически интервальный вариационный ряд распределения, приведенный в табл. 3.12 (рис. 3.6).

Гистограмма может быть преобразована в полигон распределения, если найти середины сторон прямоугольников и затем эти точки соединить прямыми линиями. Полученный полигон распределения изображен на рис.3.6 пунктирной линией.

При построении гистограммы распределения вариационного ряда с неравными интервалами по оси ординат наносят не частоты, а плотность распределения признака в соответствующих интервалах.

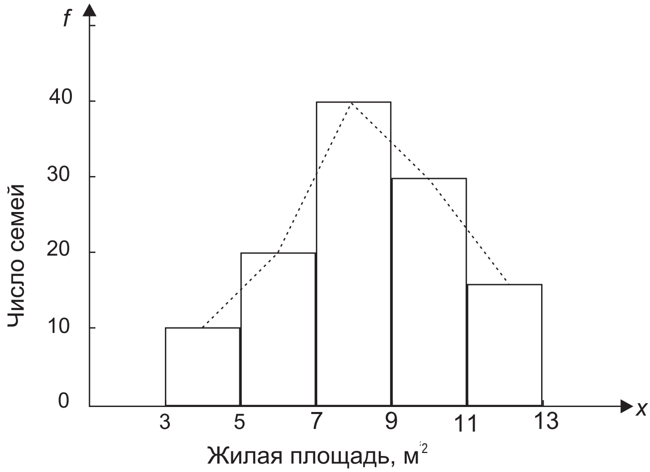

Рис. 3.6. Гистограмма распределения семей по размеру жилой площади, приходящейся на одного человека

Для графического изображения вариационных рядов может также использоваться *кумулятивная кривая*. При помощи кумуляты изображается ряд накопленных частот.

При построении кумуляты интервального вариационного ряда по оси абсцисс откладываются варианты ряда, а по оси ординат накопленные частоты, которые наносят на поле графика в виде перпендикуляров к оси абсцисс в верхних границах интервалов. Затем эти перпендикуляры соединяют и получают ломаную линию, т.е. кумуляту.

Используя данные накопленного ряда (графа 2 табл. 3.12), построим кумуляту распределения (рис.3.7).

Изображение вариационного ряда в виде кумуляты особенно эффективно для вариационных рядов, частоты которых выражены в долях или процентах к сумме частот ряда.

Если в прямоугольной системе координат построим точки, ординаты которых – варианты, а абсциссы – накопленные частоты (или частости), а затем соединим их отрезки прямой, то получим ломаную линию, которая называется *огивой*. Огива распределения семей по размеру жилой площади, приходящейся на одного человека, приведена на рис.3.8.

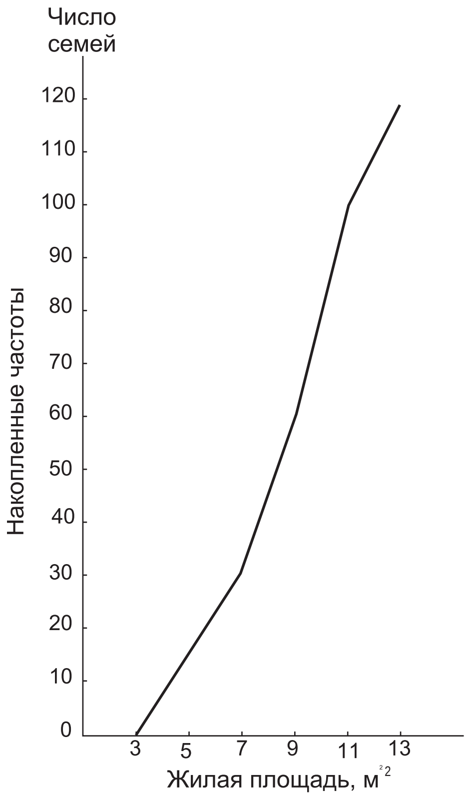

Рис. 3.7. Кумулята распределения семей по размеру жилой площади, приходящейся на одного человека

Разновидностью кумулятивной кривой является график–кривая Лоренца. График используется для характеристики процессов концентрации, дифференциации, специализации и т.д. Построение кривой Лоренца дано в главе 11, раздел 1.4.

Широкое применение компьютеров в практической работе экономиста облегчает построение рядов распределения, как и их графическое представление. Особо в этой связи следует отметить использование стандартизованных процедур определения величины интервала ряда распределения.

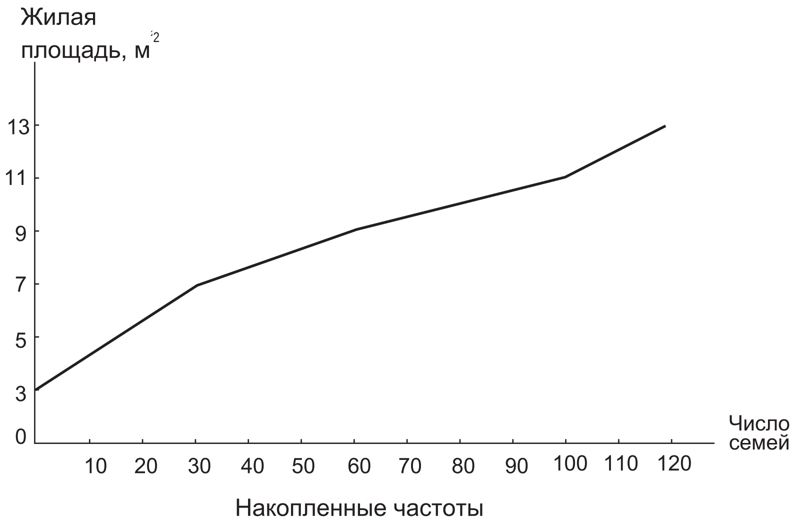

Рис. 3.8. Огива распределения семей по размеру жилой площади, приходящейся на одного человека

3.6 СРАВНИМОСТЬ СТАТИСТИЧЕСКИХ ГРУППИРОВОК

Группировки, построенные за один и тот же период времени, но для разных регионов или, наоборот, для одного региона, но за два разных периода времени, могут оказаться несопоставимыми из-за различного числа выделенных групп или неодинаковости границ интервалов. Для того чтобы привести такие группировки к сопоставимому виду (это позволяет провести их сравнительный анализ), используется метод вторичной группировки. Суть метода состоит в перегруппировке единиц объекта без обращения к первичным данным.

Вторичная группировка – операция по образованию новых групп на основе ранее построенной группировки.

Применяют два способа образования новых групп. Первым, наиболее простым и распространенным способом является *объединение первоначальных интервалов.* Способ используется, когда нужен переход от мелких интервалов к более крупным интервалам, а также когда границы новых и старых интервалов совпадают. Второй способ получил название *долевой перегруппировки;* он состоит в образовании новых групп на основе закрепления за каждой группой определенной доли единиц совокупности. Этот способ употребляется, когда необходимо в ходе перегруппировки данных определить, какая часть (доля) единиц совокупности перейдет из старых групп в новые.

Пример. Рассмотрим первый способ проведения вторичной группировки объединением первоначальных интервалов. Возьмем две группировки кредитов по сроку выдачи за ноябрь и декабрь (табл. 3.14 и 3.15).

Таблица 3.14

Группировка кредитов коммерческих банков по сроку выдачи в ноябре 2001 г. *

№ п/п	Группы кредитов по сроку выдачи, месяцев	Число заключенных договоров, % от их общего количества	Сумма выданных кредитов, % от общей суммы
	А	1	2
1	1–3	87,05	66,87
2	3–6	10,43	24,86
3	6–12	1,80	8,17
4	Более 12	0,72	0,10
	Итого	100,00	100,00

* Цифры условные.

При анализе двух группировок прежде всего их результаты необходимо привести к сопоставимому виду, перегруппировав данные первой группировки. Для этого данные (табл. 3.14) 1-й и 2-й групп объединяют вместе, образуя одну группу краткосрочных кредитов. В эту группу включают все кредиты, выданные в ноябре на срок от 1 до 6 месяцев. Данные 3-й группы (среднесрочные кредиты) и 4-й группы (долгосрочные кредиты) полностью переносятся в табл. 3.16, в которой представлены результаты вторичной группировки кредитов коммерческих банков, выданных в ноябре и декабре.

Таблица 3.15

Группировка кредитов коммерческих банков по сроку выдачи в декабре 2001 г.*

№ п/п	Группы кредитов по сроку выдачи, месяцев	Число заключенных договоров, % от их общего количества	Сумма выданных кредитов, % от общей суммы
	А	1	2
1	Краткосрочные (1–6)	86,54	97,91
2	Среднесрочные (6–12)	1,92	1,70
3	Долгосрочные (более 12)	11,54	0,39
	Итого	100,00	100,00

* Цифры условные.

Теперь можно сравнить итоги группировки ноябрьских и декабрьских кредитов. Доля заключенных договоров по краткосрочным кредитам снизилась почти на 11 проц. пунктов, доля среднесрочных кредитов осталась без изменения, а количество долгосрочных кредитов в анализируемом периоде значительно возросло. Несмотря на эти изменения, в декабре, так же как и ноябре, в структуре выданных кредитов преобладающую долю занимали краткосрочные кредиты, затем следовали среднесрочные и, наконец, – долгосрочные кредиты.

Более сложным является способ долевой перегруппировки. Рассмотрим его применение.

Пример. В табл. 3.17 приведены данные о распределении семей по размеру площади, приходящейся на одного человека по двум регионам.

Как видно из табл. 3.17, семьи первого региона разбиты на семь групп, а второго – на пять. Чтобы привести данные к сопоставимому виду, произведем перегруппировку семей второго региона. Для этого придется раздробить группы. Так как границы 1-й группы одинаковы у обеих группировок, то проведение каких-либо изменений нецелесообразно. 2-ю группу (5–10) необходимо разделить на три группы: семьи, в которых на одного человека приходится 5 и 6 м^2, должны образовать 2-ю группу; семьи, где на человека приходится 7 и 8 м^2, – 3-ю группу, а где 9–10 м^2 следует включить в 4-ю группу. Таким об-

Таблица 3.16

Группировка кредитов коммерческих банков по сроку выдачи в ноябре–декабре 2001 г.*

№ п/п	Группы кредитов по сроку выдачи, месяцев	Число заключенных договоров, % от их общего количества		Сумма выданных кредитов, % от общей суммы	
		ноябрь	декабрь	ноябрь	декабрь
	А	1	2	3	4
1	Краткосрочные (1–6)	97,48	86,54	91,73	97,91
2	Среднесрочные (6–12)	1,80	1,92	8,17	1,70
3	Долгосрочные (более 12)	0,72	11,54	0,10	0,39
	Итого	100,00	100,00	100,00	100,00

* Цифры условные.

Таблица 3.17

Группировка семей по размеру жилой площади, приходящейся на одного человека по двум регионам в 2001 г.

Первый регион			Второй регион		
№ п/п	группы семей по размеру жилой площади, приходящейся на 1 чел., $м^2$	доля семей, % к итогу	№ п/п	группы семей по размеру жилой площади, приходящейся на 1 чел., $м^2$	доля семей, % к итогу
1	До 5	3,6	1	До 5	6,2
2	5–6	11,4	2	5–10	46,3
3	7–8	19,4	3	11–15	28,5
4	9–12	37,8	4	16–19	10,8
5	13–14	11,1	5	20 и более	8,2
6	15–19	13,0			
7	20 и более	3,7			
	Итого	100,0		Итого	100,0

разом, 2-ю группу в группировке семей второго региона следует разбить на три равные по величине интервала группы. При разбивке семей по группам полагают, что их распределение внутри группы 5–10 равномерное. Тогда 1/3 семей группы 5–10 войдет в группу 5–6; 1/3 – в группу 7–8, а оставшаяся часть должна быть включена в группу 9–12. Кроме того, в эту группу следует включить и часть семей из следующей 3-й группы (11–15), т.е. семьи, в которых приходится 11 и 12 м^2 жилой площади на одного человека. Поэтому 40% семей 3-й группы надо включить в группу 9–12. Для составления группы 13–14 необходимо взять 40% семей группы 11–15. В группу 15–19 войдут оставшиеся 20% семей группы 11–15, т.е. семьи, в которых приходится на одного человека 15 м^2, и все семьи группы 16–19. Перегруппировка последней группы, как и первой, не нужна. Результаты перегруппировки представлены в табл. 3.18.

Таблица 3.18

Группировка семей по размеру жилой площади, приходящейся на одного человека по двум регионам в 2001 г.

№ п/п	Группы семей по размеру жилой площади, приходящейся на 1 чел., м^2	Доля семей, % к итогу	
		первый регион	второй регион
1	До 5	3,6	6,2
2	5–6	11,4	$1/3 \cdot 46{,}3 = 15{,}43$
3	7–8	19,4	$1/3 \cdot 46{,}3 = 15{,}43$
4	9–12	37,8	$(46{,}3 - 2 \cdot 15{,}43) + (0{,}4 \cdot 28{,}5) = 26{,}84$
5	13–14	11,1	$0{,}4 \cdot 28{,}5 = 11{,}4$
6	15–19	13,0	$0{,}2 \cdot 28{,}5 + 10{,}8 = 16{,}5$
7	20 и более	3,7	8,2
	Итого	100	100

Из приведенной группировки следует, что размер жилой площади на одного человека во втором регионе менее дифференцирован, чем в первом.

3.7 МЕТОД ГРУППИРОВОК И МНОГОМЕРНЫЕ КЛАССИФИКАЦИИ

Метод группировок позволяет получить общее представление о различных сторонах изучаемого объекта или процесса, выявить закономерности изменения основных показателей в совокупности, установить взаимосвязи и зависимости различных сторон изучаемых явлений, определить влияние факторов на изменение результативного признака.

Аналитические группировки, построенные по одному признаку, и сложные группировки позволяют установить связь и определить направление между результативными и 1–3 факторными признаками. Но часто этого бывает недостаточно, так как в действительности на изменение величины результативного признака оказывает влияние множество факторов, действующих в разных направлениях.

Для исследования таких многофакторных связей используются различные методы многомерной классификации, наибольшее распространение среди которых в практике экономико-статистического анализа социально-экономических явлений и процессов получили следующие, представленные на рис. 3.9.

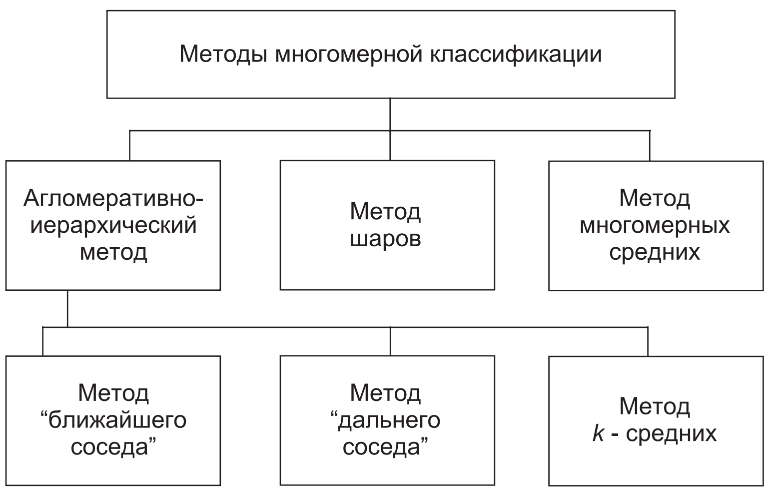

Рис. 3.9. Методы многомерной классификации

Агломеративно-иерархический метод рассмотрен в главе 13 данного учебника.

В данной главе остановимся подробнее на раскрытии сущности алгоритма реализации и возможностей экономической интерпретации выходных параметров метода многомерной группировки – метода многомерных средних.

Целью таких группировок является разбиение совокупности социально-экономических явлений на качественно однородные группы по большому числу признаков одновременно и определение на их основе связи и влияния факторных признаков на результативный.

В основу построения многомерной группировки положен принцип перехода от величин, имеющих определенную размерность, к безразмерным относительным величинам.

Сущность и этапы реализации метода

1. Все значения (абсолютные) результативного признака заменяются отношением вида:

$$Q_i = \frac{y_i}{\bar{y}}, \qquad (3.5)$$

где y_i – эмпирические значения результативного признака;

$\bar{y} = \frac{\Sigma y_i}{n}$ – средний уровень результативного признака.

2. Все значения (абсолютные) факторных признаков x_i заменяются отношениями:

$$P_{ij} = \frac{x_{ij}}{\bar{x}_j}, \qquad (3.6)$$

где $\bar{x}_j$ – средний уровень j-го факторного признака.

3. В результате такой замены (п. 1 и 2) матрица эмпирических абсолютных значений признаков y и x_j заменяется матрицей относительных величин Q_i и P_j следующего вида (табл. 3.19).

Таблица 3.19

Матрица относительных величин

№ п/п	Результативный признак $\frac{y_i}{\bar{y}}$	Факторные признаки				
		$\frac{x_{i1}}{\bar{x}_1}$	$\frac{x_{i2}}{\bar{x}_2}$	$\frac{x_{i3}}{\bar{x}_3}$	...	$\frac{x_{ik}}{\bar{x}_k}$
1	Q_1	P_{11}	P_{12}	P_{13}	…	P_{1k}
2	Q_2	P_{21}	P_{22}	P_{23}	…	P_{2k}
3	Q_3	P_{31}	P_{32}	P_{33}	…	…
.	.	.	.	.	.	.
.	.	.	.	.	.	.
.	.	.	.	.	.	.
n	Q_n	P_{n1}	P_{n2}	P_{n3}	…	P_{nk}

4. Если связь между результативными (y) и факторными (x_j) признаками обратная, то для каждой единицы объекта исследования определяется величина $1/P_{ij}$.

5. На основе относительных величин всех факторных признаков P_{ij} определяется многомерная средняя

$$\bar{P}_{ij} = \frac{\Sigma P_{ij}}{k},$$

где k – число факторных признаков.

6. Формируются общие факторы, близкие между собой по экономическому содержанию, из исходных факторных признаков: A, B, C, … При этом значимость влияния общих факторов на результативный предполагается одинаковой.

7. На основе относительных величин факторных признаков, составляющих общий фактор (A, B, C и т.д.), определяются факторные многомерные средние: $\bar{P}_A$, $\bar{P}_B$, $\bar{P}_C$ и т.д.

8. Определяется число групп исходя из объема изучаемой совокупности.

9. При анализе большого объема совокупности ширина равного интервала для построения групп на основе многомерной средней определяется по формуле:

$$i_{\overline{P}} = \frac{\overline{P}_{\max} - \overline{P}_{\min}}{\text{Число групп}}.$$

10. Определяется число единиц изучаемой совокупности по Q_i и $\overline{P}_A, \overline{P}_B, \overline{P}_C$ и т.д. в каждой выделенной группе.

11. Определяется суммарное значение результативного признака по тем единицам изучаемой совокупности, которые попали в выделенные группы в п.10.

Рассмотрим реализацию метода многомерных средних на примере предприятий пищевой промышленности в одном из регионов РФ в 2002 г. (табл. 3.20).

По каждому предприятию были рассмотрены следующие результативный (y) и факторные признаки (x_j):

y – выработка продукции на одного работающего, руб./чел.;

x_1 – электровооруженность труда, кВт·ч/чел.;

x_2 – энерговооруженность труда, кВт·ч/чел.;

x_3 – расход тепловой энергии;

x_4 – фондовооруженность труда, руб./чел.;

x_5 – рентабельность продукции, %;

x_6 – численность работников, чел.;

x_7 – балансовая прибыль предприятия, тыс. руб.;

x_8 – фонд времени без потерь, ч;

x_9 – коэффициент сменности рабочих.

Расчетные значения относительных величин по каждому из рассматриваемых признаков и многомерная средняя приведены в табл. 3.21.

Сформируем общие факторы по экономической сущности исходных переменных следующим образом:

A: x_1, x_2, x_3 – энерготопливные ресурсы предприятия;

B: x_4, x_5, x_7 – результативность (эффективность) работы предприятия;

C: x_6, x_8, x_9 – наличие и использование трудовых ресурсов предприятия.

Таким образом, матрица значений многомерных средних общих факторов будет иметь вид (табл. 3.22).

Таблица 3.20

Основные показатели работы предприятий пищевой промышленности в одном из регионов РФ в 2002 г.

№ п/п	y	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9
1	27,4	2,58	178	12	1,37	0,99	57	778	919	1,00
2	26,8	2,65	201	16	0,58	0,44	213	1995	3433	1,32
3	45,1	2,10	315	10	1,36	0,41	58	767	934	1,43
4	35,1	2,14	162	11	0,66	0,21	196	1203	3159	1,42
5	43,6	1,31	180	12	0,69	0,10	124	1017	1998	2,07
6	36,8	3,53	204	18	1,40	0,59	52	711	838	1,37
7	39,4	1,40	215	17	0,71	0,21	115	1000	1854	1,00
8	33,2	3,59	302	22	1,34	0,70	53	712	854	1,00
9	36,5	0,87	147	19	1,28	0,66	48	696	774	1,00
10	32,1	3,03	308	27	0,96	0,64	65	812	1048	1,44
Итого	356	23,2	2212	164	10,35	4,95	981	9691	15811	13,05
Средняя	35,6	2,32	221,2	16,4	1,035	0,495	98,1	969,1	1581,1	1,305

Таблица 3.21

Матрица отношений

№ п/п	Q_i	P_{i1}	P_{i2}	P_{i3}	P_{i4}	P_{i5}	P_{i6}	P_{i7}	P_{i8}	P_{i9}	$\overline{P_i}$
1	0,770	1,112	0,805	0,732	1,324	2,000	0,581	0,803	0,581	0,766	0,967
2	0,753	1,142	0,909	0,976	0,560	0,889	2,171	2,059	2,171	1,011	1,321
3	1,267	0,905	1,424	0,610	1,314	0,828	0,591	0,791	0,591	1,096	0,906
4	0,986	0,922	0,732	0,671	0,638	0,424	1,998	1,241	1,998	1,088	1,079
5	1,225	0,565	0,814	0,732	0,667	0,202	1,264	1,049	1,264	1,586	0,905
6	1,034	1,522	0,922	1,098	1,353	1,192	0,530	0,734	0,530	1,050	0,992
7	1,107	0,603	0,972	1,037	0,686	0,424	1,172	1,032	1,173	0,766	0,874
8	0,933	1,547	1,365	1,341	1,295	1,414	0,540	0,735	0,540	0,766	1,060
9	1,025	0,375	0,665	1,158	1,237	1,333	0,489	0,718	0,490	0,766	0,803
10	0,902	1,306	1,392	1,646	0,928	1,293	0,663	0,838	0,663	1,103	1,092

Таблица 3.22

Многомерные средние общих факторных признаков

№ п/п	$\overline{P_j}$	$\overline{P}_{A_j}$	$\overline{P}_{B_j}$	$\overline{P}_{C_j}$
1	0,967	0,883	1,376	0,643
2	1,321	1,009	1,169	1,784
3	0,906	0,980	0,978	0,759
4	1,079	0,775	0,768	1,695
5	0,905	0,704	0,639	1,371
6	0,992	1,181	1,093	0,703
7	0,874	0,871	0,714	1,037
8	1,060	1,418	1,148	1,846
9	0,803	0,733	1,096	1,745
10	1,092	1,448	1,020	0,810

Определим ширину интервала группировки:

$$\overline{P}\text{max} = 1{,}321;$$

$$\overline{P}\text{min} = 0{,}803;$$

$$h = \frac{R}{n} = \frac{\overline{P}\text{max} - \overline{P}\text{min}}{n} = \frac{1{,}321 - 0{,}803}{2} = 0{,}259.$$

Интервалы по многомерной средней имеют вид:

до 1,062

1,062 и более

Таким образом, многомерная группировка предприятий пищевой промышленности одного из регионов РФ имеет вид (табл. 3.23).

Таблица 3.23

Группировка предприятий пищевой промышленности одного из регионов РФ в 2002 г.

Группы предприятий по много-мерной средней	Число пред-приятий	Выработка продукции на одного работающего, руб./чел.	Обобщенные факторы					
			энерготопливные ресурсы, *A*		результативность работы, *B*		наличие и использование трудовых ресурсов, *C*	
			число пред-приятий	выработка продукции на одного работающего, руб./чел.	число пред-приятий	выработка продукции на одного работающего, руб./чел.	число пред-приятий	выработка продукции на одного работающего, руб./чел.
До 1,062	7	34,43	7	36,27	5	65,10	5	60,27
1,062 и более	3	31,33	3	34,03	5	32,14	5	35,04
Итого	10	35,60	10	35,60	10	35,60	10	35,60

ОСНОВНЫЕ ПОНЯТИЯ

Аналитическая группировка – группировка, выявляющая взаимосвязи между изучаемыми явлениями и их признаками.

Атрибутивный ряд распределения – ряд, построенный по качественному признаку.

Варианты – отдельные значения признака, которые он принимает в вариационном ряду.

Вариационный ряд распределения – ряд, построенный по количественному признаку.

Величина интервала – разность между верхней и нижней границами интервала.

Вторичная группировка – операция по образованию новых групп на основе ранее построенной группировки.

Группировка – расчленение множества единиц изучаемой совокупности на группы по определенным, существенным для них признакам.

Группировочный признак – признак, по которому производится разбиение единиц совокупности на отдельные группы.

Дискретный вариационный ряд – распределение единиц совокупности по дискретному признаку.

Закрытые интервалы – интервалы, у которых обозначены обе границы.

Интервал – значения варьирующего признака, лежащие в определенных границах.

Интервальный вариационный ряд – ряд, который отражает непрерывную вариацию признака.

Классификация – систематическое распределение явлений и объектов на определенные группы, классы, разряды на основе их сходства и различия.

Открытые интервалы — интервалы, у которых указана только одна граница.

Ряд распределения – упорядоченное распределение единиц совокупности на группы по определенному варьирующему признаку.

Сводка – комплекс последовательных операций по обобщению конкретных единичных факторов для выявления типичных черт и закономерностей, присущих изучаемому явлению в целом.

Структурная группировка – разделение исследуемой качественно однородной совокупности на группы, характеризующие ее структуру по какому-либо варьирующему признаку.

Типологическая группировка – разделение исследуемой качественно разнородной совокупности на классы, социально-экономические типы, однородные группы единиц в соответствии с правилами научной группировки.

Частоты – выраженные в долях единицы или в процентах к итогу значения изучаемого признака.

ТЕСТЫ

1. Группировка, в которой происходит разбиение однородной совокупности на группы, называется:

а) типологической группировкой;
б) структурной группировкой;
в) аналитической группировкой.

2. По технике выполнения статистическая сводка делится на:

а) простую и сложную;
б) централизованную и децентрализованную;
в) компьютерную и ручную.

3. Основанием группировки может быть:

а) качественный признак;
б) количественный признак;
в) качественный и количественный признаки.

4. Особое внимание нужно обратить на число единиц исследуемого объекта, если основанием группировки выбран:

а) качественный признак;
б) количественный признак;
в) как качественный, так и количественный признак.

5. Наибольшее значение признака в интервале называется:

а) нижней границей;
б) верхней границей интервала.

6. Величина равного интервала определяется по формуле:

а) $h_{i+1} = h_i + a$;
б) $h_{i+1} = h_i \cdot q$;
в) $h = R / n$.

7. Если величина интервала равна 0,5σ, то совокупность разбивается на:

а) 6 групп;
б) 9 групп;
в) 12 групп.

8. При непрерывной вариации признака целесообразно построить:

а) дискретный вариационный ряд;

б) интервальный вариационный ряд;

в) ряд распределения.

9. Накопленные частоты используются при построении:

а) огивы;

б) гистограммы;

в) полигона.

10. Если две группировки несопоставимы из-за различного числа выделенных групп, то они приводятся к сопоставимому виду:

а) с помощью метода вторичной группировки;

б) путем построения сложной группировки.

ЛИТЕРАТУРА

Боярский А.Я. Теоретические исследования по статистике: Сб. научных трудов. – М.: Статистика, 1974. – 303 с.

Венецкий И.Г. Вариационные ряды и их характеристики. – М.: Статистика, 1970. – 159 с.

Венецкий И.Г., Кильдишев Г.С. Теория вероятностей и математическая статистика. – М.: Статистика, 1975. – 264 с.

Грачев Н.Г. Статистические группировки. – М.: Госстатиздат, 1951. – 156 с.

Грачев Н.Г. Классификация и показатели структуры промышленности. – М.: Изд-во Академии наук СССР, 1963. – 123 с.

Елисеева И.И., Юзбашев М.М. Общая теория статистики. – М.: Финансы и статистика, 1998. – 479 с.

Иванов О.В. Теория статистической группировки. – М., 1992. – 91 с.

Плошко Б.Г. Группировка и система статистических показателей. – М., 1994. – 176 с.

Ряузов Н.Н. Сводка и группировка статистических материалов. – М., 1961. – 48 с.

Сиськов В.И. Корреляционный анализ в экономических исследованиях. – М.: Статистика, 1975. – 168 с.

ГЛАВА 4
СТАТИСТИЧЕСКИЕ ТАБЛИЦЫ

4.1
ПОНЯТИЕ О СТАТИСТИЧЕСКОЙ ТАБЛИЦЕ. ЭЛЕМЕНТЫ СТАТИСТИЧЕСКОЙ ТАБЛИЦЫ

Результаты сводки и группировки материалов статистического наблюдения, как правило, излагаются в виде таблиц. Таблица является наиболее рациональной, наглядной и компактной формой представления статистического материала. Однако не всякая таблица является статистической. Таблица умножения, опросный лист социологического обследования и т.д. могут носить табличную форму, но не являться статистическими таблицами.

Статистическая таблица от других табличных форм отличается следующим:

- содержит результаты подсчета эмпирических данных;
- является итогом сводки первоначальной информации.

Таким образом, *статистической* называется *таблица,* которая содержит сводную числовую характеристику исследуемой совокупности по одному или нескольким существенным признакам, взаимосвязанным логикой экономического анализа.

Основные элементы статистической таблицы, представленные на рис. 4.1, составляют как бы ее основу.

Табличная форма представления числовой информации – это такая, при которой число располагается на пересечении четко сформулированного заголовка по вертикальному столбцу, называемому *графой,* и сформулированного названия по соответствующей горизонтальной полосе – *строке.*

Таким образом, внешне таблица представляет собой пересечение граф и строк, которые формируют ее остов. Каждое пересечение образует клетку таблицы. Размер таблицы определяется произведением числа строк на число граф.

Статистическая таблица содержит три вида заголовков: общий, верхние и боковые. *Общий* заголовок отражает содержание всей таблицы (к какому месту и времени она относится), располагается над ее макетом

Название таблицы
(общий заголовок)

Содержание строк	Наименование граф (верхние заголовки)					
А	1	2	3	4	5	…
Наименования строк (боковые заголовки)						
Итоговая строка						Итоговая графа

Рис. 4.1. Остов (основа) статистической таблицы

по центру и является внешним заголовком. *Верхние* заголовки характеризуют содержание граф, а *боковые* – строк. Они являются внутренними заготовками.

Остов таблицы, заполненный заголовками, образует ее макет. Если на пересечении граф и строк записать цифры, то получается полная статистическая таблица. Схематично взаимодействие основных элементов статистической таблицы представлено на рис. 4.2.

В случае необходимости таблицы могут сопровождаться примечанием, используемым с целью пояснения заголовков, методики расчета некоторых показателей, источников информации и т.д.

По логическому содержанию таблица представляет собой статистическое предложение, основными элементами которого являются подлежащее и сказуемое.

Подлежащим статистической таблицы называется объект, характеризующийся цифрами. Это могут быть отдельные единицы совокупности (фирмы, объединения) в порядке их перечня или сгруппированные по каким-либо признакам. Обычно подлежащее таблицы дается в левой части, в наименовании строк.

Сказуемое статистической таблицы образует система показателей, которые характеризуют объект изучения, т.е. подлежащее таблицы. Сказуемое формирует верхние заголовки и составляет содержание граф с логически последовательным расположением показателей слева направо.

Расположение подлежащего и сказуемого может меняться местами, что зависит от достижения каждым исследователем в отдельности наиболее полного и лучшего способа прочтения и анализа исходной информации об исследуемой совокупности.

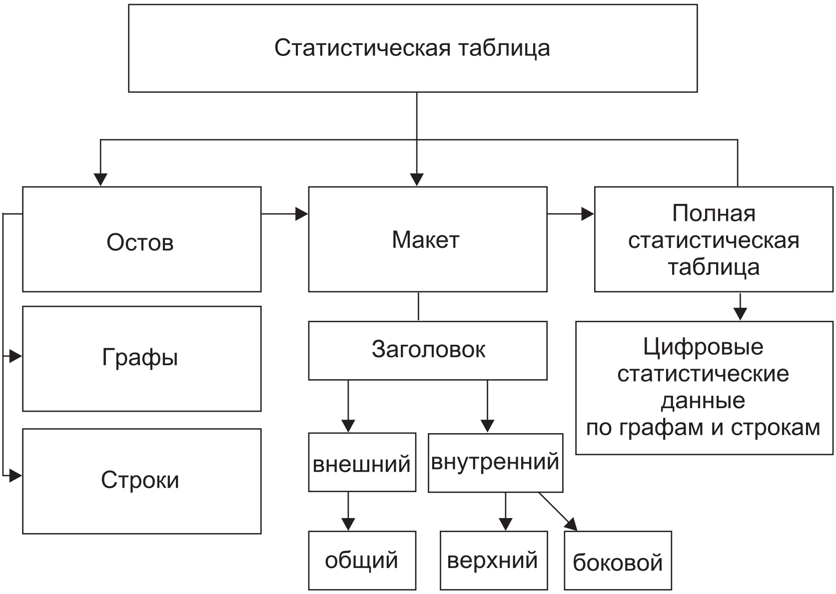

Рис. 4.2. Основные составляющие элементы статистической таблицы

4.2 ВИДЫ ТАБЛИЦ ПО ХАРАКТЕРУ ПОДЛЕЖАЩЕГО

В практике экономико-статистического анализа используются различные виды статистических таблиц, отличающихся различным строением подлежащего и сказуемого, структурой и соотношением признаков, формирующих их.

В зависимости от структуры подлежащего и группировки в нем единиц объекта различают статистические таблицы *простые* и *сложные,* а последние, в свою очередь, подразделяются на *групповые* и *комбинационные.* Общая классификация таблиц по подлежащему представлена на рис. 4.3.

В *простой* таблице в подлежащем дается простой перечень каких-либо объектов или территориальных единиц, т.е. в подлежащем нет группировки единиц совокупности. Простые таблицы бывают монографическими и перечневыми. *Монографические таблицы* характеризуют не всю совокупность единиц изучаемого объекта, а только

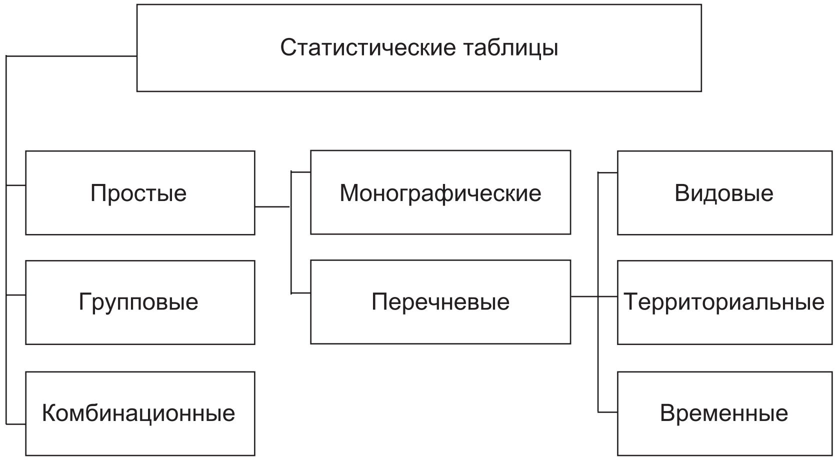

Рис. 4.3. Классификация статистических таблиц по характеру подлежащего

одну какую-либо группу из него, выделенную по определенному, заранее сформулированному признаку (табл. 4.1).

Простыми перечневыми называются таблицы, подлежащее которых содержит перечень единиц изучаемого объекта по различному признаку – видовому, территориальному, временному и т.д.

Примером видовой таблицы может служить табл. 4.2, в подлежащем которой содержится перечень основных демографических показателей.

Примером перечневой таблицы по территориальному признаку является табл. 4.3, которая дает характеристику распределения убыточных предприятий и организаций административных округов г. Москвы по различной отраслевой принадлежности. По каждому административному округу можно получить информацию о количестве убыточных предприятий и организаций и их отраслевой принадлежности.

Перечневая таблица по временному признаку показана на примере динамики среднемесячной начисленной заработной платы работников предприятий и организаций в долл. США исходя из среднегодового официального курса доллара США (табл. 4.4).

Простые таблицы содержат лишь *описательные сведения*. Они не дают возможность выявить социально-экономические типы изучаемых явлений, их структуру, а также взаимосвязи и взаимозависимости между характеризующими их признаками.

Таблица 4.1

Ввод в действие зданий нежилого назначения в РФ в 2001 г.

Показатель	Количество зданий, тыс. ед.	Общий строительный объем зданий, млн куб. м	Общая площадь зданий, млн кв. м
А	1	2	3
Ввод в действие зданий нежилого назначения	5,4	27,7	5,8

Таблица 4.2

Динамика основных демографических показателей в Российской Федерации за 1992–2000 гг.

Показатели	1992	2000
А	1	2
Среднегодовая численность населения – всего, млн чел.	148,3	145,2
Ожидаемая продолжительность жизни при рождении, число лет:		
все население	67,9	65,3
мужчин	62,0	59,0
женщин	73,8	72,2
На 1000 чел. населения:		
родившихся	10,7	8,7
умерших	12,2	15,4
Естественный прирост (+), убыль (–) населения	–1,5	–6,7
Число браков	7,1	6,2
Число разводов	4,3	4,3

Эти задачи более полно решаются с помощью групповых, и особенно комбинационных таблиц.

Групповыми называются статистические таблицы, подлежащее которых содержит группировку единиц совокупности по одному количественному или атрибутивному признаку.

Таблица 4.3

Распределение убыточных предприятий и организаций по административным округам Москвы в июне 2000 г.

Административный округ	Количество убыточных предприятий, ед.	В том числе распределение по отраслевой принадлежности		
		промышленных предприятий	строительных организаций	транспортных организаций
А	1	2	3	4
Центральный	615	63	97	28
Северный	149	26	14	18
Северо-Восточный	157	27	13	1
Восточный	132	40	10	13
Юго-Восточный	96	36	5	11
Южный	120	29	3	16
Юго-Западный	104	7	8	8
Западный	94	15	5	7
Северо-Западный	61	6	6	6
Зеленоград	24	5	-	1
Всего	1552	254	161	109

Таблица 4.4

Динамика среднемесячной начисленной заработной платы работников предприятий и организаций в долларах США за период 1992–2000 гг.

Год	1992	1993	1994	1995	1996	1997	1998	1999	2000
Долл. США	22	57	98	103	154	164	108	62	80

Простейшим видом групповых таблиц являются атрибутивные и вариационные ряды распределения. Групповая таблица может быть более сложной, если в сказуемом приводится не только число единиц в каждой группе, но и ряд других важных показателей, количественно и качественно характеризующих группы подлежащего. Такие таблицы часто используются в целях сопоставления обобщающих показателей по группам, что позволяет делать определенные практические выводы.

Примером групповой таблицы является табл. 4.5, которая отражает количественное распределение регионов РФ по уровню безработицы и показывает численность экономически активного населения, занятого в экономике, и безработных по выделенным группам.

Таблица 4.5

Группировка регионов Российской Федерации по уровню безработных за 2000 г.

№ п/п	Группы регионов по уровню безработицы, % от экономически активного населения	Количество регионов		Экономически активное население, тыс. чел.	Занятые в экономике, тыс.чел.	Безработные, тыс.чел.
		ед.	% к итогу			
	А	1	2	3	4	5
1	5,6–17,2	61	77,2	61993	54366	7627
2	17,2–28,8	15	19,0	6564	5261	1303
3	28,8–40,4	2	2,5	1068	728	340
4	40,4–52,0	1	1,3	105	51	54
	Всего	79	100,0	69730	60406	9324

Для более полной характеристики сложных социально-экономических явлений бывает недостаточно провести группировку по одному признаку. Изучаемые объекты обычно характеризуются многими свойствами, многими признаками, часто взаимосвязанными. Для того чтобы раскрыть эти связи и полнее охарактеризовать типы явлений, прибегают к сложной группировке по двум и более признакам. При статистической сводке материалов сложной группировки применяется комбинационная таблица.

Комбинационными называются статистические таблицы, подлежащее которых содержит группировку единиц совокупности одновременно по двум и более признакам: каждая из групп, построенная по одному признаку, разбита в свою очередь на подгруппы по какому-либо другому признаку и т.д.

В табл. 4.6 подлежащим являются группы предприятий одной из отраслей промышленности РФ по стоимости основных фондов и объему промышленной продукции. Из данных табл. 4.6 видно, что объем промышленной продукции зависит от стоимости основных производственных фондов предприятий данной отрасли промышленности в 2001 г.

Таблица 4.6

Группировка предприятий одной из отраслей промышленности РФ по стоимости основных фондов и объему производства промышленной продукции в 2001 г.*

№ п/п	Группы предприятий по стоимости основных производственных фондов, млн руб.	Подгруппы предприятий по объему промышленной продукции, млрд шт.	Число предприятий	Стоимость основных производственных фондов, млн руб.	Объем промышленной продукции, млрд шт.	Среднесписочная численность промышленно-производственного персонала, тыс. чел.
	А	Б	1	2	3	4
1	1,1–6,0	3,0–11,6 11,6–20,2	12 2	45,2 6,3	61,8 27,9	7,9 0,9
Итого по группе			14	51,5	89,7	8,8
2	6,0–10,9	3,0–11,6 11,6–20,2	7 2	55,7 17,4	49,9 25,0	3,7 1,1
Итого по группе			9	73,1	74,9	4,8
3	10,9–15,8	3,0–11,6 11,6–20,2	3 1	38,7 11,0	18,9 15,9	2,5 1,1
Итого по группе			4	49,7	34,8	3,6
4	15,8–20,7	3,0–11,6 11,6–20,2	1 2	30,2 35,4	10,5 27,3	1,3 1,2
Итого по группе			3	65,6	37,8	2,5
	Итого по подгруппам	3,0–11,6 11,6–20,2	23 7	169,8 70,1	141,1 96,1	15,4 4,3
Всего			30	239,9	237,2	19,7

* Цифры условные.

4.3
ВИДЫ ТАБЛИЦ ПО РАЗРАБОТКЕ СКАЗУЕМОГО

В сказуемом статистической таблицы, как уже говорилось, приводятся показатели, которые являются характеристикой изучаемого объекта. Эту характеристику можно давать небольшим числом показателей или целой системой показателей.

Статистические таблицы различаются по простой или сложной разработке (рис. 4.4).

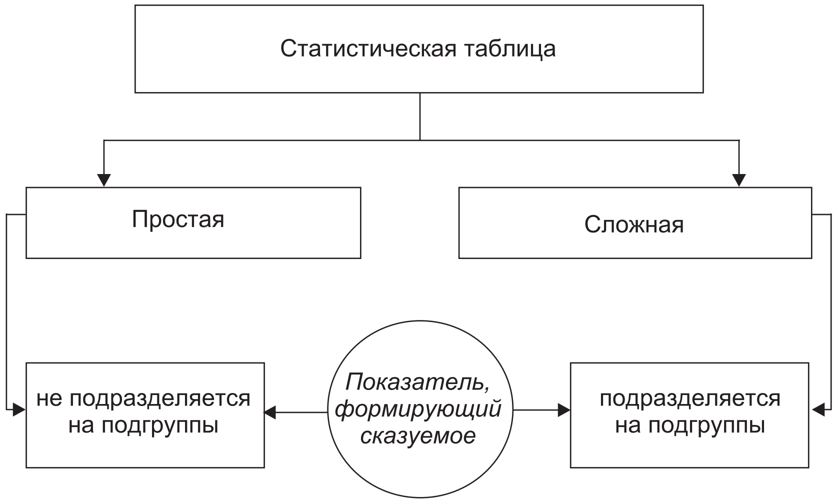

Рис. 4.4. Классификация статистических таблиц по разработке сказуемого

При *простой разработке* сказуемого показатель, определяющий его, не подразделяется на подгруппы, и итоговые значения получаются путем простого суммирования значений по каждому признаку отдельно независимо друг от друга. Примером простой разработки сказуемого могут служить табл.4.3, 4.4, 4.5, 4.6.

Сложная разработка сказуемого предполагает в сказуемом комбинацию одного признака с другим (табл. 4.7).

В табл. 4.7 сложная разработка сказуемого; она дает больше информации, чем простая разработка, так как позволяет видеть состав экономически активного населения по категориям, а внутри групп – по полу.

Таблица 4.7

Распределение численности экономически активного населения одного из регионов РФ в 2000–2001 гг.
(на конец года), млн руб.*

Год	Экономически активное население, всего	В том числе по категориям					
		занятые в экономике			безработные		
		мужчины	женщины	всего	мужчины	женщины	всего
А	1	2	3	4	5	6	7
2000	73,0	34,6	31,3	65,9	3,9	3,2	7,1
2001	72,9	33,9	30,7	64,6	4,5	3,7	8,2

* Цифры условные.

Однако сложная разработка сказуемого может привести к увеличению размерности статистической таблицы, что, в свою очередь, снижает ее наглядность, ухудшает чтение и анализ.

Поэтому исследователь при построении статистических таблиц должен руководствоваться оптимальным соотношением показателей сказуемого и учитывать как положительные, так и отрицательные моменты сложной разработки показателей сказуемого.

4.4 ОСНОВНЫЕ ПРАВИЛА ПОСТРОЕНИЯ ТАБЛИЦ

Статистические таблицы как средство наглядного и компактного представления цифровой информации должны быть статистически правильно оформлены.

Приемы, определяющие технику формирования статистических таблиц.

1. Таблица должна быть компактной и содержать только те исходные данные, которые непосредственно отражают исследуемое социально-экономическое явление в статике и динамике и необходимы для познания его сущности.

Следует избегать ненужной, второстепенной, бессодержательной к данному объекту исследования информации. Цифровой материал необходимо представлять таким образом, чтобы при анализе таблицы сущность явления раскрывалась чтением строк слева направо и сверху вниз.

2. Заголовок таблицы и названия граф и строк должны быть четкими, краткими, лаконичными, представлять собой законченное целое, органично вписываться в содержание текста.

Необходимо избегать большого количества точек и запятых в названиях таблицы и граф, затрудняющих чтение таблицы.

Если название таблицы состоит из двух и более предложений, ставится точка, чтобы отделить предложения друг от друга; после последнего предложения точка не ставится. В *заголовках граф* допускаются точки только при необходимых сокращениях. В *заголовке таблицы* должны найти отражение объект, признак, время и место совершения события. Но при этом следует помнить: чем короче и лаконичнее заголовок таблицы, тем она яснее для чтения и анализа, естественно, если краткость достигается не в ущерб точности и познавательности. Заголовки таблицы, граф и строк пишутся полностью, без сокращений.

3. Информация, располагаемая в столбцах (графах) таблицы, завершается итоговой строкой. Существуют различные способы соединения слагаемых граф с их итогом:

- строка «Итого» или «Всего» завершает статистическую таблицу;
- итоговая строка располагается первой строкой таблицы и соединяется с совокупностью ее слагаемых словами «в том числе».

В групповых и комбинационных таблицах всегда необходимо давать итоговые графы и строки.

4. Для того чтобы легче читать и анализировать достаточно большие таблицы (по количеству приведенных строк), целесообразно оставлять двойной промежуток после каждых пяти (и далее кратных пяти) строк.

5. Если названия отдельных граф повторяются между собой, содержат повторяющиеся термины или несут единую смысловую нагрузку, то им необходимо присвоить общий объединяющий заголовок.

Данный прием используется и для подлежащего, и для сказуемого таблиц.

6. Графы и строки полезно нумеровать. Графы, слева заполненные названием строк, принято обозначать заглавными буквами алфавита (А), (Б) и т.д., а все последующие графы – номерами в порядке возрастания.

7. Взаимосвязанные и взаимозависимые данные, характеризующие одну из сторон анализируемого явления (например, число предприятий и удельный вес заводов (% к итогу), абсолютный прирост и темп роста и т.д.), целесообразно располагать в соседних друг с другом графах.

8. Графы и строки должны содержать единицы измерения, соответствующие поставленным в подлежащем и сказуемом показателям. При этом используются общепринятые сокращения единиц измерения (чел., руб., кВт·ч и т.д.).

9. Лучше всего располагать в таблицах сопоставляемую в ходе анализа цифровую информацию в одной и той же графе, одну под другой, что значительно облегчает процесс их сравнения.

Поэтому в групповых таблицах, например, группы по изучаемому признаку более грамотно располагать в порядке убывания или возрастания его значений при сохранении логической связи между подлежащим и сказуемым таблицы.

10. Для удобства работы числа в таблицах следует представлять в середине граф, одно под другим: единицы под единицами, запятая под запятой, четко соблюдая при этом их разрядность.

11. По возможности числа целесообразно округлять. Округление чисел в пределах одной и той же графы или строки следует проводить с одинаковой степенью точности (до целого знака или до десятой и т.д.).

Если все числа одной и той же графы или строки даны с одним десятичным знаком, а одно из чисел имеет два и более знака после запятой, то числа с одним знаком после запятой следует дополнять нулем, тем самым подчеркивая их одинаковую точность.

12. Отсутствие данных об анализируемом социально-экономическом явлении может быть обусловлено различными причинами, что по-разному отмечается в таблице:

- если данная позиция (на пересечении соответствующих графы и строки) вообще не подлежит заполнению, то ставится знак «Х»;
- когда по какой-либо причине отсутствуют сведения, то ставится многоточие «...» или «Нет свед.», или «Н. св.»;
- при отсутствии явления клетка заполняется тире «–» и остается пустой.

Для отображения очень малых чисел используют обозначения (0,0) или (0,00), предполагающие возможность наличия числа.

13. В случае необходимости дополнительной информации – разъяснений к таблице могут даваться примечания.

Соблюдение приведенных правил построения и оформления статистических таблиц делает их основным средством представления, обработки и обобщения статистической информации о состоянии и развитии анализируемых социально-экономических явлений.

4.5
ЧТЕНИЕ И АНАЛИЗ ТАБЛИЦЫ

Анализу статистических таблиц предшествует этап ознакомления – их чтение.

Чтение и анализ таблиц должны осуществляться не хаотично, а в определенной последовательности.

Чтение предполагает, что исследователь, прочитав слова и числа таблицы, усвоил ее содержание, сформулировал первые суждения об объекте, уяснил назначение таблицы, понял ее содержание в целом, дал оценку явлению или процессу, описанному в таблице.

Анализ таблицы как метод научного исследования путем разбиения предмета изучения на части делится на структурный и содержательный. На рис. 4.5 приведено содержание основных этапов анализа статистических таблиц.

Структурный анализ предполагает анализ строения таблицы, характеристику представленных в таблице:

- совокупности и единиц наблюдения, формирующих ее;
- признаков и их комбинаций, формирующих подлежащее и сказуемое таблицы;
- признаков количественных или атрибутивных;
- соотношения признаков подлежащего с показателями сказуемого;
- вида таблицы: простая или сложная, а последняя – групповая или комбинационная;
- решаемых задач – анализ структуры, типов явлений или их взаимосвязей.

Содержательный анализ предполагает изучение внутреннего содержания таблицы: анализ отдельных групп подлежащего по соответствующим признакам сказуемого; выявление соотношения и пропорций между группами явлений по одному и разным признакам; сравнительный анализ и формулировку выводов по отдельным группам и по всей совокупности в целом; установление закономерностей и определение резервов развития изучаемого объекта.

Рис. 4.5. Этапы анализа статистических таблиц

Прежде чем приступать к анализу числовой информации, необходимо проверить ее достоверность и научную обоснованность. Исследователь должен убедиться в достоверности и надежности источника информации данных и критически оценить их цифровые значения. Следует провести логическую и счетную проверки данных (см. рис. 4.5). *Логическая проверка* состоит в возможности определения конкретных признаков теми или иными числовыми значениями (например, абсурдно, если численность работающих на фирме составила 106,7 человека). *Счетная проверка* предполагает выборочный расчет отдельных значений признаков по группе либо итоговых значений строк или граф и т.д.

Анализ данных таблиц производится по каждому признаку в отдельности, затем в логико-экономическом сочетании всей совокупности признаков в целом.

Анализ отдельных признаков и групп необходимо начинать с изучения абсолютных, затем связанных с ними относительных величин. При анализе данных следует рассматривать динамику каждого признака за весь период, переходя при этом от одного к другому.

Анализ таблиц может быть дополнен расчетными относительными и средними величинами, если этого требуют задачи исследования.

Для получения более полного и наглядного представления об изучаемых явлениях и процессах по данным статистических таблиц строятся графики, диаграммы и т.д.

Комплексный анализ таблиц, содержащих ряды динамики, дает возможность исследователю выявлять и обобщать тенденции и закономерности в развитии совокупности единиц наблюдения.

Анализ групповых и комбинационных таблиц позволяет охарактеризовать типы социально-экономических явлений, структуру совокупности, соотношения и пропорции между отдельными группами и единицами наблюдения; выявить характер и направление взаимосвязей и взаимозависимостей между различными, определенными логикой экономического анализа сочетаниями признаков и найти зависимости признаков-следствия от признаков-причин.

Соблюдение правил и последовательности работы со статистическими таблицами помогает исследователю осуществить научно обоснованный экономико-статистический анализ объектов и процессов.

4.6 ТАБЛИЦЫ И МАТРИЦЫ

В анализе данных наряду со статистическими таблицами применяются и другие виды таблиц, одним из которых является матрица.

Матрицей называется прямоугольная таблица числовой информации, состоящая из m строк и n столбцов. Таким образом матрица имеет размерность m х n:

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix},$$

где a_{ij} – элемент матрицы, стоящий на пересечении i-й строки и j-го столбца.

Различают два вида матриц:

- прямоугольную (размерность m х n);
- квадратную.

Если число строк строго равно числу столбцов (m = n), то матрица называется *квадратной порядка n*.

Квадратная матрица порядка n называется *диагональной* (D), если все элементы, стоящие вне главной диагонали (d_1, d_2,..., d_n), равны нулю.

$$D = \begin{pmatrix} d_1 & 0 & \dots & 0 \\ 0 & d_2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & d_n \end{pmatrix}.$$

Если в диагональной матрице D все $d_i = 1$, то матрица называется *единичной,* при $d_i = 0$ – *нулевой.*

Схематично названные виды матриц могут быть представлены следующим образом (рис. 4.6).

Рис. 4.6. Виды матриц

Матрицы и анализ явлений и процессов на их основе составляют базу матричного моделирования и позволяют исследовать взаимосвязи между экономическими объектами.

Таблицы-матрицы широко применяются на практике, например в экономике, в виде балансово-нормативных моделей, отражающих соотношение результатов производства, нормативов производственных затрат и т.д. Успешно используют матрицы и в межотраслевом балансе, системе национального счетоводства и т.д. (подробнее пример расчета и анализ на основе матриц см. в главе 13).

4.7 ТАБЛИЦЫ СОПРЯЖЕННОСТИ

Таблицей сопряженности называется таблица, которая содержит сводную числовую характеристику изучаемой совокупности по двум и более атрибутивным (качественным) признакам или комбинации количественных и атрибутивных признаков.

Таблицы сопряженности получили наибольшее распространение при изучении социальных явлений и процессов: общественного мнения, уровня и образа жизни, общественно-политического строя и т.д.

Наиболее простым видом таблиц сопряженности является *таблица частот 2* x *2* (табл. 4.8).

Таблица 4.8

Общая схема таблицы частот 2 x 2

Признак	B_1	B_2	Всего
A_1	f_{11}	f_{12}	f_{10}
A_2	f_{21}	f_{22}	f_{20}
Всего	f_{01}	f_{02}	f_{00}

Построение данной таблицы исходит из предположения, что ответы респондентов или анализируемые атрибутивные признаки будут принимать только два значения: A_1 и A_2, B_1 и B_2. Внутреннее цифровое наполнение таблицы представляют частоты (f_{ij}), обладающие одновременно i-м (i = 1, 2) значением одного (A_i) и j-м (j = 1, 2) значением (B_j) другого качественного признака.

Итоговая графа и строка содержат информацию о количественном распределении совокупности соответственно по A и B атрибутивным признакам.

Для более полного описания и анализа явлений и процессов, характеризуемых атрибутивными признаками, используются *таблицы сопряженности большей размерности:*

$$i \cdot j,$$

где $i = 1, 2, ..., k$ – число вариантов значений (например, ответов респондентов и т.д.) одного признака (например, признака A);

$j = 1, 2, ..., n$ – число вариантов значений другого признака (B) (табл. 4.9).

Таблица 4.9

Общая схема таблицы сопряженности большей размерности

Признак	B_1	B_2	…	B_j	Всего
A_1	f_{11}	f_{12}		f_{1j}	f_{10}
A_2	f_{21}	f_{22}		f_{2j}	f_{20}
…	…	…	…	…	…
A_i	f_{i1}	f_{i2}		f_{ij}	f_{i0}
Всего	f_{01}	f_{02}		f_{0j}	f_{00}

Принцип взаимной сопряженности наиболее эффективен при выявлении и оценке взаимосвязей и взаимозависимостей между социальными явлениями и процессами.

(Подробнее пример расчета и анализа на основе таблиц сопряженности см. в разделе 9.7).

ОСНОВНЫЕ ПОНЯТИЯ

Групповая статистическая таблица содержит группировку единиц совокупности по одному – количественному или атрибутивному – признаку.

Комбинационная статистическая таблица содержит группировку единиц совокупности одновременно по двум и более признакам.

Матрица – прямоугольная таблица числовой информации, состоящая из m строк и n столбцов.

Подлежащее статистической таблицы характеризует объект исследования. В нем дается перечень единиц совокупности либо групп исследуемого объекта по существенным признакам.

Простая разработка сказуемого – показатели в сказуемом даны параллельно один другому, без разделения на подгруппы.

Простая таблица – таблица, в подлежащем которой дается простой перечень объектов или территориальных единиц.

Сказуемое статистической таблицы – система показателей, которыми характеризуется объект изучения.

Сложная разработка сказуемого – показатели в сказуемом даны в комбинации друг с другом.

Статистическая таблица – способ рационального изложения и обобщения данных о социально-экономических явлениях при помощи цифр, расположенных в определенном порядке.

Таблица сопряженности – таблица, которая содержит сводную числовую характеристику изучаемой совокупности по двум и более атрибутивным признакам или комбинации количественных и атрибутивных признаков.

Знание вышеперечисленных основных понятий позволит в достаточной мере разобраться в теме курса.

ТЕСТЫ

1. Статистическая таблица представляет собой:

а) форму наиболее рационального изложения результатов статистического наблюдения;

б) сведения, расположенные по строкам и графам;

в) числовые характеристики, размещенные в колонках таблицы.

2. К статистической таблице можно отнести:

а) таблицу умножения;

б) опросный лист социологического обследования;

в) таблицу, характеризующую численность населения по полу и возрасту.

3. По характеру разработки подлежащего различают статистические таблицы:

а) простые;

б) перечневые;

в) комбинационные.

4. Монографические таблицы характеризуют:

а) совокупность единиц изучаемого объекта;

б) группу единиц совокупности по признаку;

в) каждую единицу совокупности.

5. Подлежащее групповых статистических таблиц содержит:

а) перечень единиц совокупности по признаку;

б) группировку единиц совокупности по одному признаку;

в) группировку единиц совокупности по нескольким признакам.

6. По характеру разработки сказуемого различают статистические таблицы:

а) монографические;

б) перечневые;

в) сложные.

7. Сказуемым статистической таблицы является:

а) исследуемый объект;

б) показатели, характеризующие исследуемый объект;

в) сведения, расположенные в верхних заголовках таблицы.

8. Имеются следующие данные о вводе в действие основных фондов, инвестиций в основной капитал и незавершенное строительство предприятий одного из регионов РФ в 2001 г.*:

№ п/п	Введено в действие основных фондов, млрд руб.	Инвестиции в основной капитал, млрд руб.	Незавершенное строительство на 01.01.2002 г., млрд руб.
	А	1	2
1	40,6	44,8	86,4
2	750,1	1704,7	4324,9
3	15,9	13,1	10,6
4	70,3	74,8	2,1
5	5,6	6,2	1,1
Итого	882,5	1843,6	4425,1

* Цифры условные.

Подлежащим таблицы является:
а) номер предприятия;
б) введено в действие основных фондов;
в) инвестиции в основной капитал;
г) незавершенное строительство на 01.01.2002 г.;
д) все показатели.

9. Имеются следующие данные о величине просроченной кредиторской задолженности предприятий некоторых отраслей экономики в декабре 2000 г. (на конец месяца), млрд руб.:

Отрасль	Количество организаций, имеющих просроченную кредиторскую задолженность, ед.	Просроченная кредиторская задолженность	Из нее		
			поставщикам	в бюджеты всех уровней	по платежам в государственные внебюджетные фонды
А	1	2	3	4	5
Строительство	7471	117,9	42,6	32,2	28,7
Промышленность	16 313	833,3	371,6	197,7	165,9
Транспорт	4201	146,5	41,4	48,1	25,6
Связь	493	3,2	1,9	0.4	0,2
Торговля и общественное питание	4721	103,3	69,0	21,1	2,7
Итого	33199	1204,2	526,5	299,5	223,1

По характеру разработки сказуемого таблица является:
а) групповой;
б) комбинационной;
в) простой;
г) сложной.

10. Таблица сопряженности строится по:
а) количественным признакам;
б) атрибутивным признакам;
в) комбинации количественных и атрибутивных признаков.

ЛИТЕРАТУРА

Долгушевский Ф.Г., Козлов В.С., Полушин М.И., Эрлих Я.М. Общая теория статистики. – М.: Статистика, 1967. – 384 с.

Колмогоров А. Предисловие к книге Г. Лебега «Общие величины». – М.: Госстатиздат, 1938. – 4 с.

Лившиц Ф.Д. Статистические таблицы. – М.: Госстатиздат, 1958. – 139 с.

Маслов П.П. Техника работы с цифрами. – М.: Статистика, 1969. – 120 с.

ГЛАВА 5

ГРАФИЧЕСКОЕ ИЗОБРАЖЕНИЕ СТАТИСТИЧЕСКИХ ДАННЫХ

5.1 ПОНЯТИЕ О СТАТИСТИЧЕСКОМ ГРАФИКЕ. ЭЛЕМЕНТЫ СТАТИСТИЧЕСКОГО ГРАФИКА

Современный анализ социально-экономических явлений немыслим без применения графического метода представления данных.

Графический метод есть метод условных изображений статистических данных при помощи геометрических фигур, линий, точек и разнообразных символических образов.

Впервые о технике составления статистических графиков упоминается в работе английского экономиста У. Плейфейра «Коммерческий и политический атлас», опубликованный в 1786 г. и положивший начало развитию приемов графического изображения статистических данных.

Главное достоинство статистических графиков – наглядность. При правильном их построении статистические показатели привлекают к себе внимание, становятся более понятными, выразительными, лаконичными, запоминающимися. Графики прочно вошли в практическую работу экономистов, статистиков и работников учета. В ряде случаев графики стали незаменимым средством обобщения статистических данных, подведения итогов сложных исследований и выявления связи между явлениями. Поэтому необходимо уметь строить и читать статистические графики.

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

Рассмотрим подробнее каждый из указанных элементов.

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

В качестве графического образа выступают и объемные фигуры. Иногда в графиках используются и негеометрические фигуры в виде силуэтов или рисунков предметов.

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

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

Построение графика – это всегда творческий процесс. Здесь необходим некоторый поиск. Лишь после составления и сравнения нескольких черновых вариантов можно определить правильную композицию графика, установить масштабы и расположение знаков на поле графика.

Пространственные ориентиры графика задаются в виде системы координатных сеток. Система координат необходима для размещения геометрических знаков в поле графика. Наиболее распространенной является система прямоугольных координат. Для построения статистических графиков используется обычно только 1-й и изредка 1-й и 4-й квадраты. В практике графического изображения применяются также полярные координаты. Они необходимы для наглядного изображения циклического движения явления во времени. В полярной системе координат (рис. 5.1) один из лучей, обычно правый горизонтальный, принимается за *ось ординат*, относительно которой определяется угол луча (первая координата). Второй координатой считается ее расстояние от центра сетки, называемое *радиусом*. В радиальных графиках лучи обозначают моменты времени, а окружность – величину изучаемого явления. На статистических картах пространственные ориентиры задаются контурной сеткой (контуры рек, береговая линия морей и океанов, границы государств) и определяют те территории, к которым относятся статистические величины.

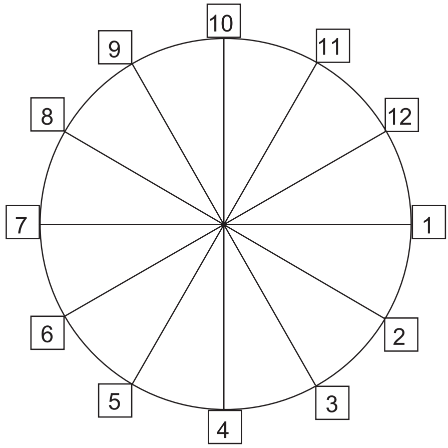

Рис. 5.1. Полярная система координат

Масштабные ориентиры статистического графика определяются масштабом и системой масштабных шкал. Масштаб статистического графика – это мера перевода числовой величины в графическую.

Масштабной шкалой называется линия, отдельные точки которой могут быть прочитаны как определенные числа. Шкала имеет большое значение в графике и включает три элемента: *линию* (или носитель шкалы); определенное *число* помеченных черточками *точек*, которые расположены на носителе шкалы в определенном порядке, *цифровое обозначение чисел*, соответствующих отдельным помеченным точкам. Как правило, цифровым обозначением снабжаются не все помеченные точки, а лишь некоторые из них, расположенные в определенном порядке. По правилам числовое значение необходимо помещать строго против соответствующих точек, а не между ними (рис.5.2.).

Носитель шкалы может представлять собой как прямую, так и кривую линии. Поэтому различают *шкалы прямолинейные* (например, миллиметровая линейка) и *криволинейные* – дуговые и круговые (циферблат часов).

Графические и числовые интервалы бывают равными и неравными. Если на всем протяжении шкалы равным графическим интервалам соответствуют равные числовые, такая шкала называется *равномерной*. Когда же равным числовым интервалам соответствуют неравные графические интервалы, и наоборот, шкала называется *неравномерной*.

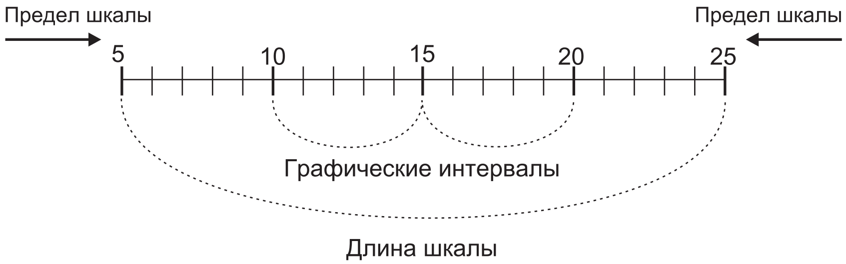

Рис. 5.2. Числовые интервалы

Масштабом равномерной шкалы называется ***длина отрезка*** (графический интервал), принятого за единицу и измеренного в каких-либо мерах. Чем меньше масштаб (рис.5.3), тем гуще располагаются на шкале точки, имеющие одно и то же значение. Построить шкалу – это значит на заданном носителе шкалы разместить точки и обозначить их соответствующими числами согласно условиям задачи.

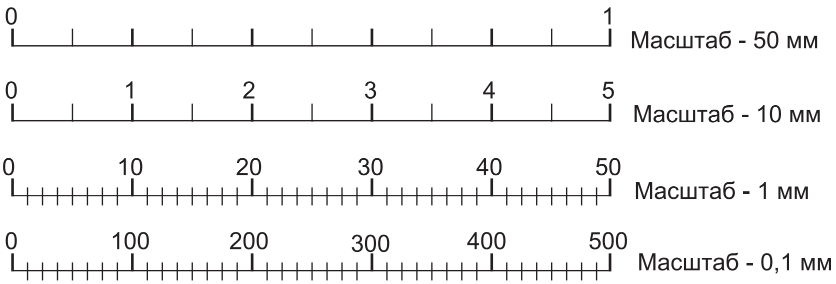

Рис. 5.3. Масштабы

Как правило, масштаб определяется примерной прикидкой возможной длины шкалы и ее пределов. Например, на поле в 20 клеток надо построить шкалу от 0 до 850. Так как 850 не делится удобно на 20, то округляем число 850 до ближайшего удобного числа, в данном случае 1000 (1000 : 20 = 50), т.е. в одной клетке 50, а в двух клетках 100; следовательно, масштаб – 100 в двух клетках.

Из неравномерных шкал наибольшее распространение имеет *логарифмическая*. Методика ее построения несколько иная, так как на этой шкале отрезки пропорциональны не изображаемым величинам, а их логарифмам. Так, при основании 101g1 = 0; 1g10 = 1; 1g100 = 2 и т.д. (рис. 5.4).

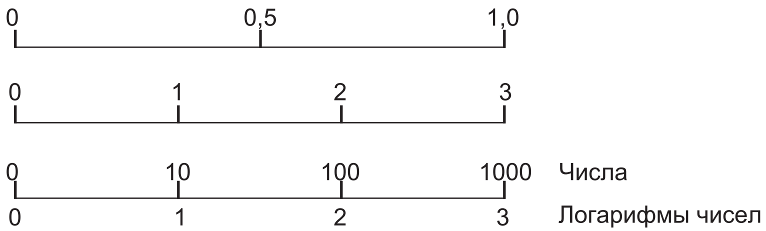

Рис. 5.4. Шкалы

Последний элемент графика – *экспликация*. Каждый график должен иметь словесное описание его содержания. Описание включает название графика, которое в краткой форме передает его содержание; надписи вдоль масштабных шкал и пояснения к отдельным частям графика.

В настоящее время задачу исследования в практическом применении графиков облегчает использование пакетов прикладных программ компьютерной графики, например, ППП: Harvard graphics, Excel, Statgraf, Supercale.

5.2 КЛАССИФИКАЦИЯ ВИДОВ ГРАФИКОВ

Существует множество видов графических изображений (рис. 5.5 и 5.6). Их классификация основана на ряде признаков, в основе которых:

- способ построения графического образа;
- геометрические знаки, изображающие статистические показатели;
- задачи, решаемые с помощью графического изображения.

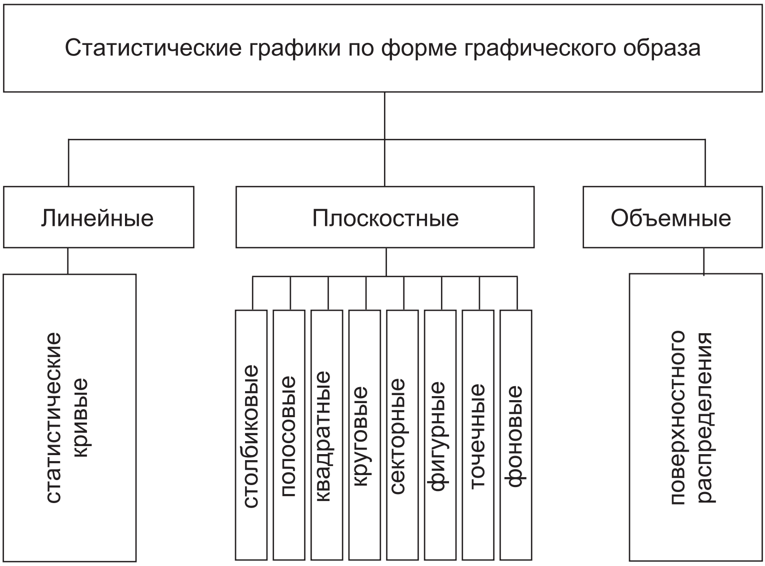

Рис. 5.5. Классификация статистических графиков по форме графического образа

По способу построения статистические графики делятся на диаграммы и статистические карты.

Диаграммы – наиболее распространенный способ графических изображений. Это графики количественных отношений. Виды и способы их построения разнообразны. Применяются диаграммы для наглядного сопоставления в различных аспектах (пространственном, временном и др.) независимых друг от друга совокупностей. При этом сравнение исследуемых совокупностей производится по какому-либо существенному варьирующему признаку.

Статистические карты – графики количественного распределения по конкретной территории. По своей основной характеристике эти графики близко примыкают к диаграммам и специфичны лишь в том отношении, что представляют собой условные изображения статистических данных на контурной географической карте. Их задачи – отражать пространственное размещение или пространственную распространенность статистических данных. Статистические карты по графическому образу делятся на *картограммы* и *картодиаграммы*.

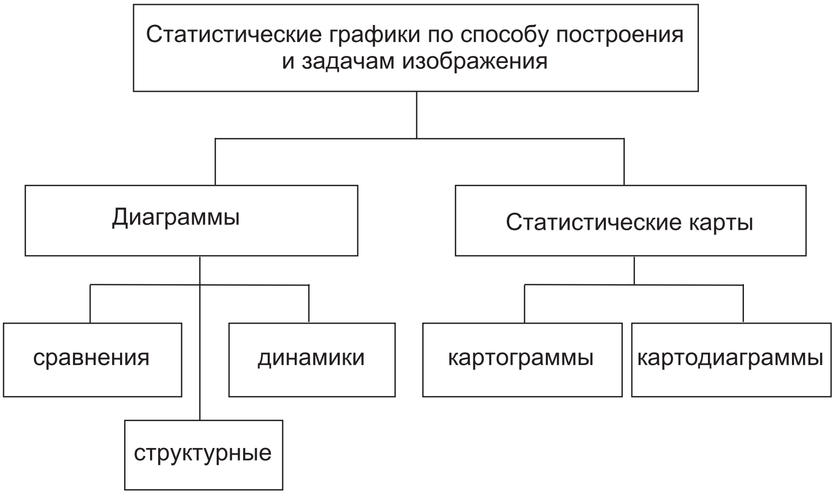

Рис. 5.6. Классификация статистических графиков по способу построения и задачам изображения

Геометрические знаки, как было сказано выше, представляют собой точки, либо линии или плоскости, либо геометрические фигуры. В соответствии с этим различают графики точечные, линейные, плоскостные и пространственные (объемные).

При построении точечных диаграмм в качестве графических образов применяются *совокупности точек*; при построении линейных – *линии*. Основной принцип построения всех плоскостных диаграмм сводится к тому, что статистические показатели изображаются в виде геометрических фигур и, в свою очередь, подразделяются на *столбиковые*, *полосовые*, *круговые, квадратные* и *фигурные*.

В зависимости от круга решаемых задач выделяют диаграммы сравнения, структурные диаграммы и диаграммы динамики.

Особым видом графиков являются диаграммы распределения величин, представленных вариационным рядом, – гистограмма, полигон, огива, кумулята[1].

В данной главе рассматриваются наиболее применяемые в статистической практике графики, на которые можно ориентироваться в известной степени при составления аналогичных графиков.

[1] Эта группа графиков рассматривалась в главе 3.

5.3
ДИАГРАММЫ СРАВНЕНИЯ

Наиболее распространенными диаграммами сравнения являются *столбиковые диаграммы*. Это графическое изображение статистических показателей в виде столбиков-прямоугольников. Данные диаграммы широко используются для наглядного сравнения объектов изучаемых явлений во времени и пространстве, а также для изображения структуры явлений.

При построении столбиковых диаграмм необходимо начертить систему прямоугольных координат, в которой расположатся столбики. На горизонтальной оси располагают основания столбиков, размер основания столбиков определяется произвольно, но должен быть одинаковым для всех.

Технику построения столбиковой диаграммы рассмотрим на данных табл. 5.1.

Таблица 5.1

Динамика численности работников научных организаций в *N*-м регионе России*

Год	1997	1998	1999	2000	2001	2002
Тыс. чел.	1943,4	1677,8	1532,6	1315,0	1106,3	990,7

* Цифры условные.

Для построения диаграммы (рис.5.7) берем систему прямоугольных координат. На оси абсцисс на одинаковом расстоянии друг от друга наносим шесть отрезков равной длины – основания для столбиков. Высота столбиков определяется в соответствии с принятым масштабом по оси ординат и значениями показателей. Учитывая размер поля графика и максимальное значение показателя, установим масштаб. Допустим, что каждым 500 тыс.чел. соответствует отрезок на оси ординат 1 см. Тогда высота столбиков будет: для 1997 г. – 3,89 см (1943,4:500); для 1998 г. – 3,36 см; для 1999 г. – 3,07 см; для 2000 г. – 2,63 см; для 2001 г. – 2,21 см и для 2002 г. – 1,98 см.

Наглядность данной диаграммы достигается сравнением высоты столбиков.

Размещение столбиков в поле графика может быть различным:
• на одинаковом расстоянии друг от друга (рис. 5.7);
• вплотную друг к другу (рис. 5.8);
• в частичном наложении друг на друга (рис. 5.9).

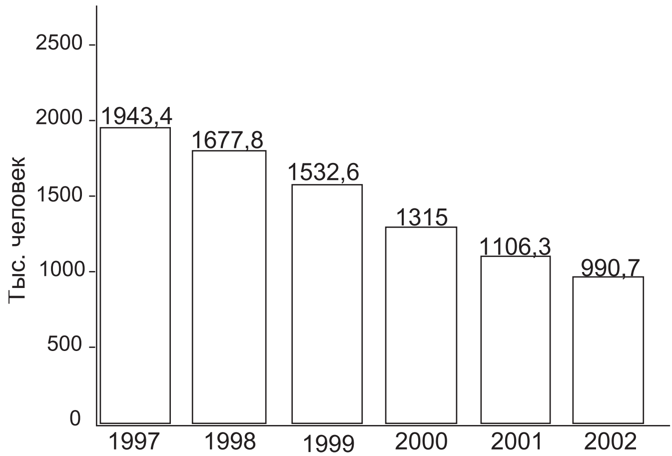

Рис. 5.7. Динамика численности работников научных организаций в *N*-м регионе России за 1997–2002 гг.

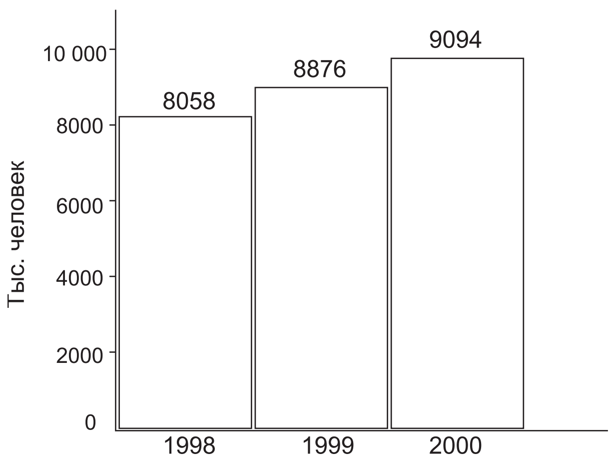

Рис. 5.8. Динамика численности безработных в РФ за 1998–2000 гг.

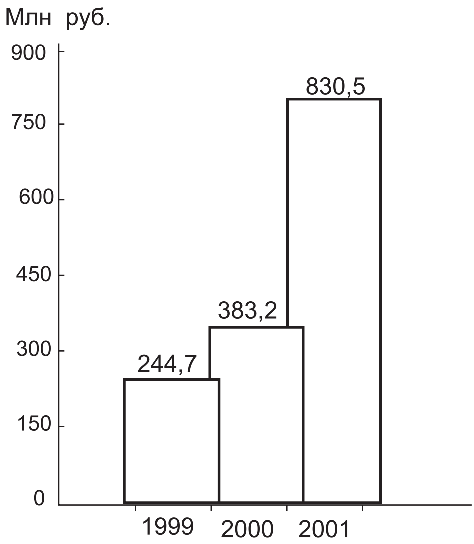

Рис. 5.9. Динамика денежных доходов населения в *N*-м регионе России за 1999–2001 гг.

Столбиковые диаграммы целесообразно применять для сравнения нескольких показателей. На рис. 5.10 посредством столбиковой диаграммы показана динамика удельного веса отдельных источников финансирования инвестиций в основной капитал за 1998–2000 годы. На этой диаграмме столбики располагаются вплотную по группам объектов. Масштаб принят такой, что каждым 5% соответствует отрезок на оси ординат 1 см. На диаграмме сделаны необходимые надписи и условные обозначения.

Разновидность столбиковых диаграмм составляют так называемые ***ленточные,*** или ***полосовые, диаграммы.*** Их отличие состоит в том, что масштабная шкала расположена по горизонтали сверху или снизу и она определяет величину полос по длине. В качестве примера приведем полосовую диаграмму сравнения, характеризующую динамику выдачи кредитов коммерческими банками России.

Для построения данной диаграммы на поле графика отложим полосы, длина которых соответствует значениям изображаемых данных на масштабной шкале (рис. 5.11).

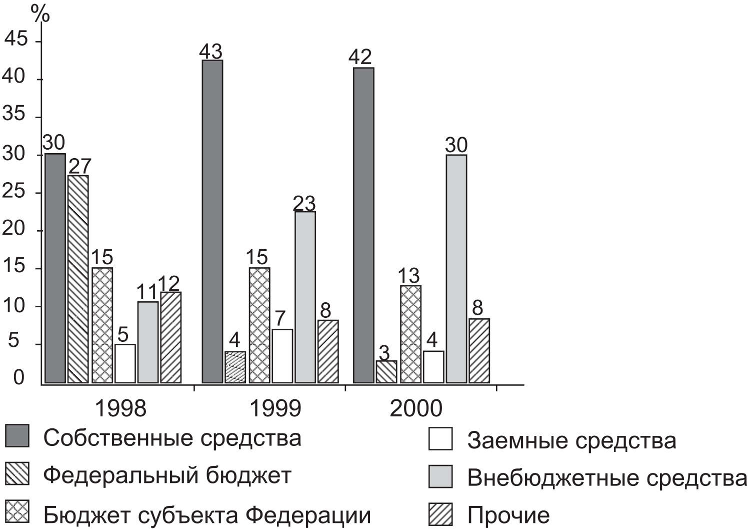

Рис. 5.10. Динамика удельного веса отдельных источников финансирования инвестиций в основной капитал за 1998–2000 гг., % к итогу

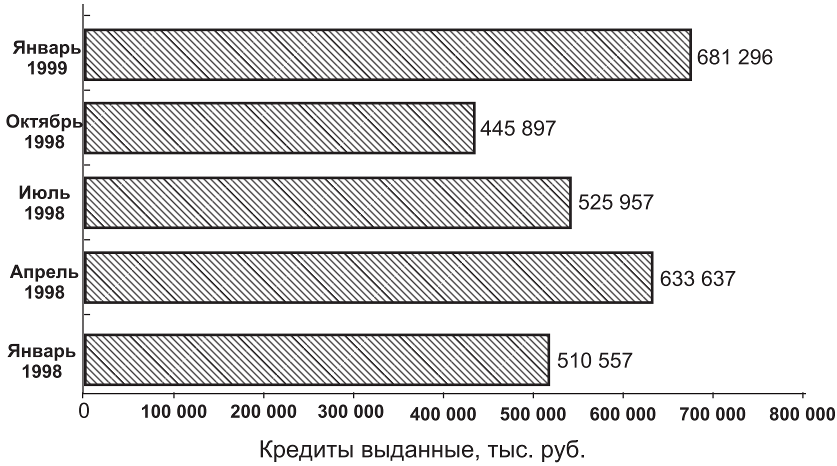

Рис. 5.11. Динамика выдачи кредитов коммерческими банками России с 01.01. 1998 г. по 01.01. 1999 г.

Из диаграммы следует, что среди выдачи кредитов коммерческими банками наибольшая их сумма приходится на 1 января 1999 г.

Столбиковые и полосовые диаграммы хорошо подходят для характеристики состава совокупности. Структура состава совокупности лучше воспринимается не в абсолютных, а в относительных величинах. При таких данных все столбики (полосы) в диаграмме имеют одинаковую высоту и соответствуют 100%. Каждый столбик разбивается на части пропорционально удельному весу отдельных частей во всей совокупности. Примером такой диаграммы является рис. 5.12, характеризующий структуру продаж по регионам за период с 1990 по 2000 г.

Рис. 5.12. Динамика структуры продаж компании «Содексо Альянс» по регионам за период с 1990 по 2000 г., %

Таким образом, область применения столбиковых и полосовых диаграмм одинакова, так как идентичны правила их построения. Одномерность изображаемых статистических показателей и их одномасштабность для различных столбиков и полос требуют выполнения единственного положения: соблюдения соразмерности (столбиков – по высоте, полос – по длине) и пропорциональности изображаемым величинам.

Для выполнения этого требования необходимо соблюдать следующие условия:

- шкала, по которой устанавливается размер столбика (полосы), должна начинаться с нуля;
- шкала должна быть непрерывной, т.е. охватывать все числа данного статистического ряда;
- разрыв шкалы и соответственно столбиков (полос) не допускается.

Невыполнение указанных правил приводит к искаженному графическому представлению анализируемого статистического материала.

Разновидностью столбиковых (ленточных) диаграмм являются *направленные диаграммы*. Они отличаются от обычных двусторонним расположением столбиков или полос и имеют начало отсчета по масштабу в середине. Обычно такие диаграммы применяются для изображения величин противоположного качественного значения. Сравнение между собой столбиков (полос), направленных в разные стороны, менее эффективно, чем расположенных рядом в одном направлении. Несмотря на это, анализ направленных диаграмм позволяет делать достаточно содержательные выводы, так как особое расположение придает графику яркое изображение. К *группе двусторонних* относятся диаграммы числовых отклонений. В них полосы направлены в обе стороны от вертикальной нулевой линии: вправо – для прироста; влево – для уменьшения. С помощью таких диаграмм удобно изображать отклонения от плана или некоторого уровня, принятого за базу сравнения. Важным достоинством рассматриваемых диаграмм является возможность видеть размах колебаний изучаемого статистического признака, что само по себе имеет большое значение для экономического анализа. Данные диаграммы широко применяются в демографии (рис. 5.13).

Для простого сравнения независимых друг от друга показателей могут также использоваться диаграммы, принцип построения которых состоит в том, что сравниваемые величины изображаются в виде правильных геометрических фигур, которые строятся так, чтобы площади их относились между собой как количества, этими фигурами изображаемые. Иными словами, эти диаграммы величину изображаемого явления выражают размером своей площади.

Для создания диаграмм такого типа используют разнообразные геометрические фигуры – квадраты, круги, реже – прямоугольники. Для построения *квадратных* и *круговых диаграмм* необходимо сна-

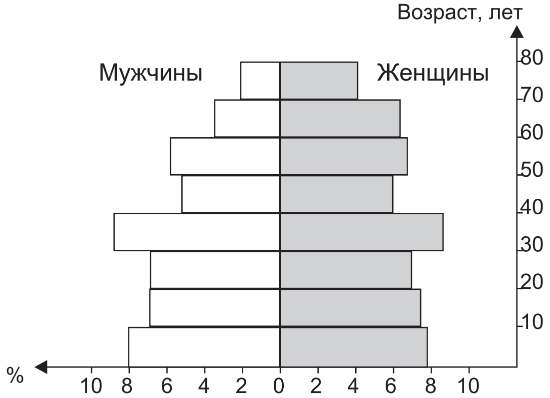

Рис. 5.13. Распределение населения одного из регионов России по полу и возрасту в 2001 г.

чала из статистических данных извлечь квадратные корни. Затем на базе полученных результатов определить сторону квадрата или радиус круга в соответствии с принятым масштабом.

Например, необходимо изобразить в виде квадрата или круга доходы от услуг связи населению России в 2000 г.: почтовая составила 4 662 млн руб., международная и междугородняя телефонная – 16 714 млн руб., городская и сельская телефонная – 14 560 млн руб.

Для построения квадратной диаграммы сначала извлечем квадратные корни из чисел: $\sqrt{4\,662} = 68,3;$ $\sqrt{16\,714} = 129,3;$ $\sqrt{14\,560} = 120,7.$

Затем установим масштаб, например, примем 1 см – 30 млн руб. Тогда сторона 1-го квадрата составит 4,3 см (129,3 : 30); 2-го – 4,0 (120,7 : 30), 3-го – 2,3 см (68,3 : 30) (рис. 5.14).

Круговая диаграмма строится аналогично квадратной с той разницей, что находим величину радиуса для каждого круга, т.е. при масштабе 1 см – 30 млн руб. радиус 1-го круга будет равен – 2,3 см, 2-го – 4,0 см и 3-го – 4,3 см (рис.5.15).

Для правильного построения диаграмм квадраты или круги необходимо расположить на одинаковом расстоянии друг от друга, а в каждой фигуре указать числовое значение, которое она изображает, не приводя масштаба измерения.

Рис. 5.14. Квадратная диаграмма доходов от услуг связи населению в России за 2000 г.

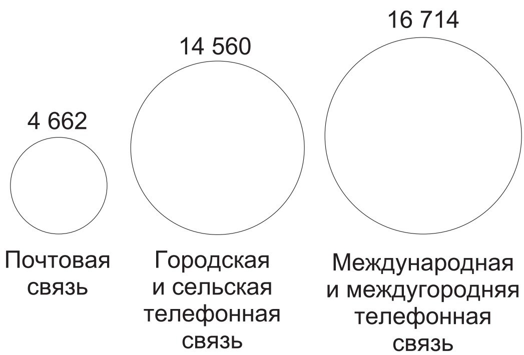

Рис. 5.15. Круговая диаграмма доходов от услуг связи населению в России за 2000 г.

К рассматриваемому виду диаграмм относится графическое изображение, полученное путем построения один в другом квадратов, кругов или прямоугольников с различной заштриховкой или закраской. Такие диаграммы также позволяют сравнивать между собой ряд исследуемых величин.

Прямоугольные диаграммы (не квадраты!) находят себе применение при графическом изображении, главным образом для двумасштабных сравнений: один масштаб для основания, другой – для высоты. Эти диаграммы называются *знаками Варзара.* Обычно они применяются в тех случаях, когда показатель является произведением двух других (например, сумма вкладов есть произведение числа вкладов на средний размер вкладов; численность населения является произведением плотности населения на территорию; валовой сбор есть произведение посевной площади на урожайность и т.д.). Такой показатель можно графически изобразить в виде сомножителей. Для этого поступают следующим образом: один сомножитель принимают за основание, другой – за высоту. Затем устанавливают масштабы: один для основания, другой для высоты. Далее, располагая значением показателя основания и высоты, строят прямоугольники. Покажем этот способ на примере данных по сбору плодов и ягод в одном из регионов России, в котором при посевной площади 936 тыс. га урожайность составила 23,6 ц/га (рис. 5.16). В нашем случае в основание прямоугольника положена урожайность плодов и ягод, высота – посевная площадь, а площадью прямоугольника является валовой сбор плодов и ягод (23,6 x 936 = 22090 тыс. т).

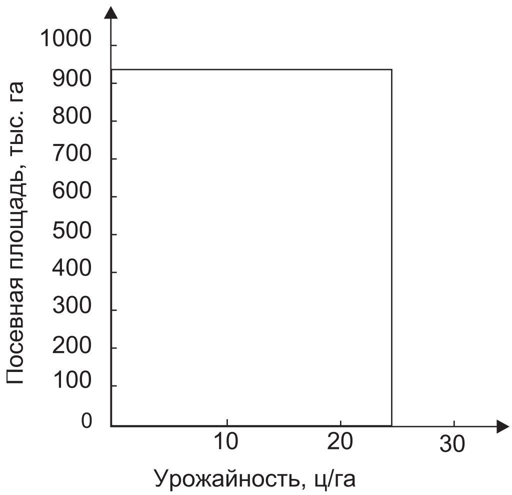

Рис. 5.16. Зависимость валового сбора плодов и ягод от урожайности и отведенной посевной площади в одном из регионов России в 2001 г.

Весьма выразителен и хорошо воспринимается способ построения *диаграмм сравнения в виде фигур-знаков*. В этом случае статистические показатели изображаются не геометрическими фигурами, а символами или знаками, воспроизводящими в какой-то степени внешний образ статистических данных. Достоинство такого способа графического изображения заключается в высокой степени наглядности, в получении подобного отображения, отражающего содержание сравниваемых совокупностей.

Важнейший признак любой диаграммы – *масштаб*. Поэтому, чтобы правильно построить фигурную диаграмму, необходимо определить *единицу счета.* В качестве последней принимается отдельная фигура (символ), которой условно присваивается конкретное численное значение. А исследуемая статистическая величина изображается отдельным количеством одинаковых по размеру фигур, последовательно располагающихся на рисунке. Однако в большинстве случаев не удается изобразить статистический показатель целым количеством фигур. Последнюю из них приходится делить на части, так как по масштабу один знак является слишком крупной единицей измерения. Обычно эта часть определяется на глаз. Сложность точного ее определения является недостатком фигурных диаграмм. Однако если большая точность представления статистических данных не преследуется, то результаты получаются вполне удовлетворительными.

Пример. Рассмотрим построение фигурной диаграммы по данным закупок крупного рогатого скота в одном из регионов России в хозяйствах всех категорий.

В 1999 г. – 6,5 млн т, в 2000 г. – 4,4 млн т, в 2001 г. – 33 млн т.

Примем условно за один знак 1,1 млн т закупок крупного рогатого скота. Разделим приведенные показатели на 1,1 и получим число фигур: для 1999 г. – 6; для 2000 г. – 4; для 2001 г. – 3. Построим диаграмму (рис.5.17).

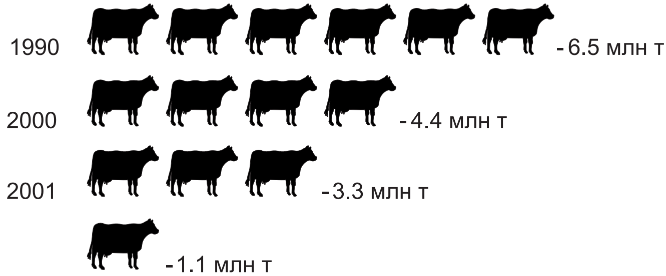

Рис. 5.17. Динамика закупок крупного рогатого скота в одном из регионов РФ в хозяйствах всех категорий (в весе живого скота) за 1999–2001 гг.

5.4
СТРУКТУРНЫЕ ДИАГРАММЫ

Основное назначение структурных диаграмм заключается в наглядной иллюстрации структуры какого-либо явления, характеристике удельных весов отдельных частей целого, выявлении структурных сдвигов.

В качестве графического образа для изображения структуры совокупностей применяются прямоугольники – для построения столбиковых и полосовых диаграмм и круги – для построения секторных диаграмм.

Круг часто используется в качестве геометрической формы при построении диаграммы. Следует различать два вида применения круга. В одном случае сравниваются площади кругов друг с другом. Такого рода диаграмма называется *круговой* (см. рис. 5.15). В другом случае круг используется для сравнения площади отдельных секторов друг с другом. Такая диаграмма именуется *секторной*.

Идея целого очень хорошо и наглядно выражается кругом, который представляет всю совокупность. Удельный вес каждой части совокупности в секторной диаграмме характеризуется величиной центрального угла (угол между радиусами круга). Сумма всех углов круга, равная 360°, приравнивается к 100%, а следовательно, 1% принимается равным 3,6°. Графическое изображение структуры с помощью столбиковой диаграммы приведено ранее (см. рис. 5.12). Построение полосовой секторной диаграммы покажем на данных, характеризующих динамику структуры денежных доходов населения в России (табл. 5.2).

Таблица 5.2

Структура денежных доходов населения России, %

Год	Оплата труда	Социальные трансферты	Доходы от собственности	Доходы от предпринимательской деятельности	Другие доходы
1992	73,6	14,3	1,0	8,4	2,7
1999	64,2	13,3	7,3	14,3	0,9

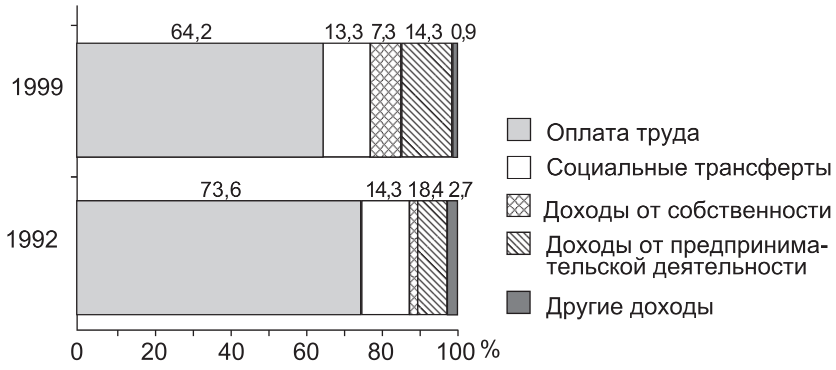

Рис. 5.18. Структура денежных доходов населения России, %

На этом рисунке каждая из полос диаграммы будет иметь одинаковую длину, так как в относительных величинах погасятся различия абсолютных размеров совокупностей. В то же время структурные различия проявятся значительно четче.

Так, приведенная на рис. 5.18 диаграмма характеризует структурные сдвиги в денежных доходах населения, в частности, в 1999 г. увеличилась доля доходов от предпринимательской деятельности в общей сумме денежных доходов.

Покажем построение секторной диаграммы по данным табл.5.3.

Таблица 5.3

Динамика структуры безработных Российской Федерации по уровню образования за 1999–2000 гг., %

Образование	1999	2000
Высшее	10,3	10,8
Незаконченное высшее	2,4	2,7
Среднее профессиональное	28,1	27,0
Среднее (полное) общее	41,9	42,3
Основное общее	14,6	13,5
Не имеет основного общего	2,7	3,7
Всего	100,0	100,0

Построение секторной диаграммы начинается с определения центральных углов секторов. Для этого процентное выражение отдельных частей совокупностей умножают на 3,6°. Например, для данных:

1999 г.: $10{,}3 \cdot 3{,}6° = 37{,}1°$; $2{,}4 \cdot 3{,}6° = 8{,}6°$; $28{,}1 \cdot 3{,}6° = 101{,}2°$;
$41{,}9 \cdot 3{,}6° = 150{,}8°$; $14{,}6 \cdot 3{,}6° = 52{,}6°$; $2{,}7 \cdot 3{,}6° = 9{,}7°$;

2000 г.: $10{,}8 \cdot 3{,}6° = 38{,}9°$; $2{,}7 \cdot 3{,}6° = 9{,}7°$; $27{,}0 \cdot 3{,}6° = 97{,}2°$;
$42{,}3 \cdot 3{,}6° = 152{,}3°$; $13{,}5 \cdot 3{,}6° = 48{,}6°$; $3{,}7 \cdot 3{,}6° = 13{,}3°$.

По найденным значениям углов круги делятся на соответствующие секторы (рис. 5.19).

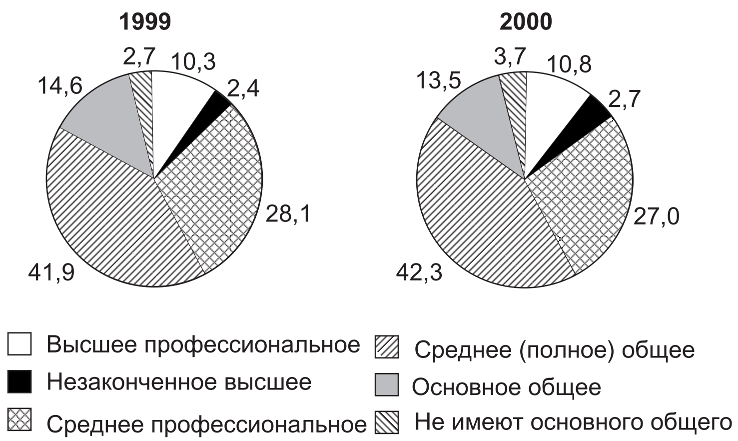

Рис. 5.19. Динамика структуры безработных РФ по уровню образования в 1999 и 2000 гг.

Применение секторных диаграмм позволяет графически не только изобразить структуру совокупности и ее изменение, но и показать динамику численности этой совокупности (см. рис. 5.19).

Рассмотренные способы графического изображения структуры совокупности имеют как достоинства, так и недостатки.

Так, секторная диаграмма сохраняет наглядность и выразительность лишь при небольшом числе частей совокупности, в противном случае ее применение малоэффективно. Кроме того, наглядность секторной диаграммы снижается при незначительных изменениях структуры изображаемых совокупностей: она выше, если имеются существенные различия сравниваемых структур. Преимуществом столбиковых (ленточных) структурных диаграмм по сравнению с секторными являются их большая емкость, возможность отразить более широкий объем полезной информации.

5.5 ДИАГРАММЫ ДИНАМИКИ

Для изображения и внесения суждений о развитии явления во времени строятся диаграммы динамики.

Для наглядного изображения явлений в рядах динамики используются диаграммы: столбиковые, ленточные, квадратные, круговые, линейные, радиальные и др. Выбор вида диаграммы зависит в основном от особенностей исходных данных, цели исследования. Например, если имеется ряд динамики с несколькими неравноотстоящими уровнями во времени (1913, 1940,1950, 1980, 1985, 2002 гг.), то часто для наглядности используют столбиковые, квадратные или круговые диаграммы. Они зрительно впечатляют, хорошо запоминаются, но не годны для изображения большого числа уровней, так как громоздки. Когда число уровней в ряду динамики велико, целесообразно применять линейные диаграммы, которые воспроизводят непрерывность процесса развития в виде непрерывной ломаной линии. Кроме того, линейные диаграммы удобно использовать, если целью исследования является изображение общей тенденции и характера развития явления; когда на одном графике необходимо изобразить несколько динамических рядов с целью их сравнения; если наиболее существенным является сопоставление темпов роста, а не уровней.

Для *построения линейных графиков* применяют систему прямоугольных координат. Обычно по оси абсцисс откладывают время (годы, месяцы и т.д.), а по оси ординат – размеры изображаемых явлений или процессов. На оси ординат наносят масштабы. Особое внимание следует обратить на их выбор, так как от этого зависит общий вид графика. Обеспечение равновесия, пропорциональности между осями координат необходимо в графике потому, что нарушение равновесия между осями координат дает неправильное изображение развития явления. Если масштаб для шкалы на оси абсцисс очень растянут по сравнению с масштабом на оси ординат, то колебания в динамике явления мало выделяются, и, наоборот, преувеличение масштаба по оси ординат по сравнению с масштабом на оси абсцисс дает резкие колебания. *Равным периодам времени* и *размерам уровня* должны соответствовать *равные отрезки масштабной шкалы*.

В статистической практике чаще всего применяются графические изображения с равномерными шкалами. По оси абсцисс они берутся пропорционально числу периодов времени, а по оси ординат – пропорционально самим уровням. Масштабом равномерной шкалы будет длина отрезка, принятого за единицу.

Пример. Рассмотрим построение линейной диаграммы на основе следующих данных (табл. 5.4).

Таблица 5.4

Динамика валового сбора зерновых культур в регионе за 1992–2001 гг.*

Год	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001
Млн т	237,4	179,2	189,1	158,2	186,8	192,2	172,6	191,7	210,1	211,3

* Цифры условные.

Изображение динамики валового сбора зерновых культур на координатной сетке с неразрывной шкалой значений, начинающихся от нуля, вряд ли целесообразно, так как 2/3 поля диаграммы остаются неиспользованными и ничего не дают для выразительности изображения. Поэтому в данных условиях рекомендуется строить шкалу без вертикального нуля, т.е. шкала значений разрывается недалеко от нулевой линии, и на диаграмму попадает лишь часть всего возможного поля графика. Это не приводит к искажениям в изображении динамики явления, и процесс его изменения рисуется диаграммой более четко (рис. 5.20).

Нередко на одном линейном графике приводится несколько кривых, которые дают сравнительную характеристику динамики различных показателей или одного и того же показателя.

Примером графического изображения сразу нескольких показателей является рис. 5.21.

Однако на одном графике не следует помещать более трех-четырех кривых, так как большое их количество неизбежно осложняет чертеж и линейная диаграмма теряет наглядность.

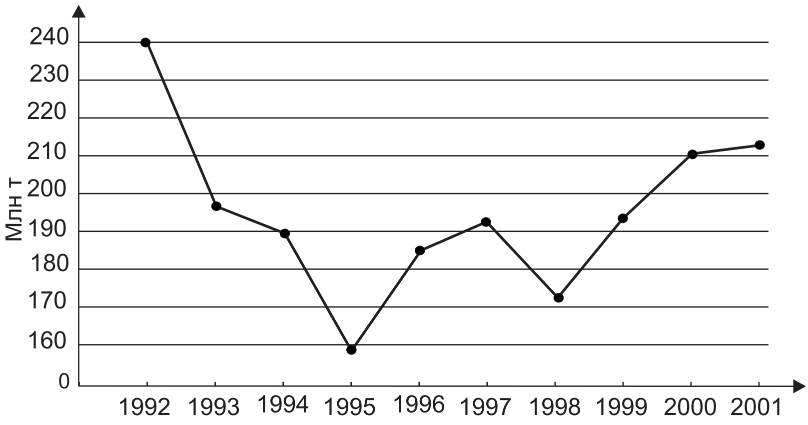

Рис. 5.20. Динамика валового сбора зерновых культур в регионе за 1992–2001 гг.

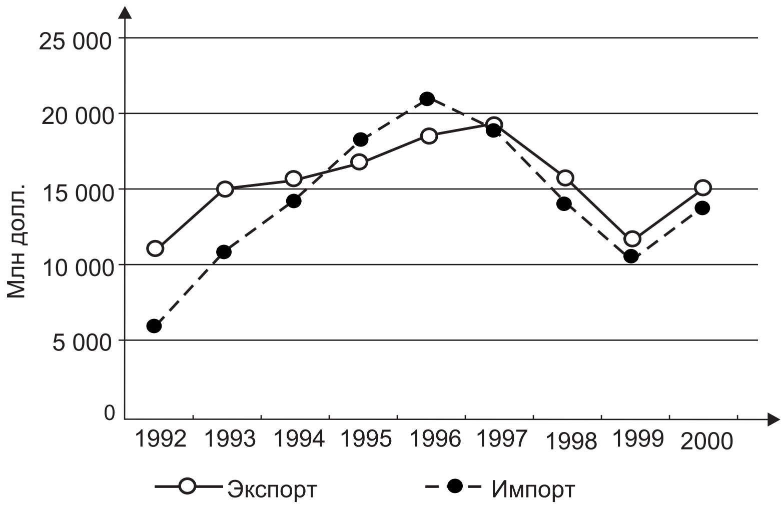

Рис. 5.21. Динамика экспорта и импорта РФ со странами СНГ за период 1992–2000 гг.

В некоторых случаях нанесения на один график двух кривых дает возможность одновременно изобразить динамику третьего показателя, если он является разностью первых двух. Например, при изображении динамики рождаемости и смертности площадь между двумя кривыми показывает величину естественного прироста или естественной убыли населения.

Иногда необходимо сравнить на графике динамику двух показателей, имеющих различные единицы измерения. В таких случаях понадобится не одна, а две масштабные шкалы. Одну из них размещают справа, другую – слева. Однако такое сравнение кривых не дает достаточно полной картины динамики этих показателей, так как масштабы произвольны. Поэтому сравнение динамики уровня двух разнородных показателей следует осуществлять на основе использования одного масштаба после преобразования абсолютных величин в относительные. Примером такой линейной диаграммы является рис. 5.22.

Линейные диаграммы с равномерной шкалой имеют один недостаток, снижающий их познавательную ценность: равномерная шкала позволяет измерять и сравнивать только отраженные на диаграм-

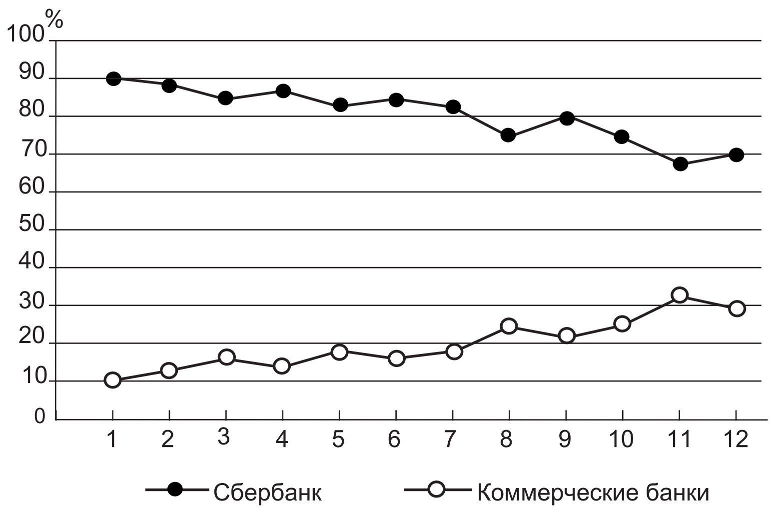

Рис. 5.22. Структура вкладов граждан в Сбербанк и коммерческие банки в одном из городов в 2001 г., %

ме абсолютные приросты или уменьшения показателей на протяжении исследуемого периода. Однако при изучении динамики важно знать относительные изменения исследуемых показателей по сравнению с достигнутым уровнем или темпы их изменения. Именно относительные изменения экономических показателей в динамике искажаются при их изображении на координатной диаграмме с равномерной вертикальной шкалой. Кроме того, в обычных координатах теряет всякую наглядность изображение для рядов динамики с резко изменяющимися уровнями, которые обычно имеют место в динамических рядах за длительный период времени.

В этих случаях следует отказаться от равномерной шкалы и положить в основу графика полулогарифмическую систему. Основная идея полулогарифмической системы состоит в том, что в ней равным линейным отрезкам соответствуют равные значения логарифмов чисел. Такой подход имеет преимущество: возможность уменьшения размеров больших чисел через их логарифмические эквиваленты. Однако с масштабной шкалой в виде логарифмов график малодоступен для понимания. Необходимо рядом с логарифмами, обозначенными на масштабной шкале, проставить сами числа, характеризующие уровни изображаемого ряда динамики, которые соответствуют указанным числам логарифмов. Такого рода графики носят название графиков на полулогарифмической сетке.

Полулогарифмической сеткой называется сетка, в которой на одной оси нанесен линейный масштаб, а на другой – логарифмический. В данном случае логарифмический масштаб наносится на ось ординат, а на оси абсцисс располагают равномерную шкалу для отсчета времени по принятым интервалам (годам, кварталам, месяцам, дням и пр.). Техника построения логарифмической шкалы следующая (рис. 5.23).

Необходимо найти логарифмы исходных чисел, начертить ординату и разделить ее на несколько равных частей. Затем нанести на ординату (или равную ей параллельную линию) отрезки, пропорциональные абсолютным приростам этих логарифмов. Далее записать соответствующие логарифмы чисел и их антилогарифмы, например (0,000; 0,3010; 0,4771; 0,6021; ...; 1,000, что дает 1, 2, 3,4, ..., 10). Полученные антилогарифмы окончательно дают вид искомой шкалы на ординате.

Пример. Построим диаграмму на полулогарифмической сетке. Допустим, что надо изобразить на графике динамику производства электроэнергии в регионе за 1995 – 2001 гг., за эти годы оно выросло в 9,1 раза. С этой целью находим логарифмы для каждого уровня ряда (табл. 5.5).

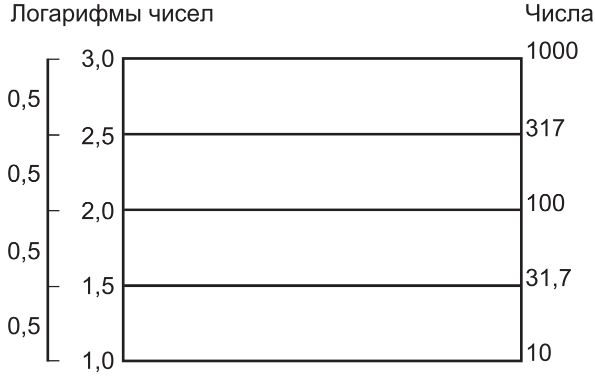

Рис. 5.23. Схема логарифмического масштаба

Определив минимальное и максимальное значение логарифмов производства электроэнергии, построим масштаб с таким расчетом, чтобы все данные разместились на графике.

Таблица 5.5

Динамика производства электроэнергии в регионе за 1995–2001 гг. (млрд кВт · ч)*

Год	y_i	$\lg y_i$	Год	y_i	$\lg y_i$
1995	170	2,23	1999	1039	3,02
1996	292	2,46	2000	1294	3,11
1997	507	2,70	2001	1544	3,19
1998	741	2,84			

* Цифры условные.

Учитывая масштаб, находим соответствующие точки, которые соединим прямыми линиями, в результате получаем график (рис. 5.24) с использованием логарифмического масштаба на оси ординат. Он называется *диаграммой на полулогарифмической сетке*. Полной логарифмической диаграммой он станет в том случае, если по оси абс-

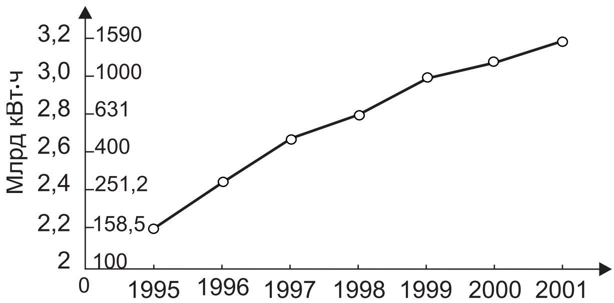

Рис. 5.24. Динамика производства электроэнергии в регионе за 1995–2001 гг.

цисс будет построен логарифмический масштаб. В рядах динамики это никогда не применяется, так как логарифмирование времени лишено всякого смысла.

Применяя логарифмический масштаб, можно без всяких вычислений характеризовать динамику уровня. Если кривая на логарифмическом масштабе несколько отклонена от прямой и становится вогнутой к оси абсцисс, значит, имеет место падение темпов; когда кривая в своем течении приближается к прямой – стабильность темпов; если она отклоняется от прямой в сторону, выпуклую к оси абсцисс, изучаемое явление имеет тенденцию к росту с увеличивающимися темпами.

Динамику изображают и *радиальные диаграммы,* строящиеся в полярных координатах. Радиальные диаграммы преследуют цель наглядного изображения определенного ритмического движения во времени. Чаще всего эти диаграммы применяются для иллюстрации сезонных колебаний. Радиальные диаграммы разделяются на замкнутые и спиральные. По технике построения радиальные диаграммы отличаются друг от друга в зависимости от того, что взято в качестве пункта отсчета – центр круга или окружность.

Замкнутые диаграммы отражают внутригодичный цикл динамики какого-либо одного года. *Спиральные диаграммы* показывают внутригодичный цикл динамики за ряд лет.

Построение замкнутых диаграмм сводится к следующему: вычерчивается круг, среднемесячный показатель приравнивается к радиусу этого круга. Затем весь круг делится на 12 радиусов, которые на графике приводятся в виде тонких линий. Каждый радиус обозначает месяц, причем расположение месяцев аналогично циферблату часов: январь – в том месте, где на часах 1, февраль – 2 и т.д. На каждом радиусе делается отметка в определенном месте согласно масштабу исходя из данных за соответствующий месяц. Если данные превышают среднемесячный уровень, отметка делается за пределами окружности на продолжении радиуса. Затем отметки различных месяцев соединяются отрезками. В приведенном примере (рис. 5.25) R = 44,8 тыс. т, длина радиуса – 3,0 см. Следовательно, 1 см = 44,8 : 3,0 ≈ 15 тыс. т. Данная замкнутая диаграмма наглядно показывает, что производство мяса подвергнуто сезонным колебаниям. Минимум производства мяса приходится на апрель, май, затем наблюдается медленное его повышение к августу, резкий подъем в сентябре, октябре и опять спад в декабре, январе. Если же в качестве базы для отсчета взять не центр круга, а окружность, то диаграммы называются спиральными.

Построение спиральных диаграмм отличается от замкнутых тем, что в них декабрь одного года соединяется не с январем данного же года, а с январем следующего года. Это дает возможность изобразить

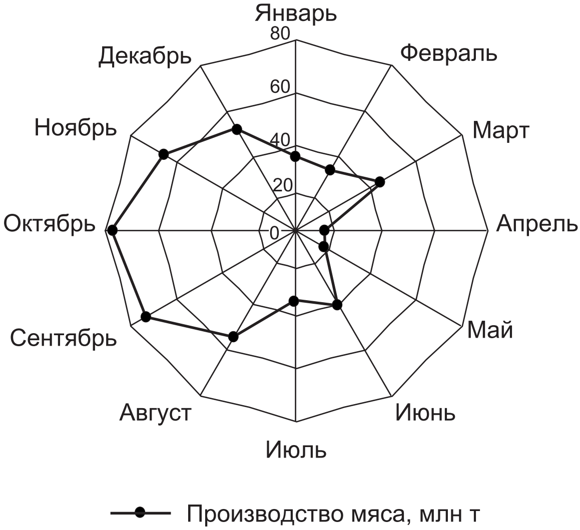

Рис. 5.25. Сезонные колебания производства мяса в одном из регионов России в 2001 г.

весь ряд динамики в виде спирали. Особенно наглядна такая диаграмма, когда наряду с сезонными изменениями происходит неуклонный рост из года в год (рис. 5.26).

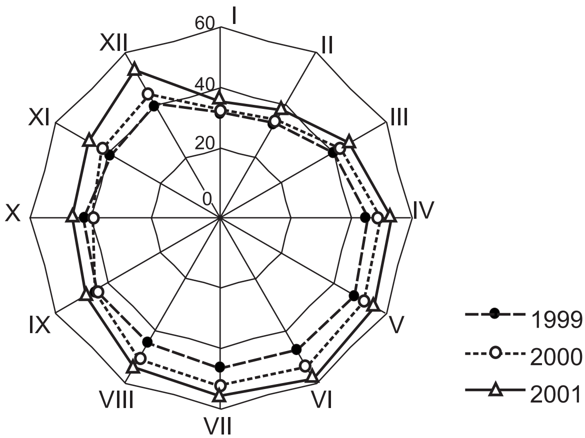

Рис.5.26. Продажа пива в розничной торговле в городе за 1999–2001 гг., л

Среди различных видов графиков особое место занимает кривая, именуемая моделью Лоренца, или *кривой Лоренца.* Данная кривая дает возможность графически изобразить уровень концентрации явления. Пример построения кривой Лоренца описан в главе 11.

5.6 СТАТИСТИЧЕСКИЕ КАРТЫ

Статистические карты представляют собой вид графических изображений статистических данных на схематической географической карте и характеризуют уровень или степень распространения того или иного явления на определенной территории.

Средствами изображения территориального размещения являются штриховка, фоновая раскраска или геометрические фигуры. Различают картограммы и картодиаграммы.

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

Картограмма фоновая – вид картограммы, на которой штриховкой различной густоты или окраской определенной степени насыщенности показывают интенсивность какого-либо показателя в пределах территориальной единицы.

Картограмма точечная – вид картограммы, где уровень выбранного явления изображается с помощью точек. Точка изображает одну единицу совокупности или некоторое их количество, показывая на географической карте плотность или частоту проявления определенного признака.

Фоновые картограммы, как правило, используются для изображения средних или относительных показателей, точечные – для объемных (количественных) показателей (численность населения, поголовье скота и т.д.).

Пример. Рассмотрим построение картограммы, используя данные табл. 5.6.

Таблица 5.6

Плотность населения восьми районов области*

№ района	1	2	3	4	5	6	7	8
Плотность населения на 1000 км^2, тыс. чел.	3,0	4,0	11,0	14,0	17,0	13,0	11,0	3,0

* Цифры условные.

Прежде чем приступить к построению картограммы, необходимо разбить районы на группы по плотности населения, а затем установить для каждой определенную окраску или штриховку.

Согласно данным табл. 5.6 все районы по плотности населения можно разбить на три группы:

1) до 4 тыс. чел.;
2) от 4 до 12 тыс. чел.;
3) от 12 до 17 тыс. чел.

Тогда к 1-й группе относятся районы № 1, 8; ко 2-й – № 2, 3, 7; к 3-й – № 4, 5, 6. Если принять для каждой группы районов окраску различной насыщенности, то на фоновой картограмме хорошо видно, как располагаются на территории области отдельные районы по плотности населения (рис. 5.27). Другим примером фоновой картограммы является рис. 5.28.

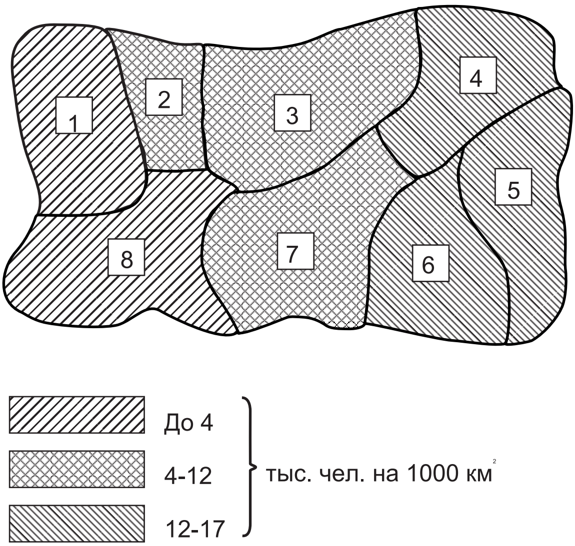

Рис. 5.27. Картограмма плотности населения восьми районов области

Вторую большую группу статистических карт составляют *картодиаграммы*, представляющие собой сочетание диаграмм с географической картой. В качестве изобразительных знаков в картодиаграммах используются диаграммные фигуры (столбики, квадраты, круги, фигуры, полосы), которые размещаются на контуре географической карты. Картодиаграммы дают возможность географически отразить более сложные статистико-географические построения, чем картограммы.

Среди картодиаграмм следует выделить картодиаграммы простого сравнения, графики пространственных перемещений, изолиний.

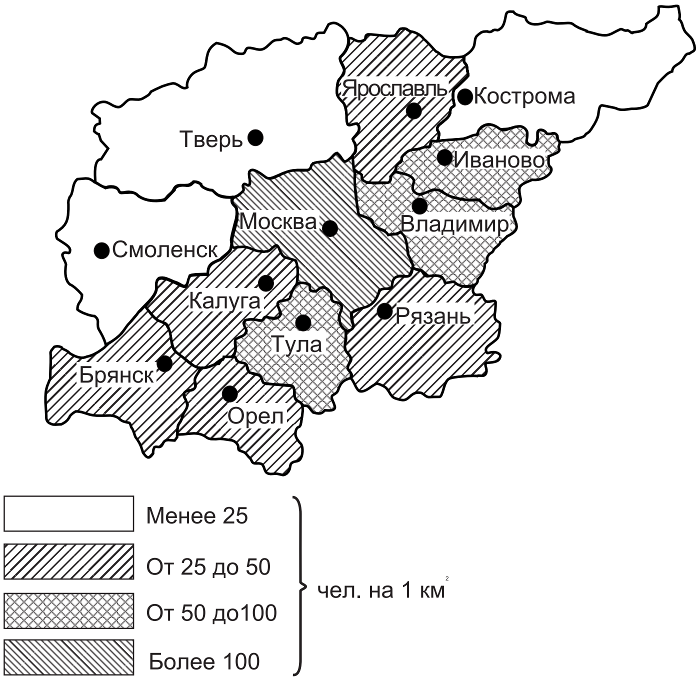

Рис. 5.28. Плотность населения в областях Центрального района России

На картодиаграмме простого сравнения в отличие от обычной диаграммы диаграммные фигуры, изображающие величины исследуемого показателя, расположены не в ряд, как на обычной диаграмме, а разносятся по всей карте в соответствии с тем районом, областью или страной, которые они представляют.

Элементы простейшей картодиаграммы можно обнаружить на политической карте, где города отличаются различными геометрическими фигурами в зависимости от числа жителей.

В качестве примера картодиаграммы возьмем изображение валового сбора зерна Центрального района России (рис. 5.29).

Изолинии (греч. isos – равный, одинаковый, подобный) – это линии равного значения какой-либо величины в ее распространении на поверхности, в частности на географической карте или графике. Изоли-

Рис. 5.29. Валовой сбор зерна в Центральном районе России (данные условные)

ния отражает непрерывное изменение исследуемой величины в зависимости от двух других переменных и применяется при картографировании природных и социально-экономических явлений. Изолинии используются для получения количественных характеристик исследуемых величин и для анализа корреляционных связей между ними.

Перечисленные виды графиков не являются исчерпывающими, но они наиболее часто употребляемы.

ОСНОВНЫЕ ПОНЯТИЯ

Абсцисса (ось x) – горизонтальная ось графика. На ней откладываются значения независимой переменной или времени, или значения признака.

Графический образ – совокупность точек, линий, фигур, с помощью которых изображаются статистические показатели.

Диаграммы динамики – линейные, спиральные, радиальные, квадратные, круговые, ленточные, фигур-знаков, секторные.

Диаграммы сравнения – столбиковые, ленточные, направленные, квадратные, круговые, фигур-знаков.

Картограмма – на схематическую географическую карту наносится штриховка различной частоты, точки или окраска определенной насыщенности, которая показывает сравнительную интенсивность какого-либо показателя в пределах каждой единицы нанесенного на карту территориального деления.

Картодиаграмма представляет собой сочетание диаграмм с географической картой.

Координаты линейной диаграммы – оси *x* и *y* графика.

Масштабная шкала – линия, отдельные точки которой могут быть прочитаны как определенные числа (прямолинейная или криволинейная).

Масштабные ориентиры – масштаб и система масштабных шкал.

Носитель шкалы – прямая или кривая линия.

Ордината (ось *y*) – вертикальная ось графика. На ней откладываются значения зависимой переменной или уровни ряда динамики, или частота повторения значений признака.

Поле графика – часть плоскости, где расположены графические образы.

Пространственные ориентиры графика – система координатных сеток.

Статистические карты – графическое изображение статистических данных на схематической географической карте, характеризующих уровень или степень распространения того или иного явления на определенной территории.

Статистический график – чертеж, на котором статистические совокупности, характеризуемые определенными показателями, описываются с помощью условных геометрических образов или знаков.

Структурные диаграммы – полосовые, столбиковые и секторные.

Экспликация – словесное описание содержания графика.

ТЕСТЫ

1. **Основными элементами статистического графика являются**:
   а) поле графика;
   б) масштабные ориентиры;
   в) геометрические знаки;
   г) экспликация графика.

2. Какие виды диаграмм можно использовать по форме геометрического образа?

а) линейные;
б) плоскостные;
в) объемные;
г) статистические карты.

3. Каковы виды статистических графиков по способу построения?

а) диаграммы;
б) статистические карты;
в) линейные;
г) плоскостные.

4. Каковы виды статистических графиков по задачам изображения социально-экономических явлений?

а) диаграммы сравнения;
б) диаграммы динамики;
в) диаграммы структуры;
г) картограммы;
д) картодиаграммы.

5. Выберите способ графического изображения данных о распределении научных работников по отраслям наук на конец года по региону:

а) картограмма;
б) картодиаграмма;
в) столбиковая;
г) секторная.

6. При изображении структуры и структурных сдвигов в совокупности явлений на графике применяются диаграммы:

а) полосовые;
б) квадратные;
в) секторные;
г) фигур-знаков.

7. При построении линейных диаграмм используются масштабные шкалы:

а) равномерные;
б) логарифмические;
в) радиальные.

8. При изображении на графике сезонных колебаний применяются диаграммы:

а) линейные;
б) радиальные;

в) спиральные;
г) столбиковые.

9. При изображении взаимосвязи между факторным и результативным признаками на графике применяются диаграммы:

а) столбиковые;
б) линейные;
в) фигур-знаков;
г) круговые.

10. При изображении социально-экономических явлений при помощи картограмм применяются их виды:

а) фоновые;
б) точечные;
в) знаков-символов.

ЛИТЕРАТУРА

Бызов Л.А. Графические методы в статистике, планировании и учете: Пособие для экономических вузов и для самообразования. – М.: Госпланиздат, 1940.

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

Едронова В.Н., Едронова М.В. Общая теория статистики. – М.: Юристъ, 2001. – 511 с.

Кан Ю. Описательная и индуктивная статистика. – М.: Финансы и статистика, 1981.

Курс лекций по общей теории статистики / Под ред. В.Е. Овсиенко. – М.: МЭСИ, 1976. – 231 с.

Ланге О., Банасиньский А. Теория статистики. – М.: Статистика, 1971. – 399 с.

ГЛАВА 6
СТАТИСТИЧЕСКИЕ ПОКАЗАТЕЛИ

6.1
ПОНЯТИЕ, ФОРМЫ ВЫРАЖЕНИЯ И ВИДЫ СТАТИСТИЧЕСКИХ ПОКАЗАТЕЛЕЙ

Статистическое исследование независимо от его масштабов и целей всегда завершается расчетом и анализом различных по виду и форме выражения статистических показателей (рис. 6.1).

Статистический показатель представляет собой количественную характеристику социально-экономических явлений и процессов в условиях качественной определенности. Качественная определенность показателя заключается в том, что он непосредственно связан с внутренним содержанием изучаемого явления или процесса, его сущностью.

Как правило, изучаемые статистикой процессы и явления достаточно сложны, и их сущность не может быть отражена посредством одного отдельно взятого показателя. В таких случаях используется система статистических показателей.

Система статистических показателей – это совокупность взаимосвязанных показателей, имеющая одноуровневую или многоуровневую структуру и нацеленная на решение конкретной статистической задачи. Так, например, сущность промышленного предприятия заключается в производстве какой-либо продукции на базе эффективного взаимодействия средств производства и трудовых ресурсов. Следовательно, для полной экономической характеристики функционирования предприятия необходимо использовать систему, включающую прежде всего такие показатели, как прибыль, рентабельность, численность промышленно-производственного персонала, производительность труда, фондовооруженность и др.

В отличие от признака, статистический показатель получается расчетным путем. Это может быть простой подсчет единиц совокупности, суммирование их значений признака, сравнение двух или нескольких величин или более сложные расчеты.

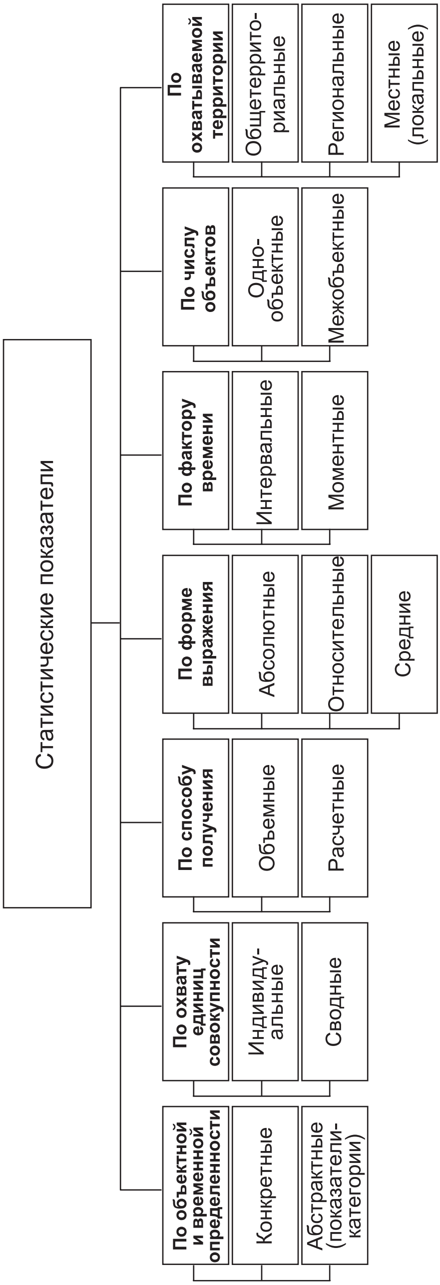

Рис. 6.1. Классификация статистических показателей

Различают конкретный статистический показатель и показатель-категорию.

Конкретный статистический показатель характеризует размер, величину изучаемого явления или процесса в данном месте и в данное время (под привязкой к месту понимается отношение показателя к какой-либо территории или объекту). Так, если мы называем конкретную величину стоимости промышленно-производственных фондов, то обязательно должны указать, к какому предприятию или отрасли и какому моменту времени она относится. Однако в теоретических работах и на этапе проектирования статистического наблюдения (при построении системы статистических показателей, обосновании методики их расчета) также оперируют и абстрактными показателями или показателями-категориями.

Показатель-категория отражает сущность, общие отличительные свойства конкретных статистических показателей одного и того же вида без указания места, времени и числового значения. Например, показатели розничного товарооборота предприятий торговли и общественного питания в Москве и Санкт-Петербурге в 2000 и 2002 гг. отличаются местом, временем и конкретными числовыми значениями, но имеют одну и ту же сущность (продажа товаров через розничную торговую сеть и сеть предприятий общественного питания), которая отражена в показателе-категории «розничный товарооборот предприятий торговли и общественного питания».

Все статистические показатели по охвату единиц совокупности разделяются на индивидуальные и сводные, а по форме выражения – на абсолютные, относительные и средние.

Индивидуальные показатели характеризуют отдельный объект или отдельную единицу совокупности – предприятие, фирму, банк, домохозяйство и т.п. Примером индивидуальных абсолютных показателей может служить численность промышленно-производственного персонала предприятия, оборот торговой фирмы, совокупный доход домохозяйства.

На основе соотнесения двух индивидуальных абсолютных показателей, характеризующих один и тот же объект или единицу, получают индивидуальный относительный показатель. В статистике рассчитываются и индивидуальные средние показатели, но только во временном измерении (среднегодовая численность персонала предприятия).

Сводные показатели, в отличие от индивидуальных, характеризуют группу единиц, представляющую собой часть статистической совокупности или всю совокупность в целом. Эти показатели, в свою очередь, подразделяются на объемные и расчетные.

Объемные показатели получают путем сложения значений признака отдельных единиц совокупности. Полученная величина, называемая объемом признака, может выступать в качестве объемного абсолютного показателя (например, стоимость основных фондов предприятий отрасли), а может сравниваться с другой объемной абсолютной величиной (например, с численностью промышленно-производственного персонала этих предприятий) или объемом совокупности (в данном примере – с числом предприятий). В последних двух случаях получают объемный относительный и объемный средний показатели (в наших примерах – фондовооруженность и средняя стоимость основных фондов).

Расчетные показатели, вычисляемые по различным формулам, служат для решения отдельных статистических задач анализа – измерения вариации, характеристики структурных сдвигов, оценки взаимосвязи и т.д. Они также делятся на абсолютные, относительные или средние. В эту группу входят индексы, коэффициенты тесноты связи, ошибки выборки и прочие показатели, подробно рассмотренные в соответствующих главах.

Охват единиц совокупности и форма выражения являются основными, но не единственными классификационными признаками статистических показателей. Важным классификационным признаком является также временной фактор. Социально-экономические процессы и явления находят свое отражение в статистических показателях либо по состоянию на определенный момент времени, как правило, на определенную дату, начало или конец месяца, года (численность населения, стоимость основных фондов, дебиторская задолженность), либо за определенный период – день, неделю, месяц, квартал, год (производство продукции, число заключенных браков, сумма страховых выплат). В первом случае показатели являются *моментными*, во втором – *интервальными*.

В зависимости от принадлежности к одному или двум объектам изучения различают *однообъектные* и *межобъектные показатели.* Если первые характеризуют только один объект, то вторые получают в результате сопоставления двух величин, относящихся к разным объектам (соотношение численности населения городов Екатеринбурга и Челябинска, соотношение численности детей дошкольного возраста и числа мест в детских дошкольных учреждениях и т.п.). Межобъектные показатели выражаются в форме относительных или средних величин.

С точки зрения пространственной определенности статистические показатели подразделяются на *общетерриториальные*, характеризующие изучаемый объект или явление в целом по стране, *региональные* и *местные (локальные),* относящиеся к какой-либо части территории или отдельному объекту.

6.2 АБСОЛЮТНЫЕ ПОКАЗАТЕЛИ

Исходной, первичной формой выражения статистических показателей являются абсолютные величины. Статистические показатели в форме абсолютных величин характеризуют абсолютные размеры изучаемых статистикой процессов и явлений: массу, площадь, объем, протяженность; отражают временные характеристики, а также могут представлять объем совокупности, т.е. число составляющих ее единиц.

Индивидуальные абсолютные показатели, как правило, получают непосредственно в процессе статистического наблюдения как результат замера, взвешивания, подсчета и оценки интересующего количественного признака. В ряде случаев индивидуальные абсолютные показатели имеют разностный характер: разность между численностью зарегистрированных безработных в данном населенном пункте на конец и на начало года, разность между выручкой от реализации торгового предприятия и общей суммой затрат и т.п.

Сводные объемные показатели, характеризующие объем признака или объем совокупности в целом как по изучаемому объекту, так и по какой-либо его части, получают в результате сводки и группировки индивидуальных значений.

Абсолютные статистические показатели всегда являются именованными числами. В зависимости от социально-экономической сущности исследуемых явлений, их физических свойств, они выражаются в натуральных, стоимостных или трудовых единицах измерения.

В международной практике используются такие *натуральные единицы измерения,* как тонны, килограммы, унции, квадратные, кубические и простые метры, мили, километры, галлоны, литры, штуки и т.д. Например, производство электроэнергии в России в 2000 г. составило 877,8 млрд кВт·ч, в этом же году добыто 313 млн т нефти и 584 млрд м^3 газа.

В группу натуральных также входят условно-натуральные измерители, используемые в тех случаях, когда какой-либо продукт имеет несколько разновидностей и общий объем можно определить только исходя из общего для всех разновидностей потребительского свойства. Так, различные виды органического топлива переводятся в условное топливо с теплотой сгорания 29,3 МДж/кг (7000 Ккал/кг); мыло разных сортов – в условное мыло с 40%-м содержанием жирных кислот; консервы различного объема – в условные консервные банки объемом 353,4 см^3 и т.д.

Перевод в условные единицы измерения осуществляется на основе специальных коэффициентов, рассчитываемых как отношение потребительских свойств отдельных разновидностей продукта к эталонному значению. Так, например, 100 т торфа, теплота сгорания которого 24 МДж/кг, будут эквивалентны 81,9 т условного топлива (100 · 24,0 / 29,3), а 100 т нефти при теплоте сгорания 45 МДж/кг оцениваются в 153,6 т условного топлива (100 · 45,0 / 29,3).

В отдельных случаях для характеристики какого-либо явления или процесса одной единицы измерения недостаточно и используется произведение двух единиц. Например, показатели грузооборота и пассажирооборота, оцениваемые соответственно в тонно-километрах и пассажиро-километрах, производство электроэнергии, измеряемое в киловатт-часах, и т.д.

В условиях рыночной экономики наибольшее значение и применение имеют *стоимостные единицы измерения,* дающие денежную оценку социально-экономическим явлениям и процессам. Так, в системе национальных счетов одним из важнейших стоимостных показателей, характеризующих общий уровень развития экономики страны, является валовой внутренний продукт, который в России в 2000 г. составил 7,1 трлн руб.

При анализе и сопоставлении стоимостных показателей в условиях высоких темпов инфляции необходимо иметь в виду, что эти показатели становятся несопоставимыми. Так, сравнивать указанный выше ВВП России в 2000 г. с его величиной в 1990 г. вряд ли целесообразно, так как содержание рубля за этот период изменилось. Для того чтобы произвести подобные сравнения, там где это возможно, осуществляют пересчет в сопоставимые цены.

К *трудовым единицам измерения,* позволяющим учитывать как общие затраты труда на предприятии, так и трудоемкость отдельных операций технологического процесса, относятся человеко-дни и человеко-часы.

6.3
ОТНОСИТЕЛЬНЫЕ ПОКАЗАТЕЛИ

Относительный показатель представляет собой результат деления одного абсолютного показателя на другой и выражает соотношение между количественными характеристиками социально-экономических процессов и явлений. Поэтому по отношению к абсолютным показателям относительные показатели или показатели в форме относительных величин являются производными (вторичными). Без относительных показателей невозможно измерить интенсивность развития изучаемого явления во времени, оценить уровень развития одного явления на фоне других взаимосвязанных с ним явлений, осуществить пространственно-территориальные сравнения, в том числе и на международном уровне.

При расчете относительного показателя абсолютный показатель, находящийся в числителе получаемого отношения, называется *текущим* или *сравниваемым*. Показатель же, с которым производится сравнение и который находится в знаменателе, называется *основанием,* или *базой сравнения*. Таким образом рассчитываемый относительный показатель указывает, во сколько раз сравниваемый абсолютный показатель больше базисного, или какую он составляет от него долю, или сколько единиц первого приходится на 1, 100, 1000 и т.д. единиц второго.

Относительные показатели могут выражаться в коэффициентах, процентах, промилле, продецимилле или быть именованными числами. Если база сравнения принимается за 1, то относительный показатель выражается в коэффициентах, если база принимается за 100, 1000 или 10 000, то относительный показатель соответственно выражается в процентах (%), промилле (‰) и продецимилле ($^0\!/_{000}$).

Проценты, как правило, используются в тех случаях, когда сравниваемый абсолютный показатель превосходит базисный не более чем в 2–3 раза. Проценты же свыше 200–300 обычно заменяются кратным отношением, коэффициентом. Так, вместо 470% говорят, что сравниваемый показатель превосходит базисный в 4,7 раза.

Относительный показатель, полученный в результате соотнесения разноименных абсолютных показателей, в большинстве случаев должен быть именованным. Его наименование представляет собой сочетание наименований сравниваемого и базисного показателей (например, производство какой-либо продукции в соответствующих единицах измерения в расчете на душу населения).

Все используемые на практике относительные статистические показатели можно подразделить на следующие виды:

- динамики;
- плана;
- реализации плана;
- структуры;
- координации;
- интенсивности и уровня экономического развития;
- сравнения.

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

$$\text{ОПД} = \frac{\text{Текущий показатель}}{\text{Предшествующий или базисный показатель}}.$$

Рассчитанная таким образом величина показывает, во сколько раз текущий уровень превышает предшествующий (базисный), или какую долю от последнего составляет. Если данный показатель выражен кратным отношением, он называется *коэффициентом роста,* при домножении этого коэффициента на 100% получают *темп роста.*

Например, если известно, что оборот торгов акциями на Московской межбанковской валютной бирже в марте 2000 г. составил 46,8 млрд руб., а в феврале – 29,0 млрд руб., то относительный показатель динамики, или темп роста, будет равен:

$$161{,}4\% = \frac{46{,}8}{29{,}0} \cdot 100\%.$$

Все субъекты финансово-хозяйственной сферы, начиная от небольших предприятий и заканчивая крупными концернами, в той или иной степени осуществляют перспективное планирование своей деятельности, а также сравнивают реально достигнутые результаты с ранее намеченными. Для этой цели используются *относительные показатели плана* (ОПП) *и реализации плана* (ОПРП):

$$\text{ОПП} = \frac{\text{Показатель, планируемый на } (i+1) \text{ период}}{\text{Показатель, достигнутый в } i\text{-м периоде}};$$

$$\text{ОПРП} = \frac{\text{Показатель, достигнутый в } (i+1) \text{ периоде}}{\text{Показатель, планируемый на } (i+1) \text{ период}}.$$

Пример. Оборот коммерческой фирмы в 2001 г. составил 2,0 млн руб. Исходя из проведенного анализа складывающихся на рынке тенденций, руководство фирмы считает реальным в следующем году довести торговый оборот до 2,8 млн руб. В этом случае относительный показатель плана, представляющий собой отношение планируемой величины к фактически достигнутой, составит 140% (2,8/2,0 · 100%). Предположим теперь, что фактический оборот фирмы за 2002 г. составил 2,6 млн руб. Тогда относительный показатель реализации плана, определяемый как отношение фактически достигнутой величины к ранее планированной, составит 92,9% (2,6/2,8 · 100%).

Между относительными показателями плана, реализации плана и динамики существует следующая взаимосвязь:

$$\text{ОПП} \cdot \text{ОПРП} = \text{ОПД}.$$

В нашем примере:

$$1{,}40 \cdot 0{,}929 = 1{,}3, \text{ или } \text{ОПД} = \frac{2{,}6}{2{,}0} = 1{,}3.$$

Основываясь на этой взаимосвязи, по любым двум известным величинам при необходимости всегда можно определить третью, неизвестную величину.

Относительный показатель структуры (ОПС) представляет собой соотношение структурных частей изучаемого объекта и их целого:

$$\text{ОПС} = \frac{\text{Показатель, характеризующий часть совокупности}}{\text{Показатель по всей совокупности в целом}}.$$

Относительный показатель структуры выражается в долях единицы или в процентах. Рассчитанные величины (d_i), соответственно называемые долями или удельными весами, показывают, какой долей обладает или какой удельный вес имеет i-я часть в общем итоге.

Рассмотрим структуру внешнеторгового оборота РФ в 2000 г. (табл. 6.1).

Рассчитанные в графе 2 табл. 6.1 проценты представляют собой относительные показатели структуры (в данном случае удельные веса).

Таблица 6.1

Структура внешнеторгового оборота РФ в 2000 г.

	Млрд долл. США	% к итогу
А	1	2
Внешнеторговый оборот – всего	150,4	100,0
в том числе:		
экспорт	105,5	70,1
импорт	44,9	29,9

Они получены как отношения объемов экспорта и импорта к общему объему внешнеторгового оборота РФ. Сумма всех удельных весов всегда должна быть строго равна 100%[1].

Относительные показатели координации (ОПК) характеризуют соотношение отдельных частей целого между собой:

$$\text{ОПК} = \frac{\text{Показатель, характеризующий } i\text{-ю часть совокупности}}{\text{Показатель, характеризующий часть совокупности, выбранную в качестве базы сравнения}}.$$

При этом в качестве базы сравнения выбирается та часть, которая имеет наибольший удельный вес или является приоритетной с экономической, социальной или какой-либо другой точки зрения. В результате определяют, сколько единиц каждой структурной части приходится на 1 единицу (иногда на 100, 1000 и т.д. единиц) базисной структурной части. Так, на основе данных приведенной выше табл. 6.1 мы можем вычислить, что на каждый миллиард долларов импорта приходилось 2,35 млрд долл. экспорта (105,5 / 44,9).

Относительный показатель интенсивности (ОПИ) характеризует степень распространения изучаемого процесса или явления в присущей ему среде:

[1] Показатели, позволяющие оценить структурные сдвиги в пространстве и во времени, будут рассмотрены в соответствующей главе.

$$\text{ОПИ} = \frac{\text{Показатель, характеризующий явление } A}{\begin{array}{c}\text{Показатель, характеризующий среду}\\ \text{распространения явления } A\end{array}}.$$

Этот показатель исчисляется, когда абсолютная величина оказывается недостаточной для обоснованных выводов о масштабе явления, его размере, насыщенности, плотности распространения. Как и в предшествующем случае, показатель может выражаться в процентах, промилле или быть именованной величиной. Например, для определения уровня рождаемости, измеряемого в ‰, рассчитывается число родившихся на 1000 человек населения, для определения плотности населения рассчитывается число людей, приходящихся на 1 км^2 территории.

Расчет относительных показателей интенсивности в ряде случаев связан с проблемой выбора наиболее обоснованной, соответствующей данному процессу или явлению базы сравнения.

Разновидностью относительных показателей интенсивности являются *относительные показатели уровня экономического развития,* характеризующие производство продукции в расчете на душу населения и играющие важную роль в оценке развития экономики государства. Так, зная лишь то, что валовой внутренний продукт России в 2000 г. составил 7,1 трлн руб., мы не можем сказать, насколько это много, оценить, «почувствовать» эту величину. Для того чтобы на основе этой цифры сделать вывод об уровне развития экономики, необходимо сопоставить ее с численностью населения страны (145,2 млн человек)[1]. В результате размер ВВП на душу населения составит 48,9 тыс. руб. (7,1 трлн руб. / 145,2 млн человек). Сделав перерасчет на доллары, данный относительный показатель уже можно использовать для временных и территориальных сравнений (между странами).

По форме выражения относительные показатели интенсивности и уровня экономического развития близки средним показателям, что нередко приводит к их смешиванию или отождествлению. Разница же между ними заключается в том, что при расчете среднего показателя мы имеем дело с совокупностью единиц, каждая из которых является носителем осредняемого признака. Например, при расчете среднедушевого дохода осредняется масса индивидуальных доходов отдельных людей. При расчете же относительного показателя интен-

[1] При отличных от нуля темпах роста численности населения необходимо использовать среднегодовое значение, исчисляемое как полусумма показателей на начало и конец периода.

сивности каждая единица не является носителем признака (при определении плотности населения отсутствует какое-либо закрепление конкретной территории за конкретными людьми).

Относительный показатель сравнения (ОПСр) представляет собой соотношение одноименных абсолютных показателей, характеризующих разные объекты (предприятия, фирмы, районы, области, страны и т.п.):

$$\text{ОПСр} = \frac{\text{Показатель, характеризующий объект } A}{\text{Показатель, характеризующий объект } Б}.$$

Например, располагая данными на конец 2000 г. о золотых запасах органов денежно-кредитного регулирования России (12,36 млн тройских унций), Канады (1,18 млн тройских унций) и США (261,61 млн тройских унций), можно на основе относительных показателей сравнения сделать вывод о том, что золотой запас нашей страны в 10,5 раза превышает золотой запас Канады, но в то же время составляет лишь 4,7% от объема золотого запаса США $\left(\frac{12{,}36}{261{,}61}\cdot 100\%\right)$.

6.4
СУЩНОСТЬ И ЗНАЧЕНИЕ СРЕДНИХ ПОКАЗАТЕЛЕЙ

Наиболее распространенной формой статистических показателей, используемых в социально-экономических исследованиях, является *средняя величина, представляющая собой обобщенную количественную характеристику признака в статистической совокупности в конкретных условиях места и времени.* Показатель в форме средней величины выражает типичные черты и дает обобщенную характеристику однотипных явлений по одному из варьирующих признаков. Он отражает уровень этого признака, отнесенный к единице совокупности. Широкое применение средних объясняется тем, что они имеют ряд положительных свойств, делающих их незаменимыми в анализе явлений и процессов общественной жизни.

Проиллюстрируем значение средних показателей на следующем примере. Одной из задач органов государственной статистики является характеристика уровня жизни населения в целом и, в частности, уровня его доходов в разрезе различных социальных групп. Очевидно, что данный объект включает столь большое число единиц, что сравнение индивидуальных доходов каждой семьи рабочего, служащего, предпринимателя, студента и т.д. является абсолютно невозможным. Не представляет особого интереса и сравнение суммарных доходов отдельных социальных групп, так как эти группы существенно различаются по численности (например, численность рабочих и численность людей, занятых в сфере предпринимательства). В данном случае мы можем использовать лишь средние показатели, а именно среднюю величину доходов в расчете на одного человека или на одну семью по каждой социальной группе.

Важнейшее свойство средней величины заключается в том, что она отражает то общее, что присуще всем единицам исследуемой совокупности. Значения признака отдельных единиц совокупности могут колебаться в ту или иную сторону под влиянием множества факторов, среди которых как основные, так и случайные. Например, доходы такой социальной группы, как студенты государственных вузов в целом определяются действующим положением о начислении стипендии. В то же время доходы отдельно взятого студента могут быть и очень большими (предположим, вследствие занятия каким-либо бизнесом в свободное от учебы время или хорошо оплачиваемых сезонных работ), и совсем отсутствовать (например, при нахождении в академическом отпуске). *Сущность средней в том и заключается, что в ней взаимопогашаются отклонения значений признака отдельных единиц совокупности, обусловленные действием случайных факторов, и учитываются изменения, вызванные действием факторов основных.* Это позволяет средней отражать типичный уровень признака и абстрагироваться от индивидуальных особенностей, присущих отдельным единицам. Возможно, что ни один студент в границах исследуемой совокупности не имеет с точностью до рубля такого дохода, какой получен на основе расчета средней. Однако эта средняя отражает тот типичный уровень доходов, который характеризует студенчество как социальную группу.

Типичность средней непосредственным образом связана с однородностью статистической совокупности. Средняя величина только тогда будет отражать типичный уровень признака, когда она *рассчитана по качественно однородной совокупности*. Так, в приведенном примере, если мы рассчитаем средний уровень доходов служащих, то получим фиктивную среднюю. Это объясняется тем, что используемая для расчета средней совокупность, включающая служащих государственных, совместных, арендных, акционерных предприятий, а также органов государственного управления, сферы науки, культуры, образования и т.п., является крайне неоднородной. В этом и подобных случаях *метод средних используется в сочетании с методом группировок*: если совокупность неоднородна – общие средние должны быть заменены или дополнены групповыми средними, т.е. средними, рассчитанными по качественно однородным группам.

Теория средних достаточно подробно разработана в отечественных и зарубежных исследованиях. Среди ученых, внесших свой вклад в ее развитие, необходимо отметить И. Зюсмильха, А. Кетле, А. Боули, К. Джини, А.Я. Боярского, Т.В. Рябушкина, И.С. Пасхавера, В.Е. Овсиенко и др.

Сущность средней можно раскрыть через *понятие ее определяющего свойства*, сформулированное А. Я. Боярским и О. Кизини: средняя, являясь обобщающей характеристикой всей статистической совокупности, должна ориентироваться на определенную величину, связанную со всеми единицами этой совокупности. Эту величину можно представить в виде функции:

$$f(x_1, x_2, \ldots, x_n). \tag{6.1}$$

Так как данная величина в большинстве случаев отражает реальную экономическую категорию, ее называют *определяющим показателем.*

Если в приведенной выше функции все величины $x_1, x_2, \ldots, x_n$ заменить их средней величиной $\bar{x}$, то значение этой функции должно остаться прежним:

$$f(x_1, x_2, \ldots, x_n) = f(\bar{x}, \bar{x}, \ldots, \bar{x}). \tag{6.2}$$

Исходя из данного равенства и определяется средняя.

Определить среднюю во многих случаях можно через *исходное соотношение средней* (ИСС) или ее логическую формулу:

$$\text{ИСС} = \frac{\text{Суммарное значение или объем осредняемого признака}}{\text{Число единиц или объем совокупности}}.$$

Так, например, для расчета средней заработной платы работников предприятия необходимо общий фонд заработной платы разделить на число работников:

$$\text{Средняя заработная плата} = \frac{\text{Фонд заработной платы, тыс. руб.}}{\text{Число работников, человек}}.$$

Числитель исходного соотношения средней представляет собой ее определяющий показатель. Для средней заработной платы таким определяющим показателем является фонд заработной платы. В любом случае независимо от того, какой первичной информацией мы располагаем (известны ли нам общий фонд заработной платы, или заработная плата и численность работников, занятых на отдельных должностях, или какие-либо другие исходные данные), среднюю заработную плату можно получить только через данное исходное соотношение средней.

Для каждого показателя, используемого в социально-экономическом анализе, можно составить только одно истинное исходное соотношение для расчета средней. Если, например, требуется рассчитать средний размер вклада в банке, то исходное соотношение будет следующим:

$$\text{Средний размер вклада} = \frac{\text{Сумма всех вкладов, тыс. руб.}}{\text{Число вкладов}}.$$

При необходимости определения средней процентной ставки по кредитам, выданным на один и тот же срок, потребуется следующее исходное соотношение:

$$\text{Средняя процентная ставка} = \frac{\text{Общая сумма выплат по процентам (из расчета за год), тыс. руб.}}{\text{Общая сумма предоставленных кредитов, тыс.руб.}} \cdot 100\%.$$

Однако от того, в каком виде представлены исходные данные для расчета средней, зависит, каким именно образом будет реализовано ее исходное соотношение. В каждом конкретном случае для реализации исходного соотношения потребуется одна из следующих *форм средней величины:*

- средняя арифметическая;
- средняя гармоническая;

- средняя геометрическая;
- средняя квадратическая, кубическая и т.д.

Перечисленные средние (кроме средней геометрической) объединяются в общей формуле *средней степенной* (при различной величине k):

$$\bar{x} = \sqrt[\kappa]{\frac{\Sigma x_i^{\kappa} \cdot f_i}{\Sigma f_i}},$$

где $\bar{x}$ – средняя величина исследуемого явления;

x_i – i-й вариант осредняемого признака ($i = \overline{1, n}$);

f_i – вес i-го варианта.

Помимо степенных средних в статистической практике также используются средние структурные, среди которых наиболее распространены мода и медиана.

При осреднении уровней динамических рядов применяются различные виды средней хронологической, которые будут рассмотрены в соответствующей главе.

6.5 СРЕДНЯЯ АРИФМЕТИЧЕСКАЯ И ЕЕ СВОЙСТВА

Наиболее распространенным видом средних величин является средняя арифметическая, которая, как и все средние, в зависимости от характера имеющихся данных может быть простой или взвешенной.

Средняя арифметическая простая (невзвешенная). Эта форма средней используется в тех случаях, когда расчет осуществляется по несгруппированным данным.

Пример. Предположим, семь членов бригады имеют следующий стаж работы:

Табельный номер рабочего	1	2	3	4	5	6	7
Стаж работы (лет)	10	3	5	12	11	7	9

Для того чтобы определить средний стаж работы, необходимо воспользоваться следующим исходным соотношением:

$$\text{Средний стаж работы} = \frac{\text{Совокупный стаж работы всех рабочих, лет}}{\text{Число рабочих}}.$$

Используя приведенные в разделе 6.4 условные обозначения, запишем формулу средней:

$$\bar{x} = \frac{x_1 + x_2 + ... + x_n}{n} = \frac{\Sigma x_i}{n} \tag{6.3}$$

С учетом имеющихся данных получим:

$$\bar{x} = \frac{10+3+5+12+11+7+9}{7} = 8{,}1 \text{ года.}$$

В этом случае мы использовали формулу средней арифметической простой (невзвешенной).

Средняя арифметическая взвешенная. При расчете средних величин отдельные значения осредняемого признака могут повторяться, встречаться по нескольку раз. В подобных случаях расчет средней производится по сгруппированным данным или вариационным рядам, которые могут быть дискретными или интервальными.

Пример. Рассмотрим следующие данные из биржевой практики (табл. 6.2).

Таблица 6.2

Продажа акций АО «Дока-хлеб» на торгах фондовой секции ТМБ «Гермес»

Сделка	Количество проданных акций, шт.	Курс продажи, руб.
1	500	1080
2	300	1050
3	1100	1145

Определим по данному дискретному вариационному ряду средний курс продажи акции, используя следующее исходное соотношение:

$$\text{Средний курс} = \frac{\text{Общая сумма сделок, руб.}}{\text{Количество проданных акций, шт.}}.$$

Чтобы получить общую сумму сделок, необходимо по каждой сделке курс продажи умножить на количество проданных акций и полученные произведения сложить. В конечном итоге результат следующий:

$$\bar{x} = \frac{1\,080 \cdot 500 + 1\,050 \cdot 300 + 1145 \cdot 1100}{500 + 300 + 1100} = \frac{2\,114\,500}{1\,900} = 1112,9 \text{ руб.}$$

Расчет среднего курса продажи произведен по формуле средней арифметической взвешенной:

$$\bar{x} = \frac{\Sigma x_i \cdot f_i}{\Sigma f_i}. \quad (6.4)$$

В отдельных случаях веса могут быть представлены не абсолютными величинами, а относительными (в процентах или долях единицы). Так, в приведенном выше примере количество проданных в ходе каждой сделки акций соответственно составляет 26,3% (0,263), 15,8% (0,158) и 57,9% (0,579) от их общего числа. Тогда с учетом несложного преобразования формулы (6.4) получим:

$$\bar{x} = \Sigma\left(x_i \cdot \frac{f_i}{\Sigma f_i}\right), \quad (6.5)$$

или $\bar{x}$ = 1080 · 0,263 + 1050 · 0,158 + 1145 · 0,579 = 1112,9 руб.

На практике наиболее часто встречающаяся при расчете средних ошибка заключается в игнорировании весов в тех случаях, когда эти веса в действительности необходимы. Предположим, имеются следующие данные (табл. 6.3).

Таблица 6.3

Заработная плата работников предприятия за май 2002 г.

Цех	Средняя заработная плата, руб.
1	4 300
2	4 100

Можно ли по имеющимся данным определить среднюю заработную плату по предприятию в целом? Можно, но только в том случае, если численность работников в 1-м и 2-м цехах совпадает. Тогда средняя заработная плата по предприятию в целом составит 4 200 руб. (доказательство этого правила будет приведено ниже). Однако в цехе 1 может быть занято, к примеру, 10 человек, а в цехе 2 – 100. Поэтому для расчета средней заработной платы потребуется средняя арифметическая взвешенная:

$$\bar{x} = \frac{4\,300 \cdot 10 + 4\,100 \cdot 100}{110} = 4\,118 \text{ руб.}$$

Общий вывод заключается в следующем: *использовать среднюю арифметическую невзвешенную можно только тогда, когда точно установлено отсутствие весов или их равенство.*

При расчете средней по *интервальному вариационному ряду* для выполнения необходимых вычислений от интервалов переходят к их серединам. Рассмотрим следующий пример (табл. 6.4).

Таблица 6.4

Распределение работников предприятия по возрасту

Возраст, лет	Число работников, чел.
До 25	7
25–30	13
30–40	38
40–50	42
50–60	16
60 и более	5
Итого	121

Для определения среднего возраста работника найдем середины возрастных интервалов. При этом величины открытых интервалов (первого и последнего) условно приравниваются к величинам интервалов, примыкающих к ним (второго и предпоследнего). Согласно вышеизложенному середины интервалов будут следующими:

22,5 27,5 35,0 45,0 55,0 65,0.

Используя среднюю арифметическую взвешенную, определим средний возраст работника данного предприятия:

$$\bar{x} = \frac{22,5 \cdot 7 + 27,5 \cdot 13 + 35 \cdot 38 + 45 \cdot 42 + 55 \cdot 16 + 65 \cdot 5}{7 + 13 + 38 + 42 + 16 + 5} = 41 \text{ год.}$$

Свойства средней арифметической. Средняя арифметическая обладает некоторыми *математическими свойствами,* более полно раскрывающими ее сущность и в ряде случаев используемыми при ее расчетах. Рассмотрим эти свойства.

1. Произведение средней на сумму частот равно сумме произведений отдельных вариантов на соответствующие им частоты:

$$\bar{x} \Sigma f_i = \Sigma x_i \cdot f_i. \qquad (6.6)$$

Действительно, если мы обратимся к приведенному выше примеру расчета среднего курса продажи акций (см. табл. 6.2), то получим следующее равенство (за счет округления среднего курса правая и левая части равенства в данном случае будут незначительно отличаться):

$$1112,9 \cdot 1900 = 1080 \cdot 500 + 1050 \cdot 300 + 1145 \cdot 1100.$$

2. Сумма отклонений индивидуальных значений признака от средней арифметической равна нулю:

$$\sum (x_i - \bar{x}) \cdot f_i = 0. \qquad (6.7)$$

Для нашего примера:

$$(1080 - 1112,9) \cdot 500 + (1050 - 1112,9) \cdot 300 + \\ + (1145 - 1112,9) \cdot 1100 = 0.$$

Математическое доказательство данного свойства сводится к следующему:

$$\Sigma(x_i-\bar{x})\cdot f_i=\Sigma x_i\cdot f_i-\Sigma\bar{x}\cdot f_i=\Sigma x_i\cdot f_i-\bar{x}\Sigma f_i=0.$$

3. Сумма квадратов отклонений индивидуальных значений признака от средней арифметической меньше, чем сумма квадратов их отклонений от любой другой произвольной величины C:

$$\begin{gathered}\Sigma(x_i-C)^2\cdot f_i=\Sigma(x_i\cdot\bar{x}+\bar{x}-C)^2\cdot f_i=\Sigma\left[(x_i-\bar{x})+(\bar{x}-C)\right]^2\cdot f_i=\\=\Sigma\left[(x_i-\bar{x})^2+2\cdot(x_i-\bar{x})\cdot(\bar{x}-C)+(\bar{x}-C)^2\right]\cdot f_i=\\=\Sigma(x_i-\bar{x})^2\cdot f_i+2\cdot(\bar{x}-C)\Sigma(x_i-\bar{x})\cdot f_i+\Sigma(\bar{x}-C)^2\cdot f_i=\\=\Sigma(x_i-\bar{x})^2\cdot f_i+2\cdot(\bar{x}-C)\cdot 0+\Sigma(\bar{x}-C)^2\cdot f_i.\end{gathered}$$

Следовательно, сумма квадратов отклонений индивидуальных значений признака от произвольной величины C больше суммы квадратов их отклонений от своей средней на величину:

$$\Sigma(\bar{x}-C)^2\cdot f_i\ \text{или}\ (\bar{x}-C)^2\cdot\Sigma f_i. \tag{6.8}$$

На использовании этого свойства базируется расчет центральных моментов, представляющих собой характеристики вариационного ряда при С $=\bar{x}$ [1].

$$\mu_k=\frac{\sum(x_i-\bar{x})^k\cdot f_i}{\Sigma f_i},$$

где k определяет порядок момента (центральный момент второго порядка представляет собой дисперсию).

4. Если все осредняемые варианты уменьшить или увеличить на постоянное число A, то средняя арифметическая соответственно уменьшится или увеличится на ту же величину:

[1] При $C = 0$ получают начальные моменты (начальный момент первого порядка – средняя арифметическая и т.д.).

$$\frac{\Sigma(x_i \pm A)\cdot f_i}{\Sigma f_i} = \frac{\Sigma x_i \cdot f_i}{\Sigma f_i} \pm \frac{\Sigma A \cdot f_i}{\Sigma f_i} = \bar{x} \pm A. \qquad (6.9)$$

Так, если все курсы продажи акций увеличить на 100 руб., то средний курс также увеличится на 100 руб.:

$$\bar{x} = \frac{1180\cdot 500 + 1150\cdot 300 + 1245\cdot 1100}{1\,900} = 1\,212,9 \text{ руб.}$$

5. Если все варианты значений признака уменьшить или увеличить в A раз, то средняя также соответственно увеличится или уменьшится в A раз:

$$\frac{\Sigma(x_i / A)\cdot f_i}{\Sigma f_i} = \frac{(1/A)\Sigma x_i \times f_i}{\Sigma f_i} = \frac{1}{A}\bar{x}. \qquad (6.10)$$

Предположим, курс продажи в каждом случае возрастет в 1,5 раза. Тогда и средний курс увеличится на 50%.

$$\bar{x} = \frac{1\,080\cdot 1,5\cdot 500 + 1\,050\cdot 1,5\cdot 300 + 1145\cdot 1,5\cdot 1100}{1900} =$$
$$= 1112,9\cdot 1,5 = 1\,669,4 \text{ руб.}$$

6. Если все веса уменьшить или увеличить в А раз, то средняя арифметическая от этого не изменится:

$$\frac{\Sigma x_i \cdot (f_i / A)}{\Sigma(f_i / A)} = \frac{(1/A)\Sigma x_i \cdot f_i}{(1/A)\Sigma f_i} = \bar{x}. \qquad (6.11)$$

Так, в нашем примере удобнее было бы рассчитывать среднюю, предварительно поделив все веса на 100:

$$\bar{x} = \frac{1\,080\cdot 5 + 1\,050\cdot 3 + 1145\cdot 11}{19} = 1112,9 \text{ руб.}$$

Исходя из данного свойства можно заключить, что в случае равенства всех весов между собой расчеты по средней арифметической взвешенной и средней арифметической простой приведут к одному и тому же результату.

6.6 ДРУГИЕ ВИДЫ СРЕДНИХ

При расчете статистических показателей помимо средней арифметической могут использоваться и другие виды средних. Однако в каждом конкретном случае в зависимости от характера имеющихся данных существует только одно истинное среднее значение показателя, являющееся следствием реализации его исходного соотношения.

Средняя гармоническая взвешенная. Рассмотрим вариант, когда известен числитель исходного соотношения средней, но неизвестен его знаменатель (табл. 6.5).

Таблица 6.5

Валовой сбор и урожайность зерновых культур по Уральскому федеральному округу в 2000 г.
(в хозяйствах всех категорий)

Область	Валовой сбор, тыс. т	Σрожайность, ц/га
Курганская	1088,3	10,9
Свердловская	603,5	12,1
Тюменская	1171,5	17,5
Челябинская	1050,7	8,8

Средняя урожайность любой сельскохозяйственной культуры по нескольким территориям, агрофирмам, крестьянским хозяйствам и т.п. может быть определена только на основе следующего исходного соотношения:

$$\text{Средняя урожайность} = \frac{\text{Общий валовой сбор, тыс. ц}}{\text{Общая посевная площадь, тыс. га}}.$$

Общий валовой сбор мы получим простым суммированием валового сбора по областям. Данные же о посевной площади в таблице отсутствуют, но их можно получить, разделив валовой сбор по каждой области на урожайность. С учетом этого определим искомую среднюю, предварительно переведя для сопоставимости тонны в центнеры:

$$\bar{x} = \frac{10\,883 + 6\,035 + 11\,715 + 10\,507}{\dfrac{10\,883}{10,9} + \dfrac{6\,035}{12,1} + \dfrac{11\,715}{17,5} + \dfrac{10\,507}{8,8}} = \frac{39\,140}{3\,360,6} = 11,6 \text{ ц/га.}$$

Таким образом, общая посевная площадь зерновых культур по Уральскому федеральному округу составила 3 360,6 тыс. га, а средняя урожайность – 11,6 ц с 1 га.

В данном случае расчет произведен по формуле средней гармонической взвешенной:

$$\bar{x} = \frac{\Sigma w_i}{\Sigma \dfrac{w_i}{x_i}} \cdot \qquad (6.12)$$

где $w_i = x_i \cdot f_i$.

Средняя гармоническая невзвешенная. Эта форма средней, используемая значительно реже, имеет следующий вид:

$$\bar{x} = \frac{n}{\sum \dfrac{1}{x_i}} . \qquad (6.13)$$

Пример. Для иллюстрации области применения средней гармонической невзвешенной воспользуемся упрощенными условными данными. Предположим, в автохозяйстве эксплуатируются два электромобиля разных моделей, работающих на однотипных подзаряжаемых за ночь аккумуляторных батареях. Первый электромобиль расходует на 1 км пути 1,0 кВт·ч электроэнергии, второй – 0,6 кВт·ч. Каков средний расход электроэнергии на 1 пройденный километр?

На первый взгляд решение этой задачи заключается в осреднении индивидуальных значений потребления электроэнергии по двум электромобилям, т.е. (1,0 + 0,6) : 2 = 0,8 кВт·ч. Проверим обоснованность такого подхода на примере одного дня работы машин, в течение которого они расходуют один заряд аккумулятора, предположим, 60,0 кВт·ч (как будет показано ниже, конкретная цифра значения не имеет). За этот день первая машина пройдет 60 км (60,0 / 1,0), пробег второй составит 100 км (60,0 / 0,6), т.е. в сумме – 160 км. Если же заменить индивидуальные значения признака их предполагаемым средним значением, то общий пробег, выступающий в данном случае в качестве определяющего показателя, сократится до 150 км (60,0 / 0,8 + 60,0 / 0,8). Следовательно, полученная средняя рассчитана неверно.

Рассмотрим решение данной задачи через исходное соотношение средней. Для того чтобы определить средний расход энергии на 1 пройденный километр, необходимо общий расход энергии за какой-либо временной промежуток (день, неделю, месяц) поделить на сделанный за это время суммарный пробег:

$$\bar{x} = \frac{60{,}0+60{,}0}{\dfrac{60{,}0}{1{,}0}+\dfrac{60{,}0}{0{,}6}} = \frac{1+1}{\dfrac{1}{1}+\dfrac{1}{0{,}6}} = \frac{2}{1+1{,}667} = 0{,}75 \text{ кВт} \cdot \text{ч}.$$

При замене индивидуального значения признака их средней величиной общий пробег не изменится:

$$\frac{60{,}0}{0{,}75} + \frac{60{,}0}{0{,}75} = 160 \text{ км}.$$

Подведем итог: средняя гармоническая невзвешенная может использоваться вместо взвешенной в тех случаях, когда значения w_j для единиц совокупности равны (машины расходуют ежедневно одно и то же количество электроэнергии).

Средняя геометрическая. Еще одной формулой, по которой может осуществляться расчет среднего показателя, является средняя геометрическая

- невзвешенная:

$$\bar{x} = \sqrt[n]{x_1 \cdot x_2 \cdot x_3 \cdot \ldots \cdot x_n} = \sqrt[n]{\Pi x_i},$$

- взвешенная:

$$\bar{x} = \sqrt[\Sigma f]{(x_1)^{f_1} \cdot (x_2)^{f_2} \cdot (x_3)^{f_3} \cdot \ldots \cdot (x_n)^{f_n}} = \sqrt[\Sigma f]{\Pi (x_i)^{f_i}}. \qquad (6.14)$$

Наиболее широкое применение этот вид средней получил в анализе динамики для определения среднего темпа роста (подробнее см. гл. 10).

Средняя квадратическая. В основе вычислений ряда сводных расчетных показателей лежит средняя квадратическая

- невзвешенная:

$$\bar{x} = \sqrt{\frac{\Sigma x_i^2}{n}}, \tag{6.15}$$

- взвешенная:

$$\bar{x} = \sqrt{\frac{\Sigma x_i^2 f_i}{\Sigma f_i}}.$$

Наиболее широко этот вид средней используется при расчете показателей вариации (глава 7).

В статистической практике также находят применение степенные средние 3-го и более высоких порядков.

ОСНОВНЫЕ ПОНЯТИЯ

Абсолютный показатель – показатель в форме абсолютной величины, отражающий физические свойства, временные или стоимостные характеристики социально-экономических процессов и явлений.

Объем признака – суммарное значение изучаемого признака по всем единицам совокупности.

Относительный показатель – показатель в форме относительной величины, получаемый как результат деления одного абсолютного показателя на другой и отражающий соотношение между количественными характеристиками изучаемых процессов и явлений.

Система статистических показателей – совокупность взаимосвязанных показателей, имеющая одноуровневую или многоуровневую структуру и нацеленная на решение конкретной статистической задачи или комплекса задач.

Средний показатель – показатель в форме средней величины, представляющий собой обобщенную количественную характеристику признака в статистической совокупности в конкретных условиях места и времени.

Статистический показатель – количественная характеристика социально-экономических явлений и процессов в условиях качественной определенности.

Средняя величина является наиболее ценной с аналитической точки зрения и универсальной формой выражения статистических показателей. Наиболее распространенная средняя – средняя арифметическая – обладает рядом математических свойств, которые могут быть использованы при ее расчете. В то же время при исчислении конкретной средней всегда целесообразно опираться на ее логическую формулу, представляющую собой отношение объема признака к объему совокупности. Для каждой средней существует только одно истинное исходное соотношение, для реализации которого, в зависимости от имеющихся данных, могут потребоваться различные формы средних. Однако во всех случаях, когда характер осредняемой величины подразумевает наличие весов, нельзя вместо взвешенных формул средних использовать их невзвешенные формулы.

ТЕСТЫ

1. К какому виду по степени охвата единиц совокупности относится показатель «активы коммерческого банка»:

а) индивидуальному;

б) сводному.

2. К какому виду по временному фактору относится показатель «число рекламаций на продукцию предприятия»:

а) моментному;

б) интервальному.

3. Чтобы получить относительный показатель динамики с переменной базой сравнения для *i*-го периода, необходимо:

а) перемножить относительные показатели динамики с постоянной базой сравнения за i-й и (i–1)-й периоды;

б) разделить относительный показатель динамики с постоянной базой сравнения за i-й период на аналогичный показатель за период $(i - 1)$;

в) разделить относительный показатель динамики с постоянной базой сравнения за i-й период на аналогичный показатель за период $(i + 1)$.

4. Относительный показатель реализации предприятием плана производства продукции составил 103%, при этом объем производства по сравнению с предшествующим периодом вырос на 2%. Что предусматривал план:

а) снижение объема производства;

б) рост объема производства.

5. Сумма относительных показателей координации, рассчитанных по одной совокупности, должна быть:

а) строго равной 100;

б) меньше 100 или равной 100;

в) меньше, больше или равной 100.

6. Объем совокупности – это:

а) сумма всех значений осредняемого признака по совокупности;

б) общее число единиц в совокупности.

7. В каких случаях взвешенные и невзвешенные средние равны между собой:

а) при отсутствии весов;

б) при равенстве весов;

в) при отсутствии или равенстве весов.

8. В каких случаях используется средняя гармоническая:

а) когда неизвестен числитель исходного соотношения;

б) когда неизвестен знаменатель исходного соотношения.

9. Если веса осредняемого показателя выражены в промилле, чему будет равен знаменатель при расчете средней арифметической:

а) 100;

б) 1000;

в) 10000.

10. Изменится ли средняя величина, если все веса уменьшить на некоторую постоянную величину:

а) изменится;

б) не изменится.

ЛИТЕРАТУРА

Боярский А.Я. Теоретические исследования по статистике: Сб. научных трудов. – С.: Статистика, 1974. –304 с.

Венецкий И.Т., Венецкая В.И. Основные математико-статистические понятия и формулы в экономическом анализе: Справочник. 2-е изд. – М.: Статистика, 1979. – 447 с.

Джини К. Средние величины. – М.: Статистика, 1970. – 447 с.

Кимбл Г. Как правильно пользоваться статистикой/Пер. с англ. – М.: Финансы и статистика, 1982. – 294 с.

Овсиенко В.Е. Выбор формы средней и о некоторых ошибках, допускаемых в этом вопросе // Вестник статистики. – 1989. – № 2. – С. 16–24.

Пасхавер И.С. Средние величины в статистике. – М.: Статистика, 1979. – 279 с.

Статистический словарь/Под ред. М.А. Королева. – М.: Финансы и статистика, 1989. – 623 с.

РАЗДЕЛ II

АНАЛИТИЧЕСКАЯ СТАТИСТИКА

ГЛАВА 7

ПОКАЗАТЕЛИ ВАРИАЦИИ И АНАЛИЗ ЧАСТОТНЫХ РАСПРЕДЕЛЕНИЙ

7.1 ВАРИАЦИЯ ПРИЗНАКА В СОВОКУПНОСТИ И ЗНАЧЕНИЕ ЕЕ ИЗУЧЕНИЯ

При изучении социально-экономических явлений и процессов статистика встречается с разнообразной вариацией признаков, характеризующих отдельные единицы совокупности. Величины признаков колеблются, варьируют под действием различных причин и условий, которые в статистике называются *факторами*. Нередко эти факторы действуют в противоположных направлениях и сами, в свою очередь, варьируют. Среди них есть существенные факторы, определяющие величину вариантов данного признака у всех единиц совокупности. Но есть и несущественные (чисто случайные), которые на одни единицы совокупности могут оказывать влияние, на другие нет. Например, вариация оценок студентов на экзамене в вузе вызывается, в частности, различными способностями студентов; временем, затраченным ими на самостоятельную работу; посещаемостью занятий; различием социально-бытовых условий и т.д. Но на оценку могут влиять и какие-либо привходящие, чисто случайные причины, например, временное недомогание. Вариация, порождаемая существенными факторами, носит систематический характер, т.е. наблюдается последовательное изменение вариантов признака в определенном направлении. Такая вариация называется *систематической*. В систематической вариации проявляются взаимосвязи между явлениями, их признаками, в такой связи – один как причина (фактор), другой как следствие (результат) его действия. Точнее говоря, проявляется зависимость вариации одного признака от вариации другого или от нескольких других.

Вариация, обусловленная случайными факторами, называется *случайной вариацией.* Здесь не наблюдается систематического изменения вариантов зависимого признака от случайных факторов; все изменения носят хаотический характер, поскольку нет устойчивой связи этих факторов с единицами изучаемой совокупности.

Вариация зависимого признака, образовавшаяся под действием всех без исключения влияющих на него факторов, называется *общей вариацией.* Следовательно, общая вариация слагается из систематической и случайной вариации. Но систематическая вариация, если между признаками имеется довольно существенная связь, в конце концов пробивает себе дорогу через хаос случайных колебаний вариантов зависимого признака и проявляет себя.

Пример. Возьмем следующие два варьирующих признака: возраст матерей и долю мальчиков среди детей, родившихся у матерей до 45-летнего возраста, полученные в результате социально-демографического обследования в 2002 г. в одном из регионов (табл. 7.1).

Таблица 7.1

Доля мальчиков, родившихся у матерей до 45 лет

Возраст матери, лет	Моложе 20	20–24	25–29	30–34	35–39	40–44
На 1000 девочек приходится мальчиков	1 057	1 054	1 051	1 046	1 044	1 035

Из данных табл. 7.1 видно, что от возраста матери зависит доля мальчиков среди новорожденных. Здесь наблюдается систематическая вариация зависимого признака. Оба признака варьируют, но вариация зависимого признака идет в противоположном направлении по сравнению с вариацией факторного признака. Чем выше возраст матерей, тем ниже доля мальчиков. Иначе говоря, по мере увеличения возраста матери вероятность рождения мальчика несколько уменьшается и в такой же мере возрастает вероятность рождения девочки.

Наличие вариации признаков, изучаемых статистикой явлений, ставит задачу определить меру вариации, ее измерение, найти соответствующие измерители – показатели, характеризующие размеры этой вариации, а также выявить сущность и методы вычисления определяющих ее факторов.

По степени вариации изучаемые явления можно рассматривать с различных аспектов, в частности, судить об однородности совокупности, устойчивости индивидуальных значений признака, типичности средней, о взаимосвязи между признаками одного и того же явления и признаками разных явлений. Статистические показатели, характеризующие вариацию, широко применяются в практической деятельности, например для оценки ритмичности работы промышленных предприятий, используются как контроль над производственными процессами, а также для определения устойчивости урожайности сельскохозяйственных культур тех или иных сортов или одного и того же сорта в определенных климатических условиях. На основе вариации в статистике разрабатываются показатели, характеризующие социально-экономические явления и процессы, например показатели тесноты связи между явлениями и их признаками, показатели оценки точности выборочного наблюдения (см. глава 8).

В настоящей главе в основном рассматриваются приемы изучения *случайной вариации*, т.е. вариации количественного признака в однородной совокупности. Совокупность значений изучаемого признака с указанием числа различных значений называется *распределением признака*. Распределение представляют в виде *вариационного ряда* (см. глава 3, раздел 3.5).

Каким же образом статистика дает количественную оценку степени колеблемости признака в совокупности, измеряет вариацию?

Для характеристики закономерностей распределения изучаемого признака недостаточно пользоваться только вариационными рядами распределения и их графическим изображением. В процессе анализа требуется вычислить различные числовые характеристики (показатели), которые в обобщенном виде отразят *особенности распределения изучаемых признаков*. Наличие таких характеристик (показателей) существенно облегчает *сравнение различных распределений* между собой.

Все показатели вариации в зависимости от характеризуемых ими особенностей можно разделить на три группы:

- показатели центра распределения – средняя арифметическая, мода и медиана;

- показатели степени вариации – вариационный размах, среднее линейное отклонение, дисперсия, среднее квадратическое отклонение и коэффициент вариации;
- показатели типа (формы) распределения – структурные характеристики, показатели асимметрии и эксцесса, кривые распределения.

7.2 ПОКАЗАТЕЛИ ЦЕНТРА РАСПРЕДЕЛЕНИЯ

Важнейшей характеристикой центра распределения является средняя арифметическая ($\bar{x}$). Для вычисления по данным *первичного ряда* применяется формула простой средней арифметической.

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}. \tag{7.1}$$

Расчет средней арифметической по формуле (7.1) см. в табл. 7.6.

При вычислении по данным *ранжированного вариационного ряда* применяется формула средней взвешенной:

$$\bar{x} = \frac{\sum_{i=1}^{k} x_i \cdot f_i}{\sum_{i=1}^{k} f_i}. \tag{7.2}$$

Расчет средней арифметической по формуле (7.2) для ранжированного вариационного ряда приведен в табл. 7.5.

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.

Модой распределения (Мо) называется такая величина изучаемого признака, которая в данной совокупности встречается наиболее часто, т.е. один из вариантов признака повторяется чаще, чем все другие.

Рассмотрим определение моды по несгруппированным данным.

Пример. Рабочие бригады из 11 человек имеют следующие тарифные разряды: 5, 4, 3, 4, 5, 5, 6, 2, 6, 3, 5. Так как в данной бригаде больше всего рабочих 5-го разряда, этот тарифный разряд и будет модальным.

Для упорядоченного дискретного ряда распределения мода, являющаяся характеристикой вариационного ряда, определяется по *частотам вариантов* и соответствует *варианту с наибольшей частотой.*

Пример. Распределение рабочих всего предприятия в целом по тарифному разряду имеет следующий вид (табл. 7.2).

Таблица 7.2

Распределение рабочих по тарифному разряду

Группы рабочих по тарифному разряду x	Численность рабочих, чел. f	Накопленные частоты S
А	1	2
2	20	20
3	50	70
4	60	130
5	70	
6	15	
Итого	215	

По данным табл. 7.2 наибольшую частоту (70 чел.) имеет 5-й тарифный разряд, следовательно, он и является модальным (Мо = 5 разряду), т.е. в данной совокупности рабочих самым распространенным является 5-й тарифный разряд.

Модальный интервал (т.е. содержащий моду) в случае интервального распределения с равными интервалами определяется по наибольшей частоте; с неравными интервалами – по наибольшей плотности, а определение моды требует проведения расчетов на основе следующих формул:

$$\text{Мо} = x_{\text{о}} + i\frac{(f_{\text{Мо}} - f_{\text{Мо}-1})}{(f_{\text{Мо}} - f_{\text{Мо}-1}) + (f_{\text{Мо}} - f_{\text{Мо}+1})}, \quad (7.3)$$

где $x_{\text{о}}$ – нижняя граница модального интервала;
i – величина модального интервала;
$f_{\text{Мо}}$ – частота модального интервала;
$f_{\text{Мо-1}}$ – частота интервала, предшествующего модальному;
$f_{\text{Мо+1}}$ – частота интервала, следующего за модальным;

$$\text{Мо} = x_o + i_{\text{Мо}}\frac{\dfrac{f_{\text{Мо}}}{i_{\text{Мо}}} - \dfrac{f_{\text{Мо}-1}}{i_{\text{Мо}-1}}}{\left(\dfrac{f_{\text{Мо}}}{i_{\text{Мо}}} - \dfrac{f_{\text{Мо}-1}}{i_{\text{Мо}-1}}\right) + \left(\dfrac{f_{\text{Мо}}}{i_{\text{Мо}}} - \dfrac{f_{\text{Мо}+1}}{i_{\text{Мо}+1}}\right)}, \quad (7.4)$$

где $x_{\text{о}}$ – начальная граница модального интервала, в котором достигает максимума величина f/i –отношение частоты интервала к его величине;
$i_{\text{Мо}}, i_{\text{Мо}-1}, i_{\text{Мо}+1}$ – величина соответствующего модального, до- и послемодального интервалов;
$f_{Mo}, f_{\text{Мо}-1}, f_{\text{Мо}+1}$ – частота модального, до- и послемодального интервалов соответственно.

Пример. Используя формулу (7.3), проведем расчет моды для вариационного ряда распределения с равными интервалами по данным табл. 7.3.

Интервал с границами 4–5 в данном распределении будет модальным, так как он имеет наибольшую частоту (графа 1). Определим моду:

$$\text{Мо} = 4 + 1\frac{25-21}{(25-21)+(25-12)} = 4{,}2 \text{ года.}$$

Таким образом, в данной совокупности банков самым распространенным сроком функционирования банка является 4,2 года.

В качестве характеристик вариационного ряда также применяется медиана (Me), т.е. величина изучаемого признака, которая нахо-

Таблица 7.3

Распределение коммерческих банков по сроку функционирования
(на начало года)

Группы банков по сроку функционирования, лет, x	Число банков, % к итогу, f	Накопленная частота, S
А	1	2
1–2	10	10
2–3	15	25
3–4	21	46
4–5	25	71
5–6	12	83
6–7	7	90
7–8	5	95
свыше 8	5	100
Итого	100,0	–

дится в середине упорядоченного вариационного ряда. Главное свойство медианы в том, что *сумма абсолютных отклонений значений признака от медианы меньше, чем от любой другой величины:*

$$\Sigma\left|x_i - \mathrm{Me}\right| = \min. \tag{7.5}$$

Если в вариационном ряду $2m$+1 случаев, то значение признака у случая m+1 будет медианным. Если в ряду четное число $2m$ случаев, то медиана равна средней арифметической из двух данных значений.

Формулы для исчисления медианы при нечетном числе вариантов

$$\mathrm{Me} = x_{m+1}; \tag{7.6}$$

и при четном числе вариантов

$$\mathrm{Me} = \frac{x_m + x_{m+1}}{2}. \tag{7.7}$$

Пример. Рассмотрим определение медианы по данным вариационного ряда из 11 рабочих, имеющих тарифный разряд: 5, 4, 3, 4, 5, 5, 6, 2, 6, 3, 5. Для определения медианы проведем ранжирование рабочих по тарифному разряду: 2 3 3 4 4 5 5 5 5 6 6.

Центральным в этом ряду будет рабочий 5-го разряда, следовательно, данный разряд и будет медианным.

Если ранжированный ряд включает 12 рабочих: 2 3 3 3 4 4 5 5 5 5 6 6, то медиана определяется как средняя арифметическая из двух центральных значений, т.е. в данном ряду медианой будет тарифный разряд, равный

$$\frac{4+5}{2} = 4,5 \text{ разряда.}$$

Если мода отражает типичный, наиболее распространенный вариант значения признака, то медиана практически выполняет функцию средней величины для неоднородной совокупности, не подчиняющейся нормальному закону распределения. Проиллюстрируем ее познавательное значение.

Таблица 7.4

Доходы исследуемой группы людей за месяц

№ п/п	1	2	3	4	…	50	51	…	99	100
Доход, долл.	100	104	104	107	…	162	164	…	200	50 000

Пример. Допустим, нам необходимо дать характеристику среднего дохода группы людей из 100 человек, 99 из которых имеют доход в интервале от 100 до 200 долл. в месяц, а месячный доход последнего человека из группы составляют 50 000 долл. (табл. 7.4).

Если мы воспользуемся формулой средней арифметической, то получим средний доход, равный примерно 600–700 долл., который не только в несколько раз меньше дохода 100-го человека, но и имеет мало общего с доходами остальных членов группы. Медиана же, равная в данном случае 163 долл., позволит дать объективную характеристику уровня дохода 99% данной группы людей.

Рассмотрим определение медианы по сгруппированным данным (рядам распределения).

Положение медианы в ряду распределения определяется ее номером:

$$\text{№}_{\text{Me}} = \frac{n+1}{2}, \tag{7.8}$$

где n – число единиц совокупности.

Пример. Используя данные табл. 7.2, определим номер медианы:

$$\text{№}_{\text{Me}} = \frac{215+1}{2} = 108.$$

Полученное значение указывает, что середина ряда приходится на 108-й номер рабочего. Необходимо определить, к какой группе относится рабочий с этим порядковым номером. Это можно сделать, рассчитав накопленные частоты (табл. 7.2 графа 2). Очевидно, что рабочих с таким номером нет в первой группе, где всего 20 человек, нет их и во второй группе (20 + 50). 108-й номер рабочего находится в третьей группе (20 + 50 + 60 = 130), следовательно, медианным является 4-й тарифный разряд.

В интервальном ряду распределения сразу можно указать только интервал, в котором будет находиться медиана. Для определения ее величины используется специальная формула:

$$\text{Me} = x_{\text{o}} + i\,\frac{\frac{1}{2}\sum f_i - S_{\text{Me}-1}}{f_{\text{Me}}}, \tag{7.9}$$

где x_{o} – нижняя граница медианного интервала;
i – величина медианного интервала;
$S_{\text{Me}-1}$ – накопленная частота интервала, предшествующего медианному;
f_{Me} – частота медианного интервала.

Пример. Используя данные табл. 7.3, рассчитаем медиану. По накопленным частотам (графа 2) определим, что медиана находится в интервале 4–5. Тогда:

$$\text{Me} = 4 + 1\frac{50-46}{25} = 4{,}2 \text{ года}.$$

Таким образом, 50% банков имеет срок функционирования менее 4,2 года, а 50% банков – более 4,2 года.

Моду и медиану в интервальном ряду распределения можно определить *графически*. Мода определяется по гистограмме распределения. Для этого выбирается самый высокий прямоугольник, который в данном случае является модальным. Затем правую вершину модального прямоугольника соединяют с правым верхним углом предыдущего прямоугольника. А левую вершину модального прямоугольника – с левым верхним углом последующего прямоугольника. Далее из точки их пересечения опускают перпендикуляр на ось абсцисс.

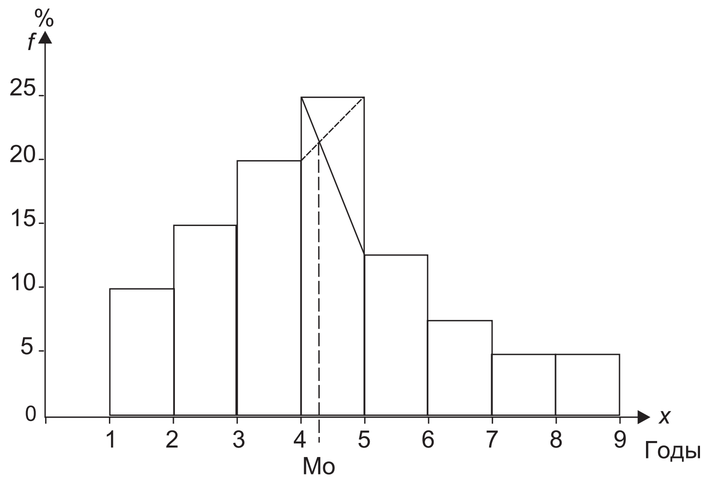

Рис. 7.1. Гистограмма распределения коммерческих банков по сроку функционирования

Абсцисса точки пересечения этих прямых и будет модой распределения (рис. 7.1).

Медиана рассчитывается по кумуляте (рис.7.2). Для ее определения из точки на шкале накопленных частот (частостей), соответствующей 50%, проводится прямая, параллельная оси абсцисс, до пересечения с кумулятой. Затем из точки пересечения указанной прямой с кумулятой опускается перпендикуляр на ось абсцисс. Абсцисса точки пересечения является медианой.

Таким образом, в качестве обобщенной характеристики значений определенного признака у единиц ранжированной совокупности могут быть использованы средняя арифметическая, мода и медиана. Каждая из них имеет свои особенности.

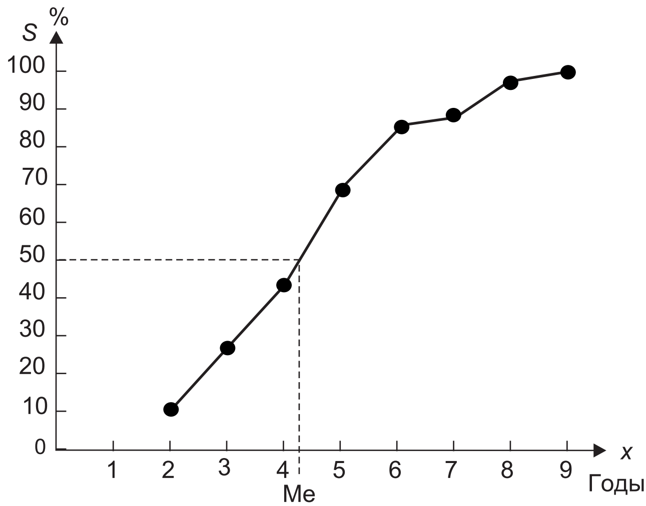

Рис. 7.2. Кумулята распределения коммерческих банков по сроку функционирования

Основной характеристикой центра распределения является средняя арифметическая, для которой характерно то, что все отклонения от нее (положительные или отрицательные) в сумме равняются нулю; для медианы характерно, что сумма отклонений от нее по модулю является минимальной, а мода представляет собой значение признака, которое наиболее часто встречается. Поэтому в зависимости от цели исследования распределения должна выбираться одна из упомянутых характеристик либо же для сравнения вычисляться все три.

Соотношение моды, медианы и средней арифметической указывает на характер распределения признака в совокупности, позволяет оценить его асимметрию.

В *симметричных распределениях* все три характеристики совпадают. Чем больше расхождение между модой и средней арифметической, тем более асимметричен ряд. Для умеренно *асимметричных рядов* разность между модой и средней примерно в три раза превышает разность между медианой и средней, т.е.

$$\left|\mathrm{Mo}-\bar{x}\right| = 3\left|Me-\bar{x}\right|. \tag{7.10}$$

7.3 ПОКАЗАТЕЛИ ВАРИАЦИИ И СПОСОБЫ ИХ РАСЧЕТА

Средняя величина дает обобщающую характеристику всей совокупности изучаемого явления. Однако, исчислив среднюю арифметическую по данным вариационного ряда, мы еще ничего не знаем о том, как отдельные значения изучаемого признака группируются вокруг средней. В этом отношении наблюдаются существенные различия. В одних случаях отдельные значения признака весьма близки к средней арифметической и мало чем от нее отличаются. В этом случае средняя хорошо представляет всю совокупность. В другом случае, наоборот, отдельные значения далеки от средней, и тогда средняя не будет представлять всю совокупность. Возьмем, например, средний уровень доходов населения. Он может быть исчислен как средняя арифметическая из доходов граждан какой-либо страны. Однако значение средней величины для стран, в которых нет резких различий в уровне доходов, будет гораздо выше, чем для стран, в которых наблюдаются резкие различия.

Поэтому нельзя ограничиться вычислением одной средней величины. Надо изучать не только среднюю, но и отклонения от нее, потому что именно в отклонениях виден весь процесс явления в его диалектическом развитии. Отклонение в одну сторону от средней для некоторых показателей следует рассматривать как ростки нового, отклонения в противоположную сторону – как пережитки старого. Для вариационного ряда важно изучать степень сплоченности всех отдельных значений признака вокруг его среднего значения, степень разбросанности этих значений, степень колеблемости их. Для этого в теории статистики используются *показатели вариации*.

Показатели вариации делятся на две группы: абсолютные и относительные. К *абсолютным показателям* относятся: размах вариации, среднее линейное отклонение, дисперсия и среднее квадратическое отклонение. К *относительным показателям* вариации относятся: коэффициенты осцилляции, вариации, относительное линейное отклонение и др. Относительные показатели вычисляются как отношение абсолютных показателей вариации к средней арифметической (или медиане).

Вариационный размах. *Вариационный размах* (R) (или, как еще говорят, амплитуда колебаний) показывает, насколько велико различие между единицами совокупности, имеющими самое маленькое и самое большое значение признака.

Размах рассчитывают как разность между наибольшим (x_{max}) и наименьшим (x_{min}) значениями варьирующего признака, т.е.:

$$R = x_{\max} - x_{\min}. \tag{7.11}$$

Пример. Рассмотрим возраст студентов какого-нибудь вуза: самому молодому студенту – 17 лет, самому старшему – 25 лет. Разность составляет 8 лет.

Значение подобного рода величины необходимо в практической и хозяйственной деятельности, а также в научных исследованиях. Например, размах вариации применяется для контроля качества продукции при определении влияния систематически действующих причин на производственный процесс. Для этого через определенные промежутки времени отбирают несколько деталей и проводят их измерение. Рассчитав по данным этих выборок показатели размаха вариации и сопоставив результаты вычислений, судят об устойчивости режима производственного процесса.

В учебной литературе по статистике обычно указывается, что размах имеет существенный недостаток. Его величина всецело зависит от крайних значений признака, и он не учитывает всех изменений варьирующего признака в пределах совокупности. Этот упрек в адрес размаха является не совсем верным. Какой же это недостаток, когда именно в этом заключается суть показателя.

Размах вариации для того и существует, чтобы измерять расстояние между крайними точками. Другое дело, что в изучении вариации нельзя ограничиться определением одного лишь ее размаха. Но это не исключает необходимости определения величины этого показателя, не умаляет его значения.

К действительным недостаткам размаха вариации можно отнести следующее: очень низкое и очень высокое значения признака по сравнению с основной массой его значений в совокупности могут быть обусловлены какими-либо сугубо случайными обстоятельствами, т.е. эти значения являются *аномальными в совокупности*. В этих случаях размах вариации даст искаженную амплитуду колебания признака против, так сказать, нормальных его размеров, так как в данную совокупность включены единицы другой совокупности с аналогичным признаком. Поэтому прежде чем определить величину размаха вариации, следует очистить совокупность от аномальных наблюдений. Например, нельзя вычислять размах вариации заработков работников какого-либо частного предприятия, если наряду с заработками наемных работников в совокупность включен «заработок» владельца.

Итак, размах вариации – важный показатель колеблемости признака, но он не исчерпывает характеристику вариации.

Среднее линейное отклонение. Для анализа вариации необходим показатель, который бы отражал все колебания варьирующего признака и давал обобщенную его характеристику. Для многих варьирующих признаков возможно допущение, что при прочих равных условиях все единицы совокупности в соответствии с основными законами своего развития имеют одинаковую и при том вполне определенную величину в данных условиях места и времени. Вполне логично в качестве такой величины условно принять *среднюю величину* из всех значений признака, поскольку в ней более или менее погашаются случайные отклонения от закономерного развития явления, и средняя тем самым отражает типичный размер признака у данной однородной совокупности единиц. Но условия существования и развития отдельных единиц совокупности в определенной степени различны, что сказывается на различии значений признака. Средняя величина отражает эти средние условия.

Следовательно, средняя применяется в качестве своего рода центра тяжести, вокруг которого происходит колебание, рассеяние значений признака. При обобщении этих колебаний необходимо прибегать к методу средних величин – искать среднюю величину этих отклонений.

Такая средняя называется *средним линейным отклонением* ($\bar{d}$). Эта величина вычисляется как средняя арифметическая из абсолютных значений отклонений вариант x_i и $\bar{x}$ (простая (формула 7.12) или взвешенная (формула 7.13), в зависимости от исходных условий):

$$\bar{d} = \frac{\sum_{i=1}^{n} |x_i - \bar{x}|}{n}; \qquad (7.12)$$

$$\bar{d} = \frac{\sum_{i=1}^{k} |x_i - \bar{x}| f_i}{\sum_{i=1}^{k} f_i}. \qquad (7.13)$$

Поскольку сумма отклонений значений признака от средней величины равна нулю, приходится все отклонения брать по модулю, на что указывают прямые скобки в числителе формул.

Пример. Покажем расчет среднего линейного отклонения по данным табл. 7.5.

Таблица 7.5

Обеспеченность населения города общей жилой площадью

Группы населения по размеру общей жилой площади на одного члена семьи, кв. м x_i	Число семей, % к итогу, f_i	Середина интервала x_i'	$x_i' f_i$	$\lvert x_i' - \bar{x} \rvert$	$\lvert x_i' - \bar{x} \rvert f_i$
А	1	2	3	4	5
До 10	30	9	270	3,06	91,8
10–12	25	11	275	1,06	26,5
12–14	26	13	338	0,94	24,4
14–16	9	15	135	2,94	26,5
16–18	4	17	68	4,94	19,8
18–20	3	19	57	6,94	20,8
Свыше 20	3	21	63	8,94	26,8
Итого	100,0	–	1206	–	236,6

Алгоритм расчета среднего линейного отклонения следующий:

1. Найдем середину интервалов (x_i') по исходным данным (графа А) и запишем в таблицу (графа 2).

2. Определим произведения значений середины интервалов (x_i') на соответствующие им веса (f_i) (графа 3). В итоге получим 1 206. Рассчитаем среднюю величину по формуле средней арифметической взвешенной:

$$\bar{x} = \frac{\sum x_i \cdot f_i}{\sum f_i} = \frac{1\,206}{100} = 12{,}06 \text{ кв. м.}$$

3. Для расчета линейного отклонения найдем абсолютные отклонения середины интервалов, принятых нами в качестве вариантов признака (x'_i) от средней величины ($\bar{x}$) (графа 4).

4. Наконец, вычислим произведения отклонений $|x'_i - \bar{x}|$ на их веса (f_i) и подсчитаем сумму их произведений. Она равна 236,6. Результаты записываем в графу 5.

Делим эту сумму на сумму весов, чтобы получить искомую величину $\bar{d}$:

$$\bar{d} = \frac{236,6}{100} = 2,366 \text{ кв.м.}$$

Таково в среднем отклонение вариантов признака от их средней величины. Это отклонение по сравнению со средней величиной признака небольшое. Оно отличается от средней на 9,694 кв. м. Это свидетельствует о том, что данная совокупность в отношении нашего признака однородна, а средняя – типична.

Таким образом, среднее линейное отклонение дает обобщенную характеристику степени колеблемости признака в совокупности. Однако при его исчислении приходится допускать некорректные с точки зрения математики действия, нарушать законы алгебры. Математики и статистики искали иной способ оценки вариации для того, чтобы иметь дело только с положительными величинами. Был найден очень простой выход – возвести все отклонения во вторую степень. Это столь простое решение привело в последующем к большим научным результатам. Оказалось, что обобщающие показатели вариации, найденные с использованием вторых степеней отклонений, обладают замечательными свойствами; позднее на их основе были разработаны новые методы исследования, а также новые показатели количественной характеристики большого класса явлений. Полученную меру вариации назвали *дисперсией* и обозначили D или σ^2.

Дисперсия. Дисперсия представляет собой средний квадрат отклонений индивидуальных значений признака от их средней величины и в зависимости от исходных данных вычисляется по формулам простой дисперсии (формула 7.14) и взвешенной дисперсии (формула 7.15):

$$\sigma^2 = \frac{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2}{n}; \qquad (7.14)$$

$$\sigma^2 = \frac{\sum_{i=1}^{k}\left(x_i - \bar{x}\right)^2 f_i}{\sum_{i=1}^{k} f_i}. \qquad (7.15)$$

Расчет дисперсии может быть упрощен. В случае равных интервалов в вариационном ряду распределения используется способ отсчета от условного нуля (способ моментов). Для его понимания необходимо знать *математические свойства дисперсии:*

1. Дисперсия постоянной величины равна нулю.

2. Уменьшение всех значений признака на одну и ту же величину A не меняет величины дисперсии:

$$\sigma^2_{(x-A)} = \sigma^2_x. \qquad (7.16)$$

Значит, средний квадрат отклонений можно вычислить не по заданным значениям признака, а по их отклонениям от какого-то постоянного числа.

3. Уменьшение всех значений признака в k раз уменьшает дисперсию в k^2 раз, а среднее квадратическое отклонение – в k раз:

$$\sigma^2_{(x/A)} = \sigma^2_x : k^2. \qquad (7.17)$$

Значит, все значения признака можно разделить на какое-то постоянное число (скажем, на величину интервала ряда), исчислить среднее квадратическое отклонение, а затем умножить его на постоянное число:

$$\sigma^2_x > \sigma^2_{x/k} k. \qquad (7.18)$$

4. Если исчислить средний квадрат отклонений от любой величины A, в той или иной степени отличающейся от средней арифметической ($\bar{x}$), то он всегда будет больше среднего квадрата отклонений, исчисленного от средней арифметической:

$$\sigma^2_A > \sigma^2_x. \qquad (7.19)$$

Средний квадрат отклонений при этом будет больше на вполне определенную величину – на квадрат разности средней и этой условно взятой величины, т.е. на $(\bar{x} - A)^2$:

$$\sigma_A^2 = \sigma_x^2 + (\bar{x} - A)^2,$$

или

$$\sigma^2 = \frac{\sum_{i=1}^{k} (x_i - A)^2 f_i}{\sum_{i=1}^{k} f_i} - (\bar{x} - A)^2. \qquad (7.20)$$

Значит, дисперсия от средней всегда меньше дисперсий, исчисленных от любых других величин, т.е. она имеет *свойство минимальности.*

В случае когда A приравнивается нулю и, следовательно, отклонения не вычисляются, формула принимает такой вид:

$$у_x^2 = \bar{x}^2 - (\bar{x})^2,$$

или

$$у^2 = \frac{\sum_{i=1}^{k} x^2 f_i}{\sum_{i=1}^{k} f_i} - \left(\frac{\sum_{i=1}^{k} x_i f_i}{\sum_{i=1}^{k} f_i} \right)^2. \qquad (7.21)$$

Значит, средний квадрат отклонений равен среднему квадрату значений признака минус квадрат среднего значения признака.

На приведенных математических свойствах дисперсии основан метод расчета дисперсии *по способу моментов,* или *способу отсчета от условного нуля*, который применялся при исчислении средней величины. Расчет производится по формуле

$$y^2 = \frac{\sum_{i=1}^{k} \left(\frac{x_i - A}{k} \right)^2 f_i}{\sum_{i=1}^{k} f_i} k^2 - \left(x - A \right)^2, \qquad (7.22)$$

где k – ширина интервала;

A – условный нуль, в качестве которого удобно использовать середину интервала, обладающего наибольшей частотой;

$\frac{\sum_{i=1}^{k} \left(\frac{x_i - A}{k} \right)^2}{\sum_{i=1}^{k} f_i}$ – момент второго порядка.

Дисперсия есть средняя величина квадратов отклонений, а варианты признака выражены в первой степени.

Среднее квадратическое отклонение (σ). Среднее квадратическое отклонение равно корню квадратному из дисперсии. Оно может быть простым (формула 7.23) или взвешенным (формула 7.24).

$$y = \sqrt{\frac{\sum_{i=1}^{n} \left(x_i - \bar{x} \right)^2}{n}} \qquad (7.23)$$

или

$$y = \sqrt{\frac{\sum_{i=1}^{k} (x_i - x)^2 f_i}{\sum_{i=1}^{k} f_i}}. \qquad (7.24)$$

Среднее квадратическое отклонение, как и среднее линейное отклонение, показывает, на сколько в среднем отклоняются конкретные варианты признака от среднего значения. Они выражаются в тех же единицах измерения, что и признак (в метрах, тоннах, рублях и т.д.).

Среднее квадратическое отклонение часто используется в качестве единицы измерения отклонений от средней арифметической. В зарубежной литературе этот показатель называется *нормированным,* или *стандартизованным, отклонением*.

По свойству мажорантности средних величин (см. глава 6) 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: $\bar{d} = 0{,}8\sigma$ или $\sigma = 1{,}25\,\bar{d}$.

Среднее квадратическое отклонение играет важную роль в анализе вариационных рядов распределения. В условиях нормального распределения существует следующая взаимосвязь между величиной среднего квадратического отклонения и количеством наблюдений:

- в пределах $\bar{x} \pm 1\sigma$ располагается 0,683, или 68,3% количества наблюдений;
- в пределах $x \pm 2\sigma$ – 0,954, или 95,4%;
- в пределах $x \pm 3\sigma$ – 0,997, или 99,7% количества наблюдений.

В действительности на практике почти не встречаются отклонения, которые превышают ±3σ. Отклонение 3σ может считаться максимально возможным. Это положение называют *правилом трех сигм*.

Пример. Рассмотрим расчет дисперсии и среднего квадратического отклонения по данным табл. 7.6 о выпуске промышленной продукции фирмами отрасли.

Таблица 7.6

Вычисление σ^2 и σ по несгруппированным данным

Номер фирмы	Выпущено промышленной продукции за год, млн.руб. x	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
А	1	2	3
1	60	+10	100
2	52	+2	4
3	40	–10	100
4	60	+10	100
5	50	0	0
6	38	–12	144
Итого	300	–	448

Алгоритм расчета следующий:

1. Определим среднюю величину по исходным данным (графа 1) по формуле средней арифметической простой:

$$\bar{x} = \frac{\sum x_i}{n} = \frac{300}{6} = 50 \text{ млн руб.}$$

2. Найдем отклонения $(x_i - \bar{x})$ и запишем их в графе 2. Возведем отклонения во вторую степень и запишем в графе 3. Определим их сумму. Она равна 448.

3. Разделив эту сумму на число единиц совокупности, получим дисперсию:

$$\text{у}^2 = \frac{448}{6} = 74,67.$$

4. Извлечем из дисперсии корень второй степени $\sqrt{74,67} = 8,64$ млн руб. и получим среднее квадратическое отклонение.

Степень вариации в данной совокупности невелика, так как средняя величина равна 50 млн руб. Это говорит об однородности рассматриваемой нами совокупности.

Пример. Рассмотрим вычисление дисперсии и среднего квадратического отклонения по данным распределения сотрудников двух фирм по тарифному разряду (табл. 7.7).

Фирма А

$$\bar{x}_1 = 15; \sigma^2 = \frac{118}{132} = 0,89; \ \sigma_1 = \sqrt{0,89} = 0,94 \text{ разряда.}$$

Фирма Б

$$\bar{x}_1 = 15; \ \sigma^2 = \frac{720}{170} = 4,24; \sigma_2 = \sqrt{4,24} = 2,05 \text{ разряда.}$$

Среднее квадратическое отклонение в фирме *Б* более чем в два раза превышает среднее квадратическое отклонение фирмы *А*. До сих пор говорилось о показателях вариации, выраженных в абсолютных

Таблица 7.7

Расчет σ^2 и σ в вариационном ряду с разным распределением частот

Тариф, разряд x_i	Число сотрудников, чел. f_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$	$(x_i - \bar{x})^2 f_i$	Тариф, разряд x_i	Число сотрудников, чел. f_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$	$(x_i - \bar{x})^2 f_i$
12	1	–3	9	9	12	30	–3	9	270
13	5	–2	4	20	13	20	–2	4	80
14	30	–1	1	30	14	10	–1	1	10
15	60	0	0	0	15	50	0	0	0
16	30	1	1	30	16	10	1	1	10
17	5	2	4	20	17	20	2	4	80
18	1	3	9	9	18	30	3	9	270
Итого	132	–	–	118	–	170	–	–	720

величинах. Но для целей сравнения колеблемости различных признаков в одной и той же совокупности или же при сравнении колеблемости одного и того же признака в нескольких совокупностях представляют интерес показатели вариации, приведенные в *относительных величинах*. Базой для сравнения должна служить средняя арифметическая. Эти показатели вычисляются как отношение размаха вариации, среднего линейного отклонения или среднего квадратического отклонения к средней арифметической или медиане. Чаще всего они выражаются в процентах и определяют не только сравнительную оценку вариации, но и дают характеристику однородности совокупности. Совокупность считается однородной, если коэффициент вариации не превышает 33% (для распределений, близких к нормальному). Различают следующие относительные показатели вариации (V).

Коэффициент осцилляции (V_R):

$$V_R = \frac{R}{\bar{x}} 100\%. \tag{7.25}$$

Линейный коэффициент вариации ($V_{\bar{d}}$):

$$V_{\bar{d}} = \frac{\bar{d}}{\bar{x}} 100\%, \tag{7.26}$$

или

$$V_{\bar{d}} = \frac{\bar{d}}{\text{Me}} 100\%.$$

Наиболее часто в практических расчетах применяется показатель относительной вариации – коэффициент вариации.

Коэффициент вариации (V_σ):

$$V_{\text{у}} = \frac{\text{у}}{\bar{x}} 100\%. \tag{7.27}$$

Пример. По данным, приведенным в табл. 7.7, рассчитаем коэффициенты вариации:

$$V_1 = \frac{0,94}{15} 100 = 6,27\%;$$

$$V_2 = \frac{2,05}{15} 100 = 13,67\%.$$

На основе полученных коэффициентов можно сделать вывод, что по тарифному разряду совокупности рабочих фирмы *А* и фирмы *Б* являются однородными. Однако вариация колеблемости тарифного разряда в фирме *Б* несколько выше, чем вариация в фирме *А*.

Характеристика степени вариации ряда может быть определена также по формуле квартильного отклонения (Q), предложенной английским биологом и антропологом Ф. Гальтоном:

$$Q = \frac{Q_3 - Q_1}{2}, \tag{7.28}$$

где Q_3 и Q_1 – соответственно 1-я и 3-я квартили распределения (см. раздел 7.6).

Эта формула дает *абсолютный квартильный показатель вариации*. 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 затруднено или невозможно. В частности, этот показатель может быть рекомендован для рядов распределения с открытыми интервалами.

В целях сравнения вариации в различных рядах вычисляется *относительный квартильный показатель вариации* по формуле

$$K_Q = \frac{Q}{\text{Me}} 100\%, \tag{7.29}$$

или

$$K_Q = \frac{Q_3 - Q_1}{2Q_2} 100\%,$$

где Ме – медиана ряда распределения.

7.4 ВАРИАЦИИ АЛЬТЕРНАТИВНОГО ПРИЗНАКА. ЭНТРОПИЯ РАСПРЕДЕЛЕНИЯ

В ряде случаев возникает необходимость в измерении дисперсии так называемых *альтернативных признаков*, тех, которыми обладают одни единицы совокупности и не обладают другие. Примером таких признаков являются: бракованная продукция, ученая степень преподавателя вуза, работа по полученной специальности и т.д. Вариация альтернативного признака количественно проявляется в значении нуля у единицы, которая этим признаком не обладает, или единицы у той, которая данный признак имеет.

Пусть p – доля единиц в совокупности, обладающих данным признаком ($p = m/n$); q – доля единиц, не обладающих данным признаком, причем $p + q = 1$. Альтернативный признак принимает всего два значения – 0 и 1 с весами соответственно q и p. Исчислим среднее значение альтернативного признака по формуле средней арифметической:

$$\bar{x} = \frac{1p + 0p}{p + q} = p. \qquad (7.30)$$

Дисперсия альтернативного признака определяется по формуле:

$$у^2 = \frac{(1-p)^2 p + (0-p)^2 q}{p+q} = \frac{q^2 p + p^2 q}{p+q} = pq. \qquad (7.31)$$

Таким образом, дисперсия альтернативного признака равна произведению доли на число, дополняющее эту долю до единицы. Корень квадратный из этого показателя, т.е. $\sqrt{pq}$, соответствует среднему квадратическому отклонению альтернативного признака. Предельное значение дисперсии альтернативного признака равно 0,25 при $p = 0{,}5$.

Показатели вариации альтернативных признаков широко используются в статистике, в частности, при проектировании выборочного наблюдения, обработке данных социологических обследований, статистическом контроле качества продукции, в ряде других случаев.

Пример. В трех партиях готовой продукции, представленной на контроль качества, была обнаружена годная и бракованная продукция (табл. 7.8).

Таблица 7.8

Продукция, представленная на контроль качества

Партия	Готовая продукция, шт.	Из них продукция	
		годная	бракованная
1	1200	800	400
2	1000	840	160
3	1100	1000	100

Определим в целом для всех партий следующие показатели:

1) средний процент годной продукции и средний процент брака;

2) дисперсию, среднее квадратическое отклонение, коэффициент вариации годной продукции.

Произведем расчет данных показателей на нашем примере.

Средний процент годной продукции в трех партиях равен:

$$p = \frac{800 + 840 + 1000}{1200 + +1000 + 1100} = \frac{2640}{3300} = 0{,}8, \text{ или } 80\%.$$

Средний процент бракованной продукции составил:

$$q = 1–0,8 = 0,2 \text{ или } 20\%.$$

Дисперсия удельного веса годной продукции:

$$у^2 = pq = 0,8 \cdot 0,2 = 0,16.$$

Среднее квадратическое отклонение удельного веса годной продукции:

$$у = \sqrt{pq} = \sqrt{0,16} = 0,4.$$

Коэффициент вариации удельного веса годной продукции в общем выпуске продукции:

$$V\frac{у}{\bar{x}} = \frac{у}{p} = \frac{0,4}{0,8} \cdot 100\% = 50\%.$$

Обобщенной характеристикой различий внутри ряда может служить *энтропия распределения.* Применительно к статистике энтропия – это мера неопределенности данных наблюдения, которая может иметь различные результаты. Энтропия зависит от числа градаций признака и от вероятности каждой из них. Энтропия показывает, имеется ли закономерность в концентрации отдельных градаций у наименьшего числа позиций или, напротив, заполненность распределения одинаковая. При этом сумма вероятностей всех возможных исходов равна единице. Энтропия измеряется в битах.

Показатель энтропии H_x представляет собой отрицательную сумму произведения вероятностей различных значений случайной величины (p_i) на логарифмы (при основании два) этих вероятностей:

$$H_x = -\Sigma \log p_i. \tag{7.32}$$

Если все варианты равновероятны, то энтропия максимальна. Если же все варианты, за исключением одного, равны нулю, то энтропия равна нулю.

Энтропия альтернативного признака (n = 2) при равновероятном распределении (p = 0,5) равна единице:

$$H_x = -(0{,}5 \log 0{,}5 + 0{,}5 \log 0{,}5) = 1. \qquad (7.33)$$

Пример. Расчет энтропии распределения можно показать на выпуске продукции различных сортов на одном из предприятий точного машиностроения (табл. 7.9).

Таблица 7.9

Вероятность выпуска различных сортов продукции

Сорт	1-й	2-й	3-й	Брак	Итого
Вероятность, p_i	0,90	0,04	0,05	0,01	1,00

Энтропия данного распределения равна:

$$H_x = -(0{,}9 \log 0{,}9 + 0{,}04 \log 0{,}04 + 0{,}05 \log 0{,}05 + 0{,}01 \log 0{,}01) =$$
$$= 0{,}6051 \text{ бита.}$$

Энтропия сложной системы вычисляется следующим образом:

$$H_{xy} = \sum_{i=1}^{n} \sum_{j=1}^{m} p_{ij} \log_2 p_{ij}, \qquad (7.34)$$

где p_{ij} – вероятность любого возможного состояния сложной системы.

Показатель энтропии позволяет также измерять количество информации. Чем больше информации о случайном событии, тем определеннее его состояние. Чем больше вероятность случайного события p_a, тем меньше информации несет его осуществление. В случае p_a = 1

$$\mathrm{H}_x = -(1 \log 1 + 0 + \ldots + 0) = 0. \qquad (7.35)$$

Следовательно, данное испытание не содержит никакой информации. Аналогично и при $p_a = 0$.

Энтропия распределения интерпретируется как мера рассредоточенности вариантов случайной переменной по ее возможным значениям, или как мера неопределенности значения реализации. Неопределенность значений реализации случайной переменной предусматривает наличие некоторого наблюдателя, находящегося в том или ином отношении к источнику неопределенности. Очевидно, можно представить ситуацию, когда для двух наблюдений степени неопределенности результата одного и того же наблюдения со случайными исходами существенно различаются. Например, различны результаты голосования при экспертных опросах для наблюдателя – участника голосования и наблюдателя, не участвующего в голосовании.

В связи с тем что верхнего предела энтропия распределения не имеет, целесообразно вычислить наряду с абсолютной и относительную величину неопределенности.

Относительная энтропия определяется как отношение ее фактической величины к максимальной, т.е.

$$H_x : H_{x\,\max}. \qquad (7.36)$$

Это отношение изменяется от 0 до 1 и может быть интерпретировано. Чем меньше относительная энтропия, тем меньше неопределенность и выше однородность.

7.5 ВИДЫ ДИСПЕРСИЙ В СОВОКУПНОСТИ, РАЗДЕЛЕННОЙ НА ГРУППЫ. ПРАВИЛО СЛОЖЕНИЯ ДИСПЕРСИЙ

Изучая вариацию по всей совокупности в целом и опираясь на общую среднюю в своих расчетах, мы не можем определить влияние отдельных факторов, характеризующих колеблемость индивидуальных значений признака. Это можно сделать при помощи аналитической группировки, разделив изучаемую совокупность на однородные группы по признаку-фактору. При этом можно определить три показателя колеблемости признака в совокупности: дисперсию общую, межгрупповую и среднюю из внутригрупповых дисперсий.

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

$$у^2 = \frac{\sum_{i=1}^{k}(x_i - \bar{x}_o)^2 n_i}{\sum_{i=1}^{k} n_i}. \qquad (7.37)$$

Межгрупповая дисперсия (δ_x^2) характеризует систематическую вариацию, т.е. различия в величине изучаемого признака, возникающие под влиянием признака-фактора, положенного в основание группировки. Она рассчитывается по формуле

$$\delta_x^2 = \frac{\sum_{j=1}^{k}\left(\bar{x}_j - \bar{x}_0\right)^2 n_j}{\sum_{j=1}^{k} n_j}, \qquad (7.38)$$

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.

Внутригрупповая дисперсия (σ_j^2) отражает случайную вариацию, т.е. часть вариации, происходящую под влиянием неучтенных факторов и не зависящую от признака-фактора, положенного в основание группировки. Она исчисляется следующим образом:

$$у_j{}^2 = \frac{\sum_{i=1}^{n_j}\left(x_{ij} - \bar{x}_j\right)^2}{n_j}. \qquad (7.39)$$

По совокупности в целом вариация значений признака под влиянием прочих факторов характеризуется *средней из внутригрупповых дисперсий* ($\bar{\sigma}^2$):

$$\bar{у}^2 = \frac{\sum\limits_{j=1}^{k} \sigma_j^2 \ n_j}{\sum\limits_{j=1}^{k} n_j}. \qquad (7.40)$$

Между общей дисперсией $\sigma^2{}_0$, средней из внутригрупповых дисперсий $\bar{\sigma}^2$ и межгрупповой $\bar{\delta}^2$ дисперсией существует соотношение, определяемое *правилом сложения дисперсий.* Согласно этому правилу *общая дисперсия равна сумме средней из внутригрупповых и межгрупповой дисперсий:*

$$у_0^2 = \bar{у}^{\,2} + д^2. \qquad (7.41)$$

Согласно этому правилу, общая дисперсия, возникающая под действием всех факторов, равна сумме дисперсий, появляющихся под влиянием всех прочих факторов, и дисперсии, возникающей за счет группировочного признака.

Зная любые два вида дисперсий, можно определить или проверить правильность расчета третьего вида.

Правило сложения дисперсий позволяет выявить зависимость результата от определяющих факторов с помощью соотношения межгрупповой дисперсии и общей дисперсии. Это соотношение называется *эмпирическим коэффициентом детерминации* (η^2):

$$з^2 = \frac{д^2}{у_0^2}. \qquad (7.42)$$

Он показывает, какая доля в общей дисперсии приходится на дисперсию, обусловленную вариацией признака, положенного в основу группировки.

Корень квадратный из эмпирического коэффициента детерминации носит название *эмпирического корреляционного отношения* (η):

$$з = \sqrt{\frac{д^2}{у_0^2}}. \qquad (7.43)$$

Это отношение характеризует влияние признака, положенного в основание группировки, на вариацию результативного признака. Эмпирическое корреляционное отношение изменяется в пределах от 0 до 1. Если $\eta = 0$, то группировочный признак не оказывает влияния на результативный. Если $\eta = 1$, то результативный признак изменяется только в зависимости от признака, положенного в основание группировки, а влияние прочих факторных признаков равно нулю. Промежуточные значения оцениваются в зависимости от их близости к предельным значениям.

Пример. Рассмотрим правило сложения дисперсий. Имеются данные об объеме выполненных работ проектно-изыскательными организациями на предприятиях разных форм собственности (табл. 7.10).

Таблица 7.10

Организация	Объем выполненных работ на предприятиях, млн руб.	
	государственных	коммерческих
1	420	3 980
2	690	6 120
3	790	6 030
4	950	7 790
5	580	5 050
Итого	3 430	28 970

Алгоритм решения следующий:

1. Определим средний объем выполненных работ на предприятиях двух форм собственности:

$$\bar{x}_0 = \frac{\sum x_i}{\sum n_i} = \frac{3\,430 + 28\,970}{5+5} = \frac{32\,400}{10} = 3\,240 \text{ млн руб.}$$

2. Определим средние объемы выполненных работ по предприятиям каждой формы собственности:

$$\overline{x}_1 = \frac{\sum x_i}{n_1} = \frac{3\,430}{5} = 686 \text{ млн руб.;}$$

$$\overline{x}_2 = \frac{\sum x_i}{n_2} = \frac{28\,970}{5} = 5\,794 \text{ млн руб.}$$

3. Рассчитаем внутригрупповые и общую дисперсии:

$$\text{у}_1^2 = \frac{\sum \left(x_i - \overline{x}_1\right)^2}{n_1} = \frac{\left(420-686\right)^2 + \left(690-686\right)^2 + \ldots + \left(580-686\right)^2}{5} =$$

$$= \frac{162\,520}{5} = 32\,504;$$

$$\text{у}_1^2 = \frac{\sum \left(x_i - \overline{x}_1\right)^2}{n_1} = \frac{\left(420-686\right)^2 + \left(690-686\right)^2 + \ldots + \left(580-686\right)^2}{5} =$$

$$= \frac{162\,520}{5} = 32\,504;$$

$$\text{у}_0^2 = \frac{\sum \left(x_i - \overline{x}_0\right)^2}{n_0} =$$

$$= \frac{\left(420-3\,240\right)^2 + \left(690-3\,240\right)^2 + \left(790-3\,240\right)^2 + (950-3\,240)^2 +}{10}$$

$$\frac{+(950-3\,240)^2 + \ldots + (5\,050-3\,240)^2}{10} = \frac{73\,381\,800}{10} = 7\,338\,180.$$

4. Рассчитаем среднюю из внутригрупповых и межгрупповую дисперсии по данным, представленным в табл. 7.11.

Таблица 7.11

Расчет $\overline{\sigma}^2$ и δ^2 по предприятиям двух форм собственности

Группы предприятий	Численность предприятий n_j	Средний объем выполненных работ, млн руб. $\overline{x}_j$	Дисперсия объема выполненных работ σ^2_j
Государственные предприятия	5	686	32 504
Коммерческие предприятия	5	5 794	1 598 040

Средняя из внутригрупповых дисперсий:

$$\overline{у^2} = \frac{1\,598\,024 \cdot 5 + 32\,504 \cdot 5}{10} = 815\,264;$$

межгрупповая дисперсия:

$$д^2 = \frac{(686 - 3\,240)^2 \cdot 5 + (5\,794 - 3\,240)^2 \cdot 5}{10} = 6\,522\,916.$$

5. Найдем общую дисперсию по правилу сложения дисперсий:

$$у_0^2 = 815\,264 + 6\,522\,916 = 7\,338\,180.$$

Сопоставив межгрупповую дисперсию с общей дисперсией, рассчитаем коэффициент детерминации:

$$з^2 = \frac{6\,522\,916}{7\,338\,180} = 0{,}889, \text{ или } 88{,}9\%.$$

Коэффициент детерминации показывает, что дисперсия объема выполненных работ зависит от формы собственности предприятия на 88,9%. Остальные 11,1% определяются множеством других неучтенных факторов.

Извлекая квадратный корень из коэффициента детерминации, определим эмпирическое корреляционное отношение:

$$з = \sqrt{0,889} = 0,94 .$$

Полученное значение эмпирического корреляционного отношения позволяет утверждать, что существует тесная связь между формой собственности предприятия и объемом выполненных проектно-изыскательных работ.

Для проверки существенности связи между группировочным признаком и вариацией исследуемого признака часто используется дисперсионное отношение F (критерий Фишера).

$$F = \frac{д^2}{\nu_1} : \frac{\overline{у^2}}{\nu_2} , \qquad (7.44)$$

где ν_1 и ν_2 – число степеней свободы для сравниваемых дисперсий, при этом $\nu_1 = m - 1$; $\nu_2 = N - m$;
m – число групп,
N – число наблюдений.

Расчетное значения критерия Фишера ($F_{расч}$) сравнивается с критическим ($F_{кр}$), которое определяется по таблице приложения 4 в зависимости от числа степеней свободы и уровня значимости. Если $F_{расч} > F_{кр}$, наличие связи доказано, так как проверяется нулевая гипотеза об отсутствии взаимосвязи признаков, т.е. об отсутствии влияния группировочного признака на исследуемый признак. В нашем примере $F_{расч}$=24, а $F_{кр}$=10,13 при уровне значимости 0,05, т.е. это говорит о наличии связи между объемом выполненных проектно-изыскательных работ и формой собственности предприятий.

Правило сложения дисперсий для доли признака. Рассмотренное правило сложения дисперсий распространяется и на дисперсии доли признака, т.е. доли единиц с определенным признаком в совокупности, разбитой на группы. При этом изучение вариации происходит непосредственно при вычислении и анализе видов дисперсий для доли признака.

Внутригрупповая дисперсия доли определяется по формуле

$$\text{у}^2_{p_i} = p_i \cdot (1 - p_i)\,, \tag{7.45}$$

где p_i – доля изучаемого признака в отдельных группах.

Средняя из внутригрупповых дисперсий имеет следующий вид:

$$\overline{\sigma^2_{p_i}} = \frac{\sum_i p_i (1 - p_i)^2 n_i}{\sum_i n_i} = p_i \cdot (1 - p_i)\,. \tag{7.46}$$

Формула *межгрупповой дисперсии* имеет следующий вид:

$$\text{д}^2_{p_i} = \frac{\sum_i (p_i - \overline{p})^2 n_i}{\sum_i n_i}\,, \tag{7.47}$$

где n_i – численность единиц в отдельных группах;
$\overline{p}$ – доля изучаемого признака во всей совокупности.

Доля признака в совокупности определяется по формуле средней арифметической взвешенной:

$$\overline{p} = \frac{\sum_i p_i n_i}{\sum_\text{i} n_i}\,. \tag{7.48}$$

Общая дисперсия определяется по формуле

$$\text{у}^2_{\overline{p}} = \overline{p} \cdot (1 - \overline{p})\,. \tag{7.49}$$

Три вида рассмотренных дисперсий связаны между собой следующим образом:

$$\text{у}^2_{p} = \overline{\text{у}^2_{p_i}} + \text{д}^2_{p_i}\,. \tag{7.50}$$

Это соотношение дисперсий называется *правилом сложения дисперсий доли признака.*

Зная любые два вида дисперсий из трех, входящих в формулу (7.50), можно определить дисперсию третьего вида или проверить правильность ее расчета.

Пример. Имеются следующие данные удельного веса основных рабочих в трех цехах фирмы (табл. 7.12).

Таблица 7.12

Удельный вес основных рабочих фирмы

Цех	Σдельный вес основных рабочих, %, p_i	Численность всех рабочих, чел., n_i
1	80	100
2	75	200
3	90	150
Итого	–	450

1. Определим долю основных рабочих в целом по фирме (формула 7.48):

$$\bar{p} = \frac{0,80 \cdot 100 + 0,75 \cdot 200 + 0,90 \cdot 150}{450} = \frac{365}{450} = 0,81 \cdot$$

2. Общая дисперсия доли основных рабочих по всей фирме в целом равна (формула 7.49):

$$\text{у}^2_{p} = 0,81 \cdot (1 - 0,81) = 0,154.$$

3. Внутрицеховые дисперсии рассчитаем, применив формулу (7.45):

$$у_{p_1}^2 = 0,8 \cdot 0,2 = 0,16;$$
$$у_{p_2}^2 = 0,75 \cdot 0,25 = 0,19;$$
$$у_{p_3}^2 = 0,9 \cdot 0,1 = 0,09.$$

4. Средняя из внутригрупповых дисперсий будет равна (формула 7.46):

$$\overline{\sigma_{p_i}^2} = \frac{0,16 \cdot 100 + 0,19 \cdot 200 + 0,09 \cdot 150}{450} = \frac{365}{450} = 0,15.$$

5. Межгрупповую дисперсию определим по формуле (7.47):

$$д_{p_i}^2 = \frac{(0,8-0,81)^2 \cdot 100 + (0,75-0,81)^2 \cdot 200 + (0,9-0,81)^2 \cdot 150}{450} =$$
$$= \frac{65}{450} = 0,004 .$$

Проверка вычислений показывает: 0,154 = 0,15 + 0,004.

7.6 СТРУКТУРНЫЕ ХАРАКТЕРИСТИКИ ВАРИАЦИОННОГО РЯДА РАСПРЕДЕЛЕНИЯ. ПОКАЗАТЕЛИ ДИФФЕРЕНЦИАЦИИ

Рассмотренные обобщающие показатели центра распределения и степени вариации не дают понятия о форме распределения, т.е. не вскрывают *характера* последовательного изменения частот. Для выражения особенностей формы распределения применяются показатели дифференциации, основанные на структурных (ранговых[1]) показателях распределения.

[1] См.: *Виноградова Н.М.* и др. Общая теория статистики. – М.: Статистика, 1968. – С. 177.

Структурные показатели. В системе структурных показателей в качестве показателей особенностей формы распределения выступают варианты, занимающие определенное место (каждое четвертое, пятое, десятое, двадцать пятое и т.д.) в ранжированном вариационном ряду. Такие показатели носят общее название *квантилей*, или *градиентов*.

Некоторые квантили имеют особые наименования: квартили, квинтили, децили и перцентили.

Квартили представляют собой значение признака, делящее ранжированную совокупность на четыре равновеликие части. Различают квартиль нижний (Q_1), отделяющий $^1/_4$ часть совокупности с наименьшими значениями признака, и квартиль верхний (Q_3), отсекающий $^1/_4$ часть с наибольшими значениями признака. Это означает, 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% превзойдут Q_3. Вторая квартиль Q_2 является медианой. Вычисление квартилей аналогично вычислению медианы (см. раздел 7.2 этой главы).

Для расчета квартилей по интервальному вариационному ряду используются формулы

$$Q_1 = x_{Q_1} + i \cdot \frac{\frac{1}{4}\sum f - S_{Q_{1-1}}}{f_{Q_1}}; \tag{7.51}$$

$$Q_3 = x_{Q_3} + i \cdot \frac{\frac{3}{4}\sum f - S_{Q_{3-1}}}{f_{Q_3}}, \tag{7.52}$$

где x_{Q_1} – нижняя граница интервала, содержащего нижний квартиль (интервал определяется по накопленной частоте, первой превышающей 25%);

x_{Q_3} – нижняя граница интервала, содержащего верхний квартиль (интервал определяется по накопленной частоте, первой превышающей 75%);

i – величина интервала;

$S_{Q_{1-1}}$ – накопленная частота интервала, предшествующего интервалу, содержащему нижний квартиль;

$S_{Q_{3-1}}$ – то же для верхнего квартиля;

f_{Q_1} – частота интервала, содержащего нижний квартиль;

f_{Q_3} – то же для верхнего квартиля.

Пример. Рассмотрим расчет нижнего и верхнего квартилей по данным, характеризующим коммерческие банки по срокам функционирования (см. табл. 7.3). Определим номер Q для 1-го и 3-го квартилей:

$$\mathrm{N}_{Q_1} = \frac{100+1}{4} = 25,5; \quad \mathrm{N}_{Q_3} = \frac{100+1}{4} \cdot 3 = 76,5.$$

Применяя способ расчета, аналогичный медиане по ряду накопленных частот, определим, что:

$$3 < Q_1 < 4, \text{ т.е. } Q_1 = 3 + 1 \cdot \frac{\frac{100}{4} - 25}{21} = 3 \text{ года};$$

$$5 < Q_3 < 6, \text{ т.е. } \cdot Q_3 = 5 + 1 \cdot \frac{\frac{100}{4} \cdot 3 - 71}{12} = 5,3 \text{ года}.$$

Итак, 25% банков имеют срок функционирования менее 3 лет, 25% банков – свыше 3 лет, а остальные имеют срок функционирования в пределах от 3 до 5,3 года.

Квинтили делят распределение на пять равных частей.

Децили (d_i) – это значения вариант, которые делят ранжированный ряд на десять равных частей: 1-й дециль (d_1) делит совокупность в соотношении $^1/_{10}$ к $^9/_{10}$, 2-й дециль (d_2) – в соотношении $^2/_{10}$ к $^8/_{10}$ и т.д.

Вычисляются децили по той же схеме, что и медиана, и квартили:

$$d_1 = x_{d_1} + i \cdot \frac{\frac{1}{10}\sum f - S_{d_1 - 1}}{f_{d_1}}; \tag{7.53}$$

$$d_2 = x_{d_2} + i \cdot \frac{\frac{2}{10}\sum f - S_{d_2 - 1}}{f_{d_2}} \tag{7.54}$$

и т.д.

$$d_9 = x_{d_9} + i \cdot \frac{\frac{9}{10}\sum f - S_{d_9 - 1}}{f_{d_9}}. \qquad (7.55)$$

Пример. Продолжим пример с распределением коммерческих банков по сроку функционирования (см. табл. 7.3). Рассчитаем 1-й и 9-й децили.

Определим номер для 1-го и 9-го децилей:

$$\text{N}d_1 = \frac{100+1}{10} = 10{,}5; \ \text{N}d_9 = \frac{100+1}{10} \cdot 9 = 94{,}5.$$

По ряду накопленных частот определяем, что:

$$2 < d_1 < 3, \text{ т.е. } d_1 = 2 + 1 \cdot \frac{\frac{100}{10} - 10}{15} = 2 \text{ года.}$$

Это означает, что 10% коммерческих банков имеют срок функционирования менее 2 лет, а 90% банков имеют срок функционирования свыше 2 лет.

$$7 < d_9 < 8, \text{ т.е. } d_9 = 7 + 1 \cdot \frac{\frac{100}{10} \cdot 9 - 90}{5} = 7 \text{ лет.}$$

Это означает, что 90% банков имеют срок функционирования меньше 7 лет, а 10% банков имеют срок функционирования свыше 7 лет.

Значения признака, делящие ряд распределения на сто частей, называются *перцентилями*. Слово «перцентиль» относится непосредственно к элементу распределения или к значению, промежуточному между двумя элементами. Для того чтобы указать местоположение конкретного наблюдения, в распределении указывается так называемый *перцентильный ранг;* он равен сумме процентов, приходящихся на наблюдения, которые в распределении стоят ниже его, и половине процентов, которые приходятся на него непосредственно.

Пример. На курсе занимается 50 студентов, студент Иванов получил оценку на экзамене выше, чем 17 его товарищей. Найдем перцентильный ранг оценки студента Иванова. Вначале отметим, что 34% оценок в распределении ниже оценки Иванова (17:50 = 0,34).

Оценка студента Иванова составляет 2% от всех оценок распределения (1 : 50 = 0,02). Таким образом, перцентильный ранг оценки Иванова равен 34 плюс половина от 2, следовательно, 35. Элемент распределения с перцентильным рангом, равным 35, называется 35-м *перцентилем*. Элемент с перцентильным рангом 74 – 74-м *перцентилем* и т.д. Перцентильные ранги дают возможность проведения некоторых сопоставлений между элементами различных распределений.

Изложенный выше метод нахождения перцентилей можно представить с помощью следующей формулы:

$$P_n = L + (S / f) \cdot (i), \tag{7.56}$$

где P_n – обозначение n-го перцентиля;
L – нижняя граница интервала;
S – число оценок, необходимое попасть в точку на горизонтальной оси, которая соответствует данному перцентилю;
i – расстояние от нижней границы L до верхней границы L +1 (шаг интервала);
f – число оценок, расположенных в интервале от L до L+1.

Пример. Чтобы использовать формулу (7.56), рассмотрим следующий ряд распределения оценок учеников за диктант в табл. 7.13.

Например, по данным табл.7.13 нужно найти 35-й перцентиль. Вначале находим точку, в которой сумма накопленных частот составит 35% от 1000, или 350.Из графы 2 видно, что эта сумма находится в интервале между значениями 47–55.

Найдем 35-й перцентиль по формуле (7.56).

$$P_{35} = 47 + \frac{5}{8} \cdot 9 = 52,6 \text{ балла.}$$

Полученный результат означает, что 35% оценок учеников за диктант имеют баллов меньше, чем 52,6, и 65% оценок учеников имеют баллов больше, чем 52,6. Таким образом, слово «меньше, чем» относится к верхней границе каждого интервала.

Таблица 7.13

Распределение оценок учеников за диктант при 100-балльной оценке*

Группы оценок учеников за диктант	Число оценок	Накопленные частоты
А	1	2
2–10	18	18
11–19	32	50
20–28	50	100
29–37	81	181
38–46	119	300
47–55	80	380
56–64	140	520
65–73	120	640
74–82	160	800
83–91	140	940
92–100	60	1 000
Итого	1 000	–

* Исходные данные взяты из книги: *Вайнберг Дж., Шумекер Дж.* – М.: Статистика, 1979. – С. 71.

Рассмотренные показатели можно представить в следующем соотношении (рис. 7.3).

Использование в анализе вариационных рядов распределения рассмотренных выше характеристик позволяет глубоко и детально охарактеризовать изучаемую совокупность.

Показатели дифференциации. В тех случаях, когда при изучении вариационного ряда возникает необходимость дать относительную характеристику степени вариации ряда и имеются уже предварительно вычисленные квартили и децили, то можно вычислить *коэффициент дифференциации (К)*.

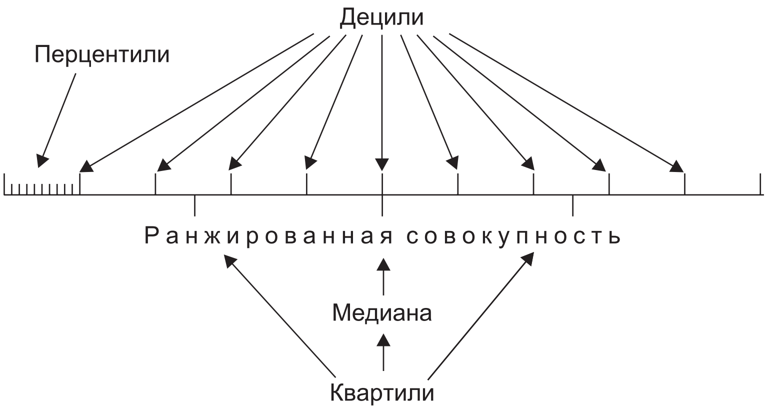

Рис. 7.3. Медиана, квартили, децили и перцентили

В зависимости от заданных ранговых показателей коэффициенты дифференциации рассчитываются по-разному.

1. Если заданы 3-я (Q_3) и 1-я (Q_1) квартили, то вместо коэффициента вариации (V), представленного в разделе 7.3, можно вычислить *коэффициент дифференциации* по формуле

$$K_v = \frac{Q_3 - Q_1}{Q_3 + Q_1} = \frac{1 - \frac{Q_1}{Q_3}}{1 + \frac{Q_1}{Q_3}}. \qquad (7.57)$$

В большинстве случаев коэффициент вариации (V) составляет примерно 1,5 коэффициента дифференциации (K_v), т.е.

$$V \approx K_v \cdot 1{,}5. \qquad (7.58)$$

Пример. По данным табл. 7.3, характеризующим коммерческие банки по сроку функционирования, имеем:

$\bar{x} = 4{,}3$ года, $\sigma = 1{,}8$ года, $Q_1 = 3$ года; $Q_3 = 5{,}3$ года.

Отсюда

$$V = \frac{1,8}{4,3} \cdot 100 = 41,9 \approx 42\%;$$

$$K = \frac{5,3-3}{5,3+3} = 0,2771 \text{ раза, или } 27,71\%,$$

$$V = 0,2771 \cdot 1,5 = 0,416 \cdot 100 = 41,6\% \approx 42\%,$$

что и требовалось доказать.

2. Если сопоставляются 9-я (d_9) и 1-я (d_1) децили, то *децильный коэффициент дифференциации* (K_d) вычисляется по формуле:

$$K_d = \frac{d_9}{d_1}. \tag{7.59}$$

Пример. По данным, представленным в табл.7.3, d_1 = 2 года; d_9 = 7 лет; отсюда K_d = 7 / 2 = 3,5 раза, т.е. минимальная величина срока функционирования 10% самых старых банков отличается от максимальной величины срока функционирования 10% самых молодых банков в 3,5 раза.

Рассмотренный выше показатель дифференциации не совсем точно измеряет уровень дифференциации, так как сопоставляется минимальная величина признака (25% или 10% самых крупных единиц совокупности) с максимальной величиной признака (25% или 10% самых мелких единиц совокупности).

3. Более точно уровень дифференциации можно измерить, сопоставив средние уровни, полученные из 10% наибольших и наименьших значений признака в совокупности. Такой показатель называется *коэффициентом фондовой дифференциации* ($K_ф$):

$$K_ф = \frac{\bar{x}_{\max}}{\bar{x}_{\min}} = \frac{\frac{1}{n}\sum_j x_j}{\frac{1}{n}\sum_s x_s} = \frac{\sum_j x_j}{\sum_s x_s}, \tag{7.60}$$

где $\sum_j x_j$ – сумма значений признака 10% самых крупных единиц в совокупности;

n – число единиц совокупности самых крупных и мелких;

$\sum x_s$ – сумма значений признака 10% самых мелких единиц в совокупности.

Пример. Рассчитаем фондовый коэффициент. Имеются данные о размере капитала 20 коммерческих банков за год, млн руб.: 6,9; 9,3; 1,3; 6,0; 13,4; 3,7; 5,1; 2,9; 1,4; 1,6; 10,9; 7,2; 3,2; 8,9; 1,2; 8,1; 2,1; 4,3; 4,5; 11,5.

Так как 10% самых крупных и 10% самых мелких банков составляют одну и ту же величину (в нашем примере 1 / 10 · 20 = 2 ед.), то для расчета фондового коэффициента подставим в формулу (7.60) соответствующие значения и получим:

$$K_{\text{ф}} = \frac{(11,5+13,4)/2}{(1,2+1,3)/2} = \frac{12,45}{1,25} = 9,96 \text{ раза.}$$

Рассчитанный коэффициент показывает, что уровень дифференциации 20 коммерческих банков по размеру капитала достаточно высок; средний размер капитала 10% самых крупных банков в 9,96 раза превышает средний размер капитала 10% самых мелких банков.

7.7 МОМЕНТЫ РАСПРЕДЕЛЕНИЯ

Для подробного описания особенностей распределения используются дополнительные характеристики, в частности, определяются *моменты распределения*. Способ моментов был разработан русским математиком П.Л. Чебышевым и успешно применен А.А. Марковым для рассмотрения возможностей использования закона нормального распределения при изучении сумм большого, но конечного числа независимых случайных величин.

Моментом k-го порядка называется средняя из $k - x$ степеней отклонений вариантов x от некоторой постоянной величины A:

$$M_k = \overline{(x - A)^k}\,. \tag{7.61}$$

При исчислении средней в качестве весов могут быть использованы частоты, частости или вероятности. При использовании в качестве весов частот или частостей моменты называются *эмпирическими*, а при использовании вероятностей – *теоретическими*.

Порядок момента определяется величиной k. Эмпирический момент k-го порядка определяется как отношение суммы произведений k-х степеней отклонений вариантов от постоянной величины A на частоты к сумме частот:

$$M_k = \frac{\sum_i (x_i - A)^k f_i}{\sum_i f_i}. \quad (7.62)$$

В зависимости от выбора постоянной величины A различают три вида моментов:

1. *Начальные моменты* (M_k) получаются, если постоянная величина A равна нулю ($A = 0$):

$$M_k = \overline{(x_i - 0)^k} = \frac{\sum_i x_i^k f_i}{\sum_i f_i}. \quad (7.63)$$

2. *Условные и начальные относительно x_0 моменты* (m_k) получаются при A равном не нулю, а некоторой производной величине x_0 (начало отсчета):

$$m_k = \overline{(x_i - x_0)^k} = \frac{\sum_i (x_i - x_0)^k f_i}{\sum_i f_i}. \quad (7.64)$$

С помощью условных моментов упрощается расчет основных характеристик ряда распределения. При подстановке различных значений k получаем начальные моменты относительно x_0. Так, например, если $k = 1$, то:

$$m_1 = \frac{\sum_i (x_i - x_0) f_i}{\sum_i f_i} = \frac{\sum_i x_i f_i}{\sum_i f_i} - \frac{x_0 \sum_i f_i}{\sum_i f_i} = \bar{x} - x_0.$$

Из этой формулы вытекает, что $\bar{x} = x_0 + m_1$, т.е. средняя арифметическая равна началу отсчета плюс начальный момент первого порядка. Если отклонения $(x_i - x_0)$ имеют общий множитель C, то на него можно разделить отклонения, а по окончании вычислить полученный момент, умножив на этот множитель в соответствующей степени, т.е.:

$$m_k = \frac{\sum\limits_{i=1} \left(\frac{x_i - x_0}{C} \right)^k \cdot f_i}{\sum\limits_{i=1} f_i} \cdot C^k. \tag{7.65}$$

Отсюда следует, что при $k = 1$ $\bar{x} = x_0 + m_1 \cdot C$.

3. *Центральные моменты* (μ_k) получаются, если за постоянную величину A взять среднюю арифметическую $(A = \bar{x})$:

$$\mu_k = \overline{(x_i - \bar{x}_k)^k} = \frac{\sum\limits_i (x_i - \bar{x})^2 f_i}{\sum\limits_i f_i}. \tag{7.66}$$

В статистической практике пользуются в основном моментами 1-го, 2-го, 3-го и 4-го порядков, которые представлены в табл. 7.14.

Таким образом, анализируя формулы моментов распределения в табл. 7.14, можно сделать следующие выводы:

- начальный момент первого порядка представляет собой среднюю арифметическую и используется как показатель центра распределения $(M_1 = \bar{x})$;
- начальные моменты 2-го, 3-го и 4-го порядков не имеют самостоятельного значения, а используются для упрощения вычислений центральных моментов. Например, используя начальные моменты 1-го и 2-го порядка, можно получить дисперсию по такой формуле[1]:

$$у^2 = м_2 = м_2 - м_1^2; \tag{7.67}$$

[1] См.: *Ефимова М.Р., Петрова Е.В., Румянцев В.Н.* Общая теория статистики. – М.: Инфра-М, 1997. – С. 137.

Таблица 7.14

Виды моментов распределения четырех порядков

Порядок \ Виды моментов	Начальные	Центральные	Σсловные
1-й	$M_1 = \frac{\Sigma x_i f_i}{\Sigma f_i} = \bar{x}$	$\mu_1 = \frac{\Sigma(x_i - \bar{x}) \cdot f_i}{\Sigma f_i}$	$m_1 = \frac{\Sigma(x_i - A) f_i}{\Sigma f_i}$
2-й	$M_2 = \frac{\Sigma x_i^2 f_i}{\Sigma f_i} = \overline{x^2}$	$\mu_2 = \frac{\Sigma(x_i - \bar{x})^2 \cdot f_i}{\Sigma f_i}$	$m_2 = \frac{\Sigma(x_i - A)^2 f_i}{\Sigma f_i}$
3-й	$M_3 = \frac{\Sigma x_i^3 f_i}{\Sigma f_i} = \overline{x^3}$	$\mu_3 = \frac{\Sigma(x_i - \bar{x})^3 \cdot f_i}{\Sigma f_i}$	$m_3 = \frac{\Sigma(x_i - A)^3 f_i}{\Sigma f_i}$
4-й	$M_4 = \frac{\Sigma x_i^4 f_i}{\Sigma f_i} = \overline{x^4}$	$\mu_4 = \frac{\Sigma(x_i - \bar{x})^4 \cdot f_i}{\Sigma f_i}$	$m_4 = \frac{\Sigma(x_i - A)^4 f_i}{\Sigma f_i}$

- центральный момент 1-го порядка всегда равен нулю в соответствии с нулевым свойством средней арифметической ($\mu_1 = 0$);
- центральный момент 2-го порядка представляет собой дисперсию и служит основной мерой колеблемости признака ($\mu_2 = \sigma^2$);
- центральный момент 3-го порядка служит мерой асимметрии распределения, а если распределение симметрично, он равен нулю ($\mu_3 = 0$);
- центральный момент четвертого порядка применяется при вычислении показателя эксцесса;
- условные моменты 1-го, 2-го, 3-го и 4-го порядков не имеют самостоятельного значения, а используются для упрощения вычислений центральных моментов.

7.8 ИЗУЧЕНИЕ ФОРМЫ РАСПРЕДЕЛЕНИЯ

Для обобщающей характеристики особенностей формы распределения применяются *кривые распределения*. Кривая распределения выражает графически (полигон, гистограмма) закономерность распределения единиц совокупности по величине варьирующего признака. Различают эмпирические и теоретические кривые распределения. *Эмпирическая кривая распределения* – это фактическая кривая распределения, полученная по данным наблюдения, в которой отражаются как общие, так и случайные условия, определяющие распределение. *Теоретическая кривая распределения* – это кривая, выражающая функциональную связь между изменением варьирующего признака и изменением частот и характеризующая определенный тип распределения. При этом теоретическое распределение играет роль некоторой идеализированной модели эмпирического распределения, а сам анализ вариационного ряда сводится к сопоставлению эмпирического и теоретического распределений.

Кривые распределения бывают симметричными и асимметричными. В зависимости от того, какая ветвь кривой вытянута – правая или левая, различают правостороннюю или левостороннюю асимметрию. Кривые распределения могут быть одно-, двух- и многовершинными.

Для однородных совокупностей, как правило, характерны одновершинные распределения. Многовершинность свидетельствует о неоднородности изучаемой совокупности. Появление двух и более вершин делает необходимой перегруппировку данных с целью выделения более однородных групп. Для симметричных распределений частоты любых двух вариант, равноотстоящих в обе стороны от центра, равны между собой. Рассчитанные для таких рядов распределений характеристики равны: $\bar{x}$ = Mo = Me, R = 6 · σ; σ = 1,25d. Если указанные соотношения нарушены, то это свидетельствует о наличии асимметрии распределения. Так, при Mo > Me > $\bar{x}$ разности между $\bar{x}$ – Mo и $\bar{x}$ – Me положительные и асимметрия правосторонняя, а при Mo < Me < $\bar{x}$, наоборот, разности $\bar{x}$ – Mo и $\bar{x}$ – Me отрицательные и асимметрия левосторонняя.

При сравнительном изучении асимметрии нескольких распределений с разными единицами измерения вычисляется относительный показатель асимметрии (As):

$$\mathrm{As} = \frac{\bar{x} - \mathrm{Mo}}{\text{у}}, \quad \text{или } \mathrm{As} = \frac{\bar{x} - \mathrm{Me}}{\text{у}}. \qquad (7.68)$$

Его величина может быть положительной и отрицательной. В первом случае речь идет о правосторонней асимметрии, а во втором – о левосторонней (рис. 7.4).

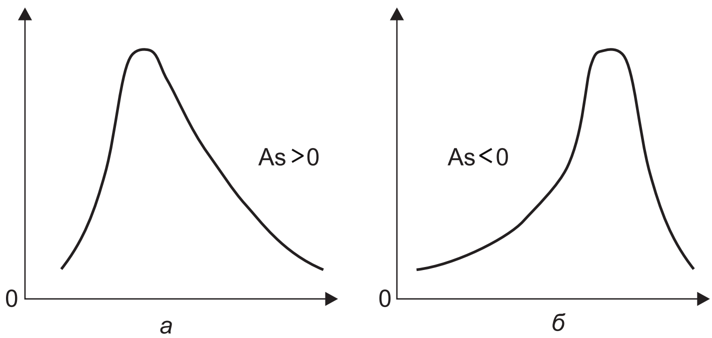

Рис. 7.4. Асимметричные ряды распределения:
а – As > 0 – с правосторонней асимметрией;
б – As < 0 – с левосторонней асимметрией

В симметричном распределении центральный момент 3-го порядка μ_3=0, поэтому чем он больше, тем больше и асимметрия. Эта особенность и используется для характеристики асимметрии. *Коэффициент асимметрии* равен отношению центрального момента 3-го порядка к среднему квадратическому отклонению в кубе, т.е.

$$\mathrm{As} = \frac{\text{м}_3}{\text{у}^3}. \qquad (7.69)$$

Если As > 0, то асимметрия правосторонняя, а если As < 0, то асимметрия левосторонняя. Чем числитель ближе к 0, тем асимметрия меньше. Этот показатель асимметрии более точен по сравнению с предыдущими и применяется более широко. Принято считать, что асимметрия выше 0,5 (независимо от знака) считается значительной; если она меньше 0,25, то незначительной.

Оценка существенности As проводится на основе средней квадратической ошибки, коэффициента асимметрии σ_{As}, которая зависит от числа наблюдений (n) и рассчитывается по формуле:

$$\sigma_{As} = \sqrt{\frac{6(n-1)}{(n+1)(n+3)}}\,. \tag{7.70}$$

В случае $\frac{|As|}{у_{As}} > 3$ асимметрия существенна и распределение признака в генеральной совокупности несимметрично. В противном случае асимметрия несущественна и ее наличие может быть вызвано случайными обстоятельствами.

Покажем расчет коэффициента асимметрии на условном примере данных размера выданных кредитов коммерческими банками (табл. 7.15).

Таблица 7.15

Распределение коммерческих банков по размеру выданных кредитов

Группы банков по размеру кредита, млн руб. x	Число банков f	Середина интервала x'	$x'_i f_i$	$x'_i - \bar{x}$	$(x'_i - \bar{x})^2 f_i$	$(x'_i - \bar{x})^3 f_i$
A	1	2	3	4	5	6
1–6	6	3,5	21	–10	600	–6000
6–11	3	8,5	25,5	–5	75	–375
11–16	11	13,5	148,5	0	0	0
16–21	5	18,5	92,5	5	125	625
21–26	5	23,5	117,5	10	500	5000
Итого	30	–	405	–	1300	–750

На основе данных табл. 7.15 определим коэффициент асимметрии по формуле (7.69), для этого сделаем следующее: определим центральный момент второго и третьего порядков по формуле (7.66).

Все расчеты сделаем в табл. 7.15 и найдем:

$$\bar{x} = \frac{\Sigma x_i' f_i}{\Sigma f_i} = \frac{405}{30} = 13,5;$$

$$м_2 = у^2 = \frac{\Sigma (x_i' - \bar{x})^2 f_i}{\Sigma f_i} = \frac{1300}{30} = 43,33;$$

$$м_3 = \frac{\Sigma (x_i' - \bar{x})^3 f_i}{\Sigma f_i} = \frac{-750}{30} = -25;$$

$$A_s = \frac{м_3}{у^3} = \frac{-25}{(6,58)^3} = \frac{-25}{284,89} = -0,09.$$

Полученный результат свидетельствует о наличии незначительной по величине и отрицательной по своему характеру асимметрии.

Для симметричных распределений может быть рассчитан показатель эксцесса (Ek). Наиболее точно он определяется по формуле с использованием центрального момента 4-го порядка $\left(м_4 = \frac{\Sigma (x_i - \bar{x})^4 f_i}{\sum_i f_i} \right)$:

$$\mathrm{Ek} = \frac{м_4}{у^4} - 3. \tag{7.71}$$

На рис. 7.5 и 7.6 представлены два распределения: островершинное (Ek положительный) и плосковершинное (Ek отрицательный). В нормальном распределении Ek = 0.

Среднеквадратическая ошибка эксцесса (σ_{Ek}) рассчитывается по формуле:

$$у_{\mathrm{Ek}} = \sqrt{\frac{24n \cdot (n-2) \cdot (n-3)}{(n-1)^2 \cdot (n+3) \cdot (n+5)}}, \tag{7.72}$$

где n – число наблюдений.

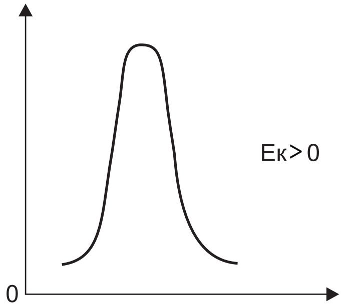

Рис. 7.5. Островершинное распределение

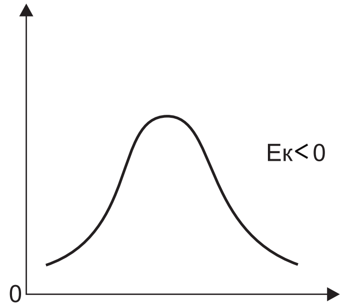

Рис. 7.6. Плосковершинное распределение

Для определения асимметрии и эксцесса можно пользоваться упрощенными формулами, предложенными Линдбергом:

$$\mathrm{As} = P - 50, \tag{7.73}$$

где P – удельный вес (%) количества тех вариант, которые превосходят среднюю арифметическую, в общем количестве вариант данного ряда;
50 – удельный вес (%) вариант, превосходящих среднюю арифметическую ряда нормального распределения.

$$\mathrm{Ek} = P - 38{,}29, \tag{7.74}$$

где P – доля (%) количества вариант, лежащих в интервале, равном половине среднего квадратического отклонения (в ту или другую сторону от величины средней в общем количестве вариант данного ряда);
38,29 – доля (%) количества вариант, лежащих в интервале, равном половине среднего квадратического отклонения (в ту или другую сторону от величины средней), в общем количестве вариант ряда нормального распределения.

Необходимо отметить, что хотя показатели асимметрии и эксцесса характеризуют непосредственно лишь форму распределения признака в пределах изучаемой совокупности, однако их определение имеет не только описательное значение. Часто асимметрия и эксцесс дают определенные указания для дальнейшего исследования социально-экономических явлений. Например, появление значительного отрицательного эксцесса может указывать на качественную неоднородность исследуемой совокупности. Кроме того, эти показатели позволяют сделать вывод о возможности отнесения данного эмпирического распределения к типу кривых нормального распределения.

7.9 ТЕОРЕТИЧЕСКИЕ РАСПРЕДЕЛЕНИЯ В АНАЛИЗЕ ВАРИАЦИОННЫХ РЯДОВ

Для аппроксимации (выравнивания) эмпирических кривых распределения и сопоставления их с теоретическими в статистике часто пользуются *нормальным распределением*, функция которого равняется

$$y_t = \frac{1}{\sqrt{2p}} \cdot e^{\frac{1}{2}t^2}, \qquad (7.75)$$

где y_t – ордината кривой нормального распределения;

$t = \frac{x - \bar{x}}{y}$ – стандартизованное отклонение;

е и π – математические постоянные;

x – варианты вариационного ряда;

$\bar{x}$ – их средняя величина;

σ – среднее квадратическое отклонение.

Нормальное распределение полностью определяется двумя параметрами – средней арифметической $(\bar{x})$ и средним квадратическим отклонением σ. Подчиненность закону нормального распределения проявляется тем точнее, чем больше случайных величин действуют вместе. Если ни одна из случайно действующих причин по своему действию не окажется преобладающей над другими, то закон распределения очень близко подходит к нормальному.

Такая закономерность проявляется, например, в распределении отклонений в производственном процессе при нормальном уровне организации и технологии, в распределении населения определенного возраста по размеру обуви и т.д.

Часто возникают распределения, хотя и не отвечающие строго нормальному распределению, но имеющие с ним сходство. Такие сходные черты часто обусловлены тем, что крайние значения вариантов, близкие к x_{min} и x_{max}, встречаются много реже, чем срединные.

Рассмотрим некоторые *свойства кривой нормального распределения:*

- $f(t)$ – функция нормального распределения – четная, т.е. $f(-t) = f(+t)$. Следовательно, изображающая ее кривая распределена симметрично относительно оси ординат, т.е. $\bar{x}$ = Мо = Ме;
- функция имеет бесконечно малые значения при $t = \pm\infty$. Это означает, что ветви кривой удалены в бесконечность и асимптотически приближаются к оси абсцисс. При этом чем больше значения признака отклоняются от $\bar{x}$, тем реже встречаются;
- функция имеет максимум при $t = 0$.Отсюда следует, что модального значения кривая достигает при $t = 0$ или при $x = \bar{x}$. Величина максимума составляет $1/\sqrt{2\pi}$;
- при $t = \pm1$ функция дает точки перегиба. Следовательно, при отклонении значений признака (x) от среднего значения ($\bar{x}$) в положительном и отрицательном направлениях на одно стандартное (нормированное) отклонение ($\pm\sigma$ от $\bar{x}$) кривая дает переход от выпуклости к вогнутости;
- если случайная величина представляет сумму двух независимых случайных величин, следующих каждая нормальному закону, то она тоже следует нормальному закону;
- площадь между кривой и осью оt равна единице, как интеграл Пуассона.

Пример. Рассмотрим данные распределения образцов крепости одиночной нити (табл. 7.16).

Поскольку нормальное распределение зависит от двух параметров: $\bar{x}$ и σ, прежде всего определим соответствующие характеристики.

В графах 1 и 2 табл.7.16 приведены фактические варианты и частоты. Расчет $\bar{x}$ и σ произведен обычным способом.

Для расчета частот нормального распределения 500 образцов крепости нити со средней $\bar{x}$ = 64,66 г и средним квадратическим отклонением σ = 3,1 г необходимо использовать формулу плотности вероятности:

$$P_x = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{\frac{(x-\bar{x})^2}{2\sigma^2}}. \qquad (7.76)$$

Таблица 7.16

Расчет теоретических частот нормального распределения

Крепость одиночной нити, г x	Количество образцов f	Середина интервала x'	$x - \bar{x}$	$t = \frac{x - \bar{x}}{у}$	$f(t) = \frac{1}{\sqrt{2\pi}} \cdot e^{\frac{1}{2}t^2}$	Теоретические частоты $f_m = \frac{k\Sigma f}{у} f(t)$	
						исчисленные	округленные
1	2	3	4	5	6	7	8
56–58	5	57	–7,66	2,47	0,01888	6,09	6
58–60	29	59	–6,66	1,83	0,07477	24,12	24
60–62	63	61	–3,66	1,18	0,19886	64,15	64
62–64	117	65	0,34	0,11	0,39654	127,92	128
64–66	116	63	–1,66	0,54	0,34482	127,92	128
66–68	102	67	2,34	0,75	0,30114	97,14	97
68–70	48	69	4,34	1,40	0,14973	48,30	48
70–72	14	71	6,34	2,04	0,04980	16,06	16
72–74	6	73	8,34	2,69	0,01071	3,55	4
Итого	500	–	–	–	–	499,06	499

Чтобы прийти к частотам нормального распределения f_m, необходимо выразить их через P_x.

Для удобства вычислений вероятностей случайные величины нормируются, а затем используются заранее табулированные значения плотности функции распределения нормированной случайной величины. Первый множитель такой функции – величина постоянная для данного распределения. В нашем случае: $\frac{k\Sigma f}{у} = \frac{2 \cdot 500}{3{,}1} = 322{,}6$, во втором множителе выражение $\frac{x - \bar{x}}{у}$ обозначим через t, тогда получим:

$$\frac{1}{\sqrt{2р}} \cdot e^{\frac{(x-\bar{x})^2}{2σ^2}} = \frac{1}{\sqrt{2р}} \cdot e^{\frac{t^2}{2}}.$$

Полученную функцию от t обозначим $f(t)$:

$$f(t) = \frac{1}{\sqrt{2р}} \cdot e^{\frac{t^2}{2}}. \qquad (7.77)$$

В математической статистике существуют специальные таблицы для любых значений $f(t)$ (приложение 8).

Таким образом, $f_m = \frac{k\Sigma f}{у} \cdot f(t)$ очень легко рассчитать, определив для каждого значения варианта x' величину $t = \frac{x - \bar{x}}{у}$ (графа 5) и найдя по таблицам соответствующие $f(t)$ (графа 6). Умножив $f(t)$ на постоянный для всех частот множитель $\frac{k\Sigma f}{у}$, получим теоретические частоты нормального распределения f_m (графа 7).

Сравнивая полученные f_m (графа 7) с фактическими частотами f (графа 2), убеждаемся, что их расхождения невелики. На графике, представленном на рис.7.7, видна довольно большая близость фактических частот распределения к теоретическим.

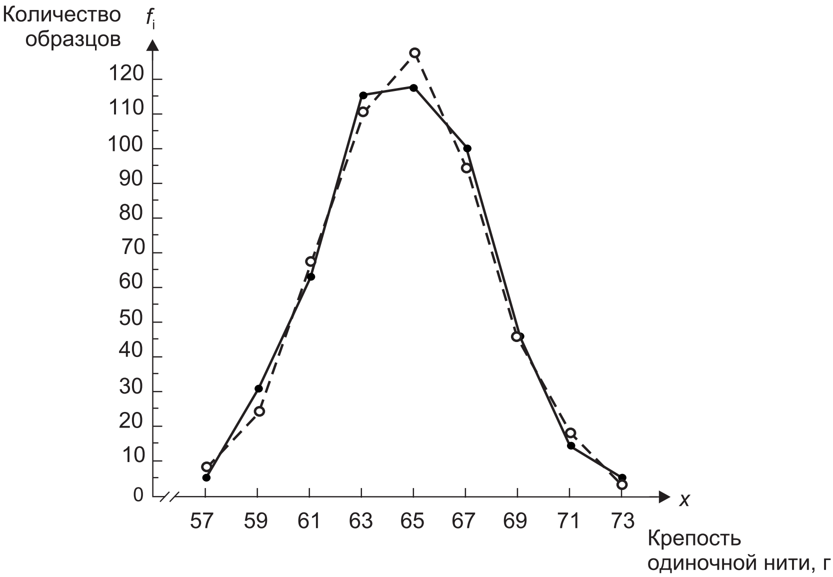

Рис. 7.7. Эмпирические и теоретические данные распределения крепости одиночной нити в 500 образцах

В то же время нельзя не отметить, что сопоставление графика эмпирических частот с теоретическими в целях определения соответствия эмпирического распределения нормальному позволяет оценивать эти расхождения только субъективно. Объективная характеристика соответствия может быть получена с помощью особых статистических показателей – критериев согласия. Известны критерии согласия К. Пирсона (хи-квадрат), В.И. Романовского, А.Н. Колмогорова и Б.С. Ястремского.

Критерий согласия Пирсона (χ^2) вычисляется по формуле:

$$ч^2 = \frac{(f_э - f_т)^2}{f_т}, \qquad (7.78)$$

где $f_э$ и $f_т$ – эмпирические и теоретические частоты соответственно.

С помощью величины χ^2 по специальным таблицам приложения 3 определяется вероятность $P(\chi^2)$. Входами в таблицу являются значения χ^2 и число степеней свободы $y = n - 1$. На основе P выносится суждение о существенности или несущественности расхождения между эмпирическим и теоретическим распределением. При $P > 0{,}5$ считается, что эмпирическое и теоретическое распределения близки, при $P \in [0{,}2;\ 0{,}5]$ совпадение между ними удовлетворительное, в остальных случаях – недостаточное.

Если число степеней свободы большое, то применяется соотношение, равное $\sqrt{2ч^2} - \sqrt{2г\text{-}1}$. Расхождение между эмпирическим и теоретическим распределениями существенно при значениях этой разности, заметно превосходящих 2.

Критерий Романовского (C), также используемый для проверки близости эмпирического и теоретического распределений, определяется следующим образом:

$$C = \frac{ч^2 - г}{\sqrt{2г}}, \tag{7.79}$$

где χ^2 – критерий Пирсона, рассчитываемый по формуле (7.78);
γ – число степеней свободы (при проверке гипотезы о нормальности распределения равно числу групп минус три).

При $C < 3$ различие несущественно, что позволяет считать эмпирическое распределение близким к нормальному.

Критерий Ястремского (L) может быть найден на основе следующего соотношения:

$$L = \frac{\Sigma \dfrac{(f_3 - f_T)^2}{N \cdot pq} - K}{\sqrt{2K + 4Q}}, \tag{7.80}$$

где N – объем совокупности;
pq – дисперсия альтернативного признака;
K – число вариантов или групп;
Q – принимает значение 0,6 при числе вариантов или групп от 8 до 20.

Если $L > 3$, то эмпирическое распределение соответствует теоретическому.

Критерий Колмогорова (λ) вычисляется по формуле:

$$л = \frac{D}{\sqrt{\Sigma f}}, \quad (7.81)$$

где D – максимальное значение разности между накопленными эмпирическими и теоретическими частотами;

Σf – сумма эмпирических частот.

Необходимым условием использования этого критерия является достаточно большое число наблюдений (не меньше ста).

Рассмотрим применение критерия Колмогорова (табл. 7.17).

Пример. Как видно из табл. 7.17, максимальное значение разности между эмпирическими и теоретическими частотами составляет 7, т.е. $D = 7$.

Следовательно, величина критерия Колмогорова в нашем случае равна:

$$л = \frac{7}{\sqrt{500}} \approx 0,31.$$

По специальным таблицам вероятностей $P(\lambda)$ определяем, что λ соответствует $P(x)$, близкой к 1,000. Это означает, что с вероятностью, близкой к 1, можно утверждать, что отклонения фактических частот от теоретических в нашем примере являются случайными. Следовательно, в основе фактического распределения образцов по крепости нити лежит закон нормального распределения.

В статистике широко используются различные виды теоретических распределений, кроме нормального распределения, – биномиальное распределение, распределение Пуассона и др. Каждое из теоретических распределений имеет специфику и свою область применения в различных отраслях знания.

Рассмотрение других видов распределения, кроме нормального, не является предметом изучения в данной главе.

Таблица 7.17

Расчет критерия Колмогорова по данным крепости одиночной нити в 500 образцах

Крепость одиночной нити, г	Частоты ряда распределения		Накопленные частоты		$\lvert \Sigma f_э - \Sigma f_т \rvert$
	эмпирические $f_э$	теоретические $f_т$	эмпирические $\Sigma f_э$	теоретические $\Sigma f_т$	
56–58	5	6	5	6	1
58–60	29	24	34	30	4
60–62	63	64	97	94	3
62–64	116	112	213	206	7
64–66	117	128	330	334	4
66–68	102	97	432	431	1
68–70	48	48	480	479	1
70–72	14	16	494	495	1
72–74	6	4	500	499	1
Итого	500	499	–	–	-

ОСНОВНЫЕ ПОНЯТИЯ

Абсолютные показатели вариации – это размах вариации, среднее линейное отклонение, среднее квадратическое отклонение и дисперсия.

Вариация – колеблемость, многообразие, изменяемость величины признака у отдельных единиц совокупности.

Внутригрупповая дисперсия отражает случайную вариацию, т.е. часть вариации, происходящей под влиянием неучтенных факторов и не зависящую от признака-фактора.

Децили – значения признака, делящие ранжированную совокупность на десять равных частей.

Дисперсия – средний квадрат отклонений индивидуальных значений признака от их средней величины.

Закономерности распределения – закономерности изменения частот в вариационных рядах.

Квартили – значения признака, делящие ранжированную совокупность на четыре равновеликие части.

Коэффициент вариации – процентное отношение среднего квадратического отклонения к средней величине признака.

Коэффициент осцилляции – процентное отношение размаха вариации к средней величине признака.

Кривая распределения – графическое изображение в виде непрерывной линии изменения частот в вариационном ряду, функционально связанном с изменением вариант.

Критерии согласия – особые статистические показатели, характеризующие соответствие эмпирического и теоретического распределений. Известны критерии согласия К. Пирсона, В.И. Романовского, А.Н. Колмогорова, Б. С. Ястремского.

Линейный коэффициент вариации – процентное отношение среднего линейного отклонения к средней величине признака.

Межгрупповая дисперсия характеризует систематическую вариацию, т.е. различия в величине изучаемого признака, возникающие под действием признака-фактора, положенного в основу группировки.

Мода и медиана – структурные средние. *Мода* – значение изучаемого признака, повторяющееся с наибольшей частотой. *Медиана* – значение признака, приходящееся на середину ранжированной совокупности. Структурные средние могут быть определены по дискретным и интервальным рядам распределения.

Общая дисперсия измеряет вариацию признака во всей совокупности под влиянием всех факторов, обусловивших эту вариацию.

Относительные показатели вариации – это коэффициенты осцилляции, вариации, относительное линейное отклонение и др.

Перцентили – значения признака, делящие ранжированную совокупность на сто равных частей.

Размах вариации – разность между наибольшим и наименьшим значениями варьирующего признака.

Среднее квадратическое отклонение рассчитывается как корень квадратный из дисперсии. Среднее квадратическое отклонение, дисперсия и среднее линейное отклонение могут определяться по формулам простой и взвешенной (в зависимости от исходных данных).

Среднее линейное отклонение – средняя арифметическая из абсолютных значений отклонений вариант признака от их средней.

Теоретическая кривая распределения – кривая, выражающая общую закономерность данного типа распределения в чистом виде, исключающем влияние случайных факторов.

Эмпирический коэффициент детерминации – доля межгрупповой дисперсии в общей дисперсии.

Эмпирическое корреляционное отношение – корень квадратный из эмпирического коэффициента детерминации.

Энтропия – мера неопределенности данных наблюдения, которая может иметь различные результаты. Зависит от числа градаций признака и вероятности каждой из них.

ТЕСТЫ

1. Вариация – это:

а) изменение массовых явлений во времени;

б) изменение структуры статистической совокупности в пространстве;

в) изменение значений признака во времени и в пространстве;

г) изменение состава совокупности.

2. Какой из показателей вариации характеризует абсолютный размер колеблемости признака около средней величины:

а) коэффициент вариации;

б) дисперсия;

в) размах вариации;

г) среднее квадратическое отклонение.

3. Что характеризует коэффициент вариации:

а) диапазон вариации признака;

б) степень вариации признака;

в) тесноту связи между признаками;

г) пределы колеблемости признака.

4. Если все значения признака увеличить в 16 раз, то дисперсия:

а) не изменится;

б) увеличится в 16 раз;

в) увеличится в 256 раз;

г) увеличится в 4 раза;

д) предсказать изменение дисперсии нельзя.

5. Чему равна межгрупповая дисперсия, если отсутствуют различия между вариантами внутри групп:

а) единице;

б) нулю;

в) колеблется от нуля до единицы;

г) общей дисперсии;

д) средней из групповых дисперсий.

6. Коэффициент детерминации измеряет:

а) степень тесноты связи между исследуемыми явлениями;

б) вариацию, сложившуюся под влиянием всех факторов;

в) долю вариации признака-результата, сложившуюся под влиянием изучаемого (изучаемых) фактора (факторов);

г) вариацию, связанную с влиянием всех остальных факторов, кроме исследуемого (исследуемых).

7. При анализе данных о росте призывников получен коэффициент асимметрии As = 0,732 и показатель эксцесса Ek = 3,456. Это значит, что распределение:

а) нормальное;

б) имеет правостороннюю асимметрию;

в) имеет левостороннюю асимметрию;

г) островершинное;

д) плосковершинное.

8. Проверяется соответствие эмпирического распределения нормальному. Статистическая совокупность из 245 единиц разделена на 16 групп. Число степеней свободы для критерия χ^2 равно:

а) 244;

б) 242;

в) 16;

г) 15;

д) 13.

9. Критерий Колмогорова может быть рассчитан на основе:

а) индивидуальных данных;

б) частот;

в) частостей.

10. Средний размер реализованной торговой фирмой спортивной обуви равен 39, Mo = 39, Me = 39. На основе этого можно сделать вывод, что распределение проданной спортивной обуви по размеру:

а) симметричное;

б) приближенно симметричное;

в) с левосторонней асимметрией;

г) с правосторонней асимметрией;

д) данные не позволяют сделать вывод.

ЛИТЕРАТУРА

Вайнберг Дж., Шумакер Дж. Статистика. – М.: Статистика, 1979.

Виноградова Н.М. и др. Общая теория статистики. – М.: Статистика, 1968.

Венецкий И.Г. , Кильдишев Г.С. Основы теории вероятностей и математической статистики. – М.: Статистика, 1968.

Ежов А.И. Выравнивание и вычисление рядов распределений. – М.: Госстатиздат, 1961.

Ефимова М.Р. и др. Общая теория статистика. – М.: Инфра-М, 1997.

Карамзин Н. Изучение вариации сублинейными отклонениями//Вестник статистики. – 1986. – № 6. – С. 63–65.

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

Крылова Н.И. Ряды распределения. – М.: МИНХ, 1972.

Методы математической статистики в экономических исследованиях. – М.: МЭСИ, 1983.

Миллс Ф. Статистические методы. – М.: Госстатиздат, 1958.

Мхитарян В.С., Трошин Л.И. Дисперсионный анализ. – М.: МЭСИ, 1990.

Теория статистики / Под ред. Г.Л. Громыко. – М.: Инфра-М, 2000.

Френкель А.А., Адамова Е.В. Вариационные ряды и их статистические характеристики. – М.: МЭСИ, 1986.

Юзбашев М.М., Манелля А.И. Статистический анализ тенденций и колеблемости. – М.: Финансы и статистика, 1983.

ГЛАВА 8
ВЫБОРОЧНОЕ НАБЛЮДЕНИЕ

8.1
ЗНАЧЕНИЕ И ТЕОРЕТИЧЕСКИЕ ОСНОВЫ ВЫБОРОЧНОГО НАБЛЮДЕНИЯ

Выборочным называется такое несплошное наблюдение, при котором признаки регистрируются у отдельных единиц изучаемой статистической совокупности, отобранных с использованием специальных методов, а полученные в процессе обследования результаты с определенным уровнем вероятности распространяются на всю исходную совокупность.

Выборочное наблюдение нельзя отождествлять с несплошным обследованием вообще, так как оно является лишь одним из видов последнего, наиболее проработанным с методологической и организационной точек зрения. Помимо выборочного наблюдения несплошное обследование может осуществляться путем *монографического описания*, *методом основного массива* или на основе различных видов *анкетирования*, когда отсутствуют какие-либо специальные методы отбора респондентов и процент заполненных и возвращенных анкет заранее не известен.

Преимущества выборочного наблюдения заключаются в существенной экономии различного вида ресурсов, а именно:

- финансовых средств, затрачиваемых на сбор и обработку данных, подготовку и оплату кадров;
- материально-технических ресурсов (канцелярские товары, оргтехника, расходные материалы, транспортное обслуживание и т.п.);
- трудовых ресурсов, привлекаемых к обследованию на всех его этапах;
- времени, затрачиваемого как на получение первичной информации, так и на последующую ее обработку, вплоть до публикации итоговых материалов.

В то же время необходимо четко представлять, что выборочное наблюдение, как бы грамотно с методологической точки зрения оно ни было организовано, всегда связано с определенными, пусть небольшими и измеряемыми ошибками. Поэтому, когда вариация регистрируемых признаков очень сильная и процент отбора для получения выборочных значений с заданной точностью достигает 20–25%, следует правильно оценить целесообразность несплошного обследования, сопоставив достаточно большие затраты всех ресурсов на такую объемную выборку и ожидаемые погрешности статистических характеристик. Вполне вероятно, что проведение сплошного обследования в подобных случаях будет более оправданным.

В то же время при решении ряда задач выборочное наблюдение является единственно возможным способом получения необходмой информации. Так, контроль многих видов продукции связан с их порчей, потерей товарного вида, нарушением герметизации и т.п. Например, нельзя проверить каждую производимую предприятием электролампу на соблюдение требований по продолжительности горения. Нельзя проверить на соответствие стандартам каждого пакета с соком или молочной продукцией, так как это связано с вскрытием их упаковки. В подобных случаях контроль качества может осуществляться только с использованием выборочного метода.

Реализация выборочного метода базируется на понятиях генеральной и выборочной совокупностей.

Генеральной совокупностью называется вся исходная изучаемая статистическая совокупность, из которой на основе отбора единиц или групп единиц формируется ***совокупность выборочная***. Поэтому генеральную совокупность также называют основой выборки.

Отбор единиц в выборочную совокупность может быть повторным или бесповторным.

При *повторном отборе* попавшая в выборку единица подвергается обследованию, т.е. регистрации значений ее признаков, возвращается в генеральную совокупность и наравне с другими единицами участвует в дальнейшей процедуре отбора. Таким образом некоторые единицы могут попадать в выборку дважды, трижды или даже большее число раз. И при изучении выборочной совокупности они будут рассматриваться как отдельные независимые наблюдения.

Отметим, что число единиц генеральной совокупности, участвующих в отборе, при таком подходе остается постоянным. Поэтому вероятность попадания в выборку для всех единиц совокупности на протяжении всего процесса отбора также не меняется.

На практике методология повторного отбора обычно используется в тех случаях, когда объем генеральной совокупности не известен и теоретически возможно повторение единиц с уже встречавшимися значениями всех регистрируемых признаков.

Например, при проведении маркетинговых исследований мы не можем сколько-нибудь точно оценить, какое число потребителей предпочитает стиральный порошок конкретной торговой марки, сколько покупателей предпочитает делать покупки именно в данном супермаркете и т.д. Поэтому возможно повторение совершенно идентичных единиц по причине как практически неограниченных объемов совокупности, так и возможной повторной регистрации. Предположим, при проведении обследования один и тот же покупатель может дважды прийти в магазин и дважды подвергнуться обследованию.

При выборочном контроле качества продукции объем генеральной совокупности также часто не определен, так как процесс производства может осуществляться постоянно, каждый день дополняя генеральную совокупность новыми единицами-изделиями. Поэтому в выборочную совокупность могут попасть два и более изделий с абсолютно одинаковыми характеристиками. Следовательно, и в этом случае при обработке результатов выборки необходимо ориентироваться на методологию, используемую при повторном отборе.

При *бесповторном отборе* попавшая в выборку единица подвергается обследованию и в дальнейшей процедуре отбора не участвует. Такой отбор целесообразен и практически возможен в тех случаях, когда объем генеральной совокупности четко определен. Получаемые при этом результаты, как правило, являются более точными по сравнению с результатами, основанными на повторной выборке.

Необходимо отметить, что в выборочную совокупность могут отбираться не только отдельные единицы, но и группы единиц. В первом случае отбор называется *индивидуальным*, во втором случае – *групповым*.

Как уже отмечалось выше, выборочное наблюдение всегда связано с определенными ошибками получаемых характеристик. Классификация этих ошибок представлена на рис. 8.1.

Ошибки регистрации являются следствием неправильного установления значения наблюдаемого признака или неправильной записи. Они свойственны не только выборочному, но и сплошному наблюдению (эти ошибки были рассмотрены в разд. 2.2).

Ошибки репрезентативности обусловлены тем, что выборочная совокупность не может по всем параметрам в точности воспроизвести генеральную совокупность. Получаемые расхождения называются ошибками репрезентативности, или представительности, так как

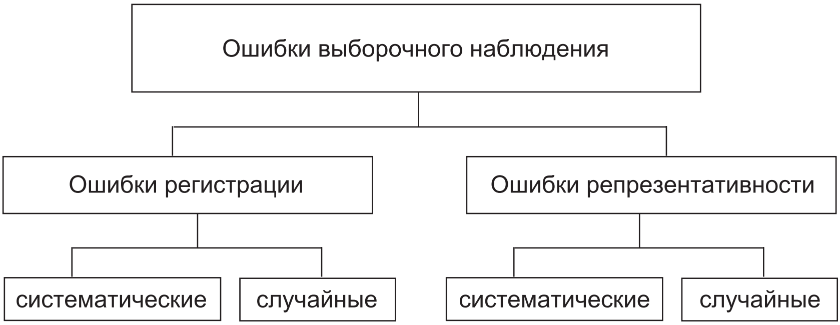

Рис. 8.1. Классификация ошибок выборочного наблюдения

они отражают, в какой степени попавшие в выборку единицы могут представлять всю генеральную совокупность. При этом следует различать систематические и случайные ошибки репрезентативности.

Систематические ошибки репрезентативности связаны с нарушением принципов формирования выборочной совокупности. Например, вследствие каких-либо причин, связанных с организацией отбора, в выборку попали единицы, характеризующиеся несколько большими или, наоборот, несколько меньшими по сравнению с другими единицами значениями наблюдаемых признаков. В этом случае и рассчитанные выборочные характеристики будут завышенными или заниженными.

Случайные ошибки репрезентативности обусловлены действием случайных факторов, не содержащих каких-либо элементов системности в направлении воздействия на рассчитываемые выборочные характеристики. Но даже при строгом соблюдении всех принципов формирования выборочной совокупности выборочные и генеральные характеристики будут несколько различаться. Получаемые случайные ошибки могут быть статистически оценены и учтены при распространении результатов выборочного наблюдения на всю генеральную совокупность. Оценка ошибок выборочного наблюдения основана на теоремах теории вероятностей.

При дальнейшем рассмотрении теории и методов выборочного наблюдения мы будем использовать следующие общепринятые условные обозначения:

N – объем (число единиц) генеральной совокупности;

n – объем (число единиц) выборочной совокупности;

$\overline{x}$ – генеральная средняя, т.е. среднее значение изучаемого признака по генеральной совокупности (средняя прибыль, средняя величина активов, средняя численность работников предприятия и т.п);

$\widetilde{x}$ – выборочная средняя, т.е. среднее значение изучаемого признака по выборочной совокупности;

M – численность единиц генеральной совокупности, обладающих определенным вариантом или вариантами изучаемого признака (численность городского населения, численность сельского населения, количество бракованных изделий, число нерентабельных предприятий и т.п.);

p – генеральная доля, т.е. доля единиц, обладающих определенным вариантом или вариантами изучаемого признака, во всей генеральной совокупности (доля городского населения в общей численоcти населения, доля бракованной продукции в общем выпуске, доля нерентабельных предприятий в общей численности предприятий и т.п.); определяется как M/N;

m – численность единиц выборочной совокупности, обладающих определенным вариантом или вариантами изучаемого признака;

w – выборочная доля, т.е. доля единиц, обладающих определенным вариантом или вариантами изучаемого признака, в выборочной совокупности; определяется как m/n;

μ – средняя ошибка выборки;

Δ – предельная ошибка выборки.

Ошибка выборки, или отклонение выборочной средней от средней генеральной, находится в прямой зависимости от дисперсии изучаемого признака в генеральной совокупности и в обратной зависимости от объема выборки.

Пример. Предположим, изучаемая генеральная совокупность включает всего три единицы, характеризующиеся следующими значениями признака: 4, 6, 8.

Рассчитаем генеральную среднюю и дисперсию данного признака по генеральной совокупности:

$$\overline{x} = \frac{4+6+8}{3} = 6;$$

$$у_{ген}^2 = \frac{\sum(x_i - \bar{x})^2}{N} = \frac{(4-6)^2 + (6-6)^2 + (8-6)^2}{3} = 2,66.$$

Произведем теперь из данной генеральной совокупности все возможные повторные выборки объемом 2 единицы и рассчитаем выборочные средние. Полученные результаты представим в табл. 8.1.

Таблица 8.1

Варианты повторной выборки из генеральной совокупности

Номер выборки	1	2	3	4	5	6	7	8	9
Отобранные значения	4 4	4 6	4 8	6 4	6 6	6 8	8 4	8 6	8 8
Выборочная средняя	4	5	6	5	6	7	6	7	8

Средняя ошибка выборки представляет собой среднее квадратическое отклонение выборочных средних относительно генеральной средней:

$$м = \sqrt{\frac{\sum(\tilde{x}_i - \bar{x})^2}{k}},$$

где k – число всех возможных выборок данного объема из генеральной совокупности.

Определим подкоренное выражение этой формулы, т.е. дисперсию выборочных средних:

$$у_{\tilde{x}}^2 = \frac{\sum(\tilde{x}_i - \bar{x})^2}{k} = \frac{(4-6)^2 + (5-6)^2 + ... + (8-6)^2}{9} = 1,33.$$

Между дисперсией выборочных средних и дисперсией изучаемого признака в генеральной совокупности следующая взаимосвязь:

$$\frac{\sum(\tilde{x}_i - \bar{x})^2}{k} = \frac{\text{у}^2_{\text{ген}}}{n}.$$

Для нашего примера получим:

$$1,33 = \frac{2,66}{2}.$$

Таким образом, среднюю ошибку выборки можно представить как

$$\text{м} = \sqrt{\frac{\text{у}^2_{\text{ген}}}{n}}.$$

При проведении выборочного наблюдения дисперсия изучаемого признака в генеральной совокупности, как правило, неизвестна. В то же время между генеральной дисперсией и средней из всех возможных выборочных дисперсий существует следующее соотношение:

$$\text{у}^2_{\text{ген}} = \overline{\text{у}^2}\,\frac{n}{n-1}.$$

В связи с тем, что на практике в большинстве случаев из генеральной совокупности в определенный момент времени производится только одна выборка, дисперсия изучаемого признака по этой выборке и используется при расчете ошибки. Учитывая, что при достаточно большом объеме выборки отношение n / n –1 близко 1, формула *средней ошибки повторной выборки* принимает следующий вид:

$$\text{м} = \sqrt{\frac{\text{у}^2}{n}},$$

где у^2 – дисперсия изучаемого признака по выборочной совокупности.

При определении возможных границ значений характеристик генеральной совокупности рассчитывается *предельная ошибка выборки*, которая зависит от величины ее средней ошибки и уровня вероятности, с которым гарантируется, что генеральная средняя не выйдет за указанные границы. Согласно теореме А.М. Ляпунова, вероятность той или иной величины предельной ошибки, при достаточно большом объеме выборочной совокупности, подчиняется нормальному закону распределения и может быть определена на основе интеграла Лапласа:

$$P(\Delta \le t\text{м}) = \frac{1}{\sqrt{2\text{р}}} \int_{-t}^{t} e^{-\frac{t^2}{2}} dt = F(t).$$

Значения интеграла Лапласа при различных t приведены в приложении 1. При обобщении результатов выборочного наблюдения наиболее часто используют следующие уровни вероятности и соответствующие им значения t:

P	0,683	0,950	0,954	0,997
t	1	1,96	2	3

Например, если при определении предельной ошибки выборки мы используем $t = 2$, то с вероятностью $P = 0{,}954$ можно утверждать, что расхождение между выборочной и генеральной средними не превысит двухкратной величины рассчитанной средней ошибки выборки.

Расчет ошибок при определении границ генеральной доли, т.е. доли единиц, обладающих тем или иным вариантом изучаемого признака, основан на теореме Бернулли. Согласно этой теореме, вероятность сколь угодно малого расхождения между выборочной долей и генеральной долей при достаточно большом объеме выборки будет стремиться к единице. С учетом того, что вероятность расхождения между выборочной и генеральной долями подчиняется нормальному закону распределения, при определении предельной ошибки выборочной доли также используется функция $F(t)$ при заданном значении t.

В целом процесс подготовки и проведения выборочного наблюдения включает ряд последовательных этапов, представленных на рис.8.2.

В зависимости от состава и структуры генеральной совокупности выбирается *вид выборки*, или способ отбора. К наиболее распространенным на практике видам относятся:

- собственно-случайная (простая случайная) выборка;
- механическая (систематическая) выборка;

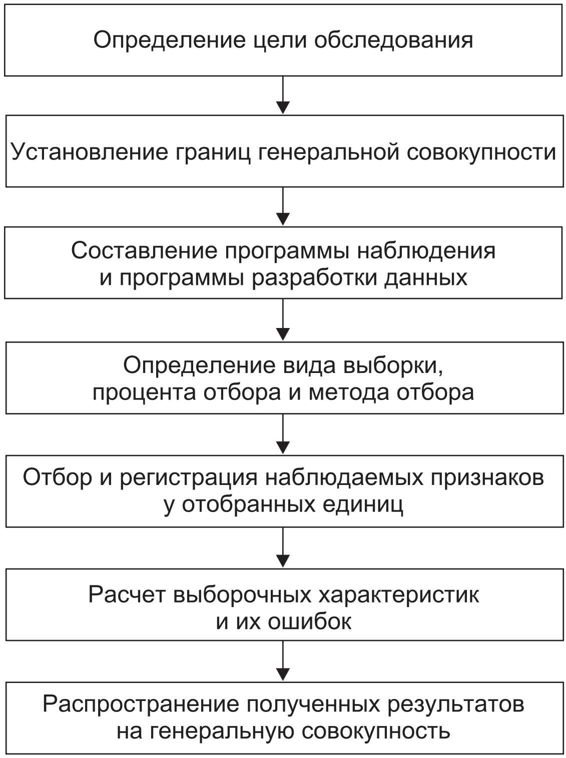

Рис. 8.2. Этапы проведения выборочного наблюдения

- типическая (стратифицированная, расслоенная) выборка;
- серийная (гнездовая) выборка.

Отбор единиц из генеральной совокупности может быть комбинированным, многоступенчатым и многофазным.

Комбинированный отбор предполагает объединение нескольких видов выборки. Так, например, можно комбинировать типическую и серийную, серийную и собственно-случайную выборки. Ошибка такой выборки определяется ступенчатостью отбора.

Многоступенчатым называется отбор, при котором из генеральной совокупности сначала извлекаются укрупненные группы, потом – более мелкие и так до тех пор, пока не будут отобраны те единицы, которые подвергаются обследованию.

В отличие от многоступенчатой *многофазная выборка* предполагает сохранение одной и той же единицы отбора на всех этапах его проведения; при этом отобранные на каждой стадии единицы подвергаются обследованию (программа обследования на каждой последующей стадии отбора расширяется).

Любой вид выборки или их комбинация предполагает использование тех или иных методов непосредственного отбора единиц (групп единиц), основанных на специальных алгоритмах, реализующих принцип случайности. Рассмотрению этих методов и посвящен раздел 8.2.

8.2 МЕТОДЫ (АЛГОРИТМЫ) ОТБОРА ЕДИНИЦ В ВЫБОРОЧНУЮ СОВОКУПНОСТЬ

Процесс формирования выборочной совокупности основан на принципе случайности, реализация которого обеспечивается применением соответствующих методов, или алгоритмов, отбора единиц[1].

В простейшем варианте отбор единиц в выборочную совокупность может быть проведен *методом жеребьевки*. Для этого необходимо располагать достаточным количеством жребиев (фишек, карточек), соответствующих объему генеральной совокупности. Каждый жребий должен содержать информацию об отдельной единице совокупности – номер, название, фамилию лица, адрес или какой-либо другой отличительный признак. Требуемое в соответствии с установленным процентом отбора число жребиев извлекается из общей совокупности в случайном порядке.

Жеребьевка является в большей степени теоретическим методом формирования выборки, так как ее техническая реализация при большом объеме генеральной совокупности затруднительна. Используемые же на практике методы отбора единиц в выборочную совокупность базируются на специальных алгоритмах, реализующих принцип случайности. Рассмотрим некоторые из них.

Метод случайной сортировки включает три шага:

1. Каждой единице генеральной совокупности присваивается случайное число u, полученное с помощью процессора случайных чисел

[1] В данном случае мы не рассматриваем менее распространенные методы неравновероятностного, или направленного, отбора.

в интервале от 0 до 1 (полученные случайные числа должны в той или иной степени соответствовать закону равномерного распределения). Отметим, что генерация случайных чисел может быть произведена в Microsoft Excel (Вставка функции – Математические – Случайное число).

2. Единицы генеральной совокупности ранжируются в соответствии с полученным значением *u*.

3. Отбираются *n* первых единиц.

Достоинства данного метода заключаются в простом алгоритме отбора единиц, а также в возможности формирования нескольких выборок без перекрытия. К недостатку данного метода относят наличие процедуры сортировки единиц генеральной совокупности, которая при достаточно большом ее объеме нежелательна.

Метод прямой реализации предполагает следующую последовательность действий:

1. Все единицы генеральной совокупности, расположенные в случайном порядке или ранжированные по какому-либо признаку, нумеруются от 1 до *N*.

2. С помощью процессора случайных чисел получают *n* значений в интервале от 1 до *N*. Если первоначально случайные числа получены в интервале от 0 до 1, их необходимо умножить на *N* и округлить по правилам до целого значения.

3. Из сформированного списка единиц генеральной совокупности отбираются единицы, соответствующие по номеру полученным случайным числам.

Отметим, что если полученные в п. 2 случайные числа ранжировать, то реализация данного алгоритма потребует только одного считывания файла единиц генеральной совокупности.

Упрощенным вариантом метода прямой реализации является отбор единиц в выборочную совокупность на основе *таблицы случайных чисел* (см. приложение 15). Для проведения отбора могут быть использованы цифры любого столбца данной таблицы, при этом необходимо учитывать объем генеральной совокупности.

Рассмотрим процедуру отбора на основе фрагмента таблицы случайных чисел. Предположим, объем генеральной совокупности составляет 70 000 ед. и требуется сформировать выборку объемом 500 ед.; тогда цифры таблицы следует перегруппировать для получения пятизначных чисел следующим образом:

5489	5	583	31	56	083	5	1988
3522	0	935	78	77	566	5	7020
7555	7	579	25	50	248	7	9477
5759	3	554	50	80	907	4	7001
6303	6	895	33	71	319	6	7231

Для формирования выборки мы должны взять 500 чисел в интервале от 00 001 до 70 000. Таким образом, нам следует из списка единиц генеральной совокупности отобрать единицы под номером 54 895, 35 220, 57 593 и т.д. При этом номера свыше 70 000 (75 557, 93 578 и подобные) будут проигнорированы.

При проведении бесповторного отбора повторяющиеся номера следует учитывать только один раз. При повторном отборе, если тот или иной номер случайно встретится еще один или более раз, соответствующая этому номеру единица в каждом случае повторно включается в выборочную совокупность.

Метод отбора-отказа включает следующие итерации:

- последовательно образуют случайные числа u_1, u_2 ... в соответствии с законом равномерного распределения в интервале от 0 до 1;
- для первой единицы генеральной совокупности проверяется выполнение следующего неравенства:

$$u_1 < \frac{n}{N}. \tag{8.1}$$

  Если данное неравенство выполняется, то первая единица включается в выборку, в противном случае – нет;
- для оставшихся единиц последовательно проверяется выполнение неравенства

$$u_{k+1} < \frac{n - n_k}{N - k}, \tag{8.2}$$

где n_k – число отобранных в выборку единиц среди первых k просмотренных единиц.

Если для (k+1)-й единицы это неравенство выполняется, то данная единица включается в выборку, в противном случае – нет;

- процедура заканчивается, когда $n_k = n$, т. е. когда выборка необходимого объема полностью сформирована. Этот момент вполне может наступить и до завершения полного просмотра всех единиц генеральной совокупности.

Следует отметить, что данный метод основан на алгоритме последовательного извлечения единиц, не требующем ни предварительной сортировки единиц генеральной совокупности или образованных случайных чисел, ни многократного считывания исходного файла.

Рассмотрим на условных примерах, как действует метод отбора-отказа и докажем, что положенный в его основу алгоритм действительно приводит к формированию выборки требуемого объема вне зависимости от значений получаемых случайных чисел.

Пример. Требуется сформировать выборку объемом 100 единиц из генеральной совокупности объемом 1000 единиц, т.е. n = 100 и N = 1000.

Предположим, на 1-м шаге для совокупности A образованное случайное число составило 0,03, тогда неравенство (8.1) выполняется, так как

$$0,03 < \frac{100}{1000},$$

и 1-я единица генеральной совокупности будет включена в совокупность выборочную.

Допустим теперь, что значения последующих случайных чисел по той или иной причине также не превысили 0,1 (см. графу 1 табл. 8.2). Результаты проверки выполнения неравенства (8.2) для соответствующих единиц генеральной совокупности представлены в графах 2 и 3 табл. 8.2.

На 2-м шаге из первых 2 единиц обе были включены в выборку, на 3-м шаге – из первых 3 единиц все три были включены в выборку, и так далее до 100-го шага, на котором из первых 100 единиц генеральной совокупности все 100 вошли в совокупность выборочную. Начиная со 101-го шага, числитель дроби в правой части неравенства (8.2) становится равным нулю, а следовательно, и вся дробь также равна нулю. Тогда, какими бы малыми ни были случайные числа, образованные для оставшихся 900 единиц генеральной совокупности, они не могут быть

Таблица 8.2

Реализация метода отбора-отказа для совокупности *A*

Номер единицы генеральной совокупности	Полученное случайное число	Проверка выполнения условия включения единицы в выборку	Включение единицы в выборку
A	1	2	3
2	0,05	$0,05 < \frac{100-1}{1000-1}$	Включена
3	0,07	$0,07 < \frac{100-2}{1000-2}$	Включена
4	0,04	$0,04 < \frac{100-3}{1000-3}$	Включена
. . .	. . .	. . .	. . .
100	0,001	$0,001 < \frac{100-99}{1000-99}$	Включена
101	0,001	$0,001 > \frac{100-100}{1000-100}$	Не включена
102	0,003	$0,003 > \frac{100-100}{1000-101}$	Не включена
103	0,002	$0,002 > \frac{100-100}{1000-102}$	Не включена
. . .	. . .	. . .	. . .
1000	0,001	$0,001 > \frac{100-100}{1000-999}$	Не включена

меньше нуля, и поэтому ни одна из этих единиц не войдет в выборочную совокупность. Таким образом, в результате реализации данного метода была сформирована требуемая выборка объемом 100 единиц.

Предположим теперь, что для 1-й единицы такой же по объему совокупности *Б* (N = 1000) образованное случайное число составило 0,91. Тогда при формировании выборки n = 100 получим:

$$0,91 > \frac{100}{1000},$$

и, следовательно, 1-я единица генеральной совокупности не будет включена в выборочную совокупность.

Допустим, что последующие полученные случайные числа по той или иной причине имеют относительно большие значения (см. графу 1 табл. 8.3). Соответствующая таким значениям процедура реализации метода отбора-отказа представлена в графах 2 и 3 табл. 8.3.

Таблица 8.3

Реализация метода отбора-отказа для совокупности *Б*

Номер единицы генеральной совокупности	Полученное случайное число	Проверка выполнения условия включения единицы в выборку	Включение единицы в выборку
А	1	2	3
2	0,93	$0,93 > \frac{100-0}{1000-1}$	Не включена
3	0,32	$0,32 > \frac{100-0}{1000-2}$	Не включена
4	0,47	$0,47 > \frac{100-0}{1000-3}$	Не включена
. . .	. . .	. . .	. . .
900	0,994	$0,994 > \frac{100-0}{1000-899}$	Не включена
901	0,992	$0,992 < \frac{100-0}{1000-900} = 1$	Включена

Продолжение

Номер единицы генеральной совокупности	Полученное случайное число	Проверка выполнения условия включения единицы в выборку	Включение единицы в выборку
A	1	2	3
902	0,997	$0,997 < \frac{100-1}{1000-901} = 1$	Включена
903	0,115	$0,115 < \frac{100-2}{1000-902} = 1$	Включена
. . .	. . .	. . .	. . .
1000	0,732	$0,732 < \frac{100-99}{1000-999} = 1$	Включена

Итак, если допустить, что по случайным причинам для первых 900 ед. генеральной совокупности условие включения в выборку не выполнялось, то, начиная с 901-й единицы, ситуация принципиально меняется. Правая часть неравенства (8.2) становится равной 1, и 901-я единица генеральной совокупности включается в выборку при любом значении присвоенного ей случайного числа. Далее на каждом шаге числитель и знаменатель правой части неравенства пропорционально уменьшаются, и поэтому значение дроби в целом остается неизменным, т.е. постоянно равным 1. Вследствие этого, вне зависимости от полученных значений случайных чисел, оставшиеся 99 ед. генеральной совокупности войдут в выборочную совокупность и требуемый объем выборки будет обеспечен.

Мы рассмотрели принцип работы метода отбора-отказа и показали, что при любых условиях положенный в его основу алгоритм приведет к формированию выборки желаемого объема. Наилучшие же результаты, безусловно, будут получены тогда, когда генерируемые случайные числа подчиняются закону равномерного распределения.

Методы отбора единиц в выборочную совокупность используются при различных способах (видах) выборки, которые рассмотрены в последующих разделах.

8.3
СОБСТВЕННО-СЛУЧАЙНАЯ (ПРОСТАЯ СЛУЧАЙНАЯ) ВЫБОРКА

Собственно-случайная выборка заключается в отборе единиц из генеральной совокупности в целом, без разделения ее на группы, подгруппы или серии отдельных единиц. При этом единицы отбираются в случайном порядке, не зависящем ни от последовательности расположения единиц в совокупности, ни от значений их признаков.

Прежде чем производить собственно-случайный отбор, необходимо убедиться, что все без исключения единицы генеральной совокупности имеют абсолютно равные шансы попадания в выборку, в списках или перечне отсутствуют пропуски, нет игнорирования отдельных единиц и т.п. Следует также установить четкие границы генеральной совокупности таким образом, чтобы включение или невключение в нее отдельных единиц не вызывало сомнений. Так, например, при обследовании торговых предприятий необходимо указать, включит ли генеральная совокупность торговые павильоны, коммерческие палатки, передвижные торговые точки и прочие подобные объекты; при обследовании студентов важно определить, будут ли приняты во внимание студенты-заочники, экстерны, учащиеся в магистратуре, лица, находящиеся в академическом отпуске, и т.п.

После проведения отбора с использованием какого-либо алгоритма, реализующего принцип случайности (некоторые из которых были рассмотрены в разделе 8.2), или на основе таблицы случайных чисел, необходимо определить границы генеральных характеристик. Для этого рассчитываются средняя и предельная ошибки выборки.

Средняя ошибка повторной собственно-случайной выборки определяется по формуле

$$м = \frac{у}{\sqrt{n}}. \qquad (8.3)$$

С учетом выбранного уровня вероятности и соответствующего ему значения t предельная ошибка выборки составит:

$$\Delta_{\tilde{x}} = t \cdot \mu_{\tilde{x}}. \qquad (8.4)$$

Тогда можно утверждать, что при заданной вероятности генеральная средняя будет находиться в следующих границах:

$$\tilde{x} - \Delta_{\tilde{x}} \le \bar{x} \le \tilde{x} + \Delta_{\tilde{x}}. \qquad (8.5)$$

Пример. Предположим, в результате выборочного обследования жилищных условий жителей города, осуществленного на основе собственно-случайной повторной выборки, получен следующий ряд распределения (табл. 8.4).

Таблица 8.4

Результаты выборочного обследования жилищных условий жителей города

Общая (полезная) площадь жилищ, приходящаяся на 1 чел., кв. м	До 5,0	5,0–10,0	10,0–15,0	15,0–20,0	20,0–25,0	25,0–30,0	30,0 и более
Число жителей	8	95	204	270	210	130	83

Рассмотрим определение *границ генеральной средней*, в данном случае – средней площади жилищ в расчете на 1 чел. в целом по городу, опираясь только на результаты выборочного обследования. Для определения средней ошибки выборки нам необходимо, прежде всего, рассчитать выборочную среднюю величину и дисперсию изучаемого признака (табл. 8.5).

Таблица 8.5

Расчет средней общей (полезной) площади жилищ, приходящейся на 1 чел., и дисперсии

Общая (полезная) площадь жилищ, приходящаяся на 1 чел., кв. м	Число жителей f	Середина интервала x	xf	x^2f
До 5,0	8	2,5	20,0	50,0
5,0–10,0	95	7,5	712,5	5343,75
10,0–15,0	204	12,5	2550,0	31875,0

Продолжение

Общая (полезная) площадь жилищ, приходящаяся на 1 чел., кв. м	Число жителей f	Середина интервала x	xf	x^2f
15,0–20,0	270	17,5	4725,0	82687,5
20,0–25,0	210	22,5	4725,0	106312,5
25,0–30,0	130	27,5	3575,0	98312,5
30,0 и более	83	32,5	2697,5	87668,75
Итого	1000	–	19005,0	412250,0

$$\tilde{x} = \frac{19005,0}{1000} = 19,0;$$

$$у^2 = \frac{412\,250}{1\,000} - 19,0^2 = 51,25;$$

$$у = \sqrt{51,25} = 7,16.$$

Средняя ошибка выборки составит:

$$м_{\tilde{x}} = \frac{7,16}{\sqrt{1\,000}} = 0,23 \text{ кв. м.}$$

Определим предельную ошибку выборки с вероятностью 0,954 ($t = 2$):

$$\Delta_{\tilde{x}} = 2 \cdot 0,23 = 0,46 \text{ кв. м.}$$

Установим границы генеральной средней:

$$19,0 - 0,46 \le \bar{x} \le 19,0 + 0,46,$$

или

$$18,54 \le \bar{x} \le 19,46.$$

Таким образом, на основании проведенного выборочного обследования с вероятностью 0,954 можно заключить, что средний размер общей площади, приходящейся на 1 чел., в целом по городу лежит в пределах от 18,5 до 19,5 кв. м.

При расчете средней ошибки собственно-случайной бесповторной выборки необходимо учитывать поправку на бесповторность отбора:

$$м_{\tilde{x}} = \sqrt{\frac{у^2}{n}\left(1 - \frac{n}{N}\right)}. \qquad (8.6)$$

Если предположить, что представленные в табл. 8.4 данные являются результатом 5%-го бесповторного отбора (следовательно, генеральная совокупность включает 20 000 ед.), то средняя ошибка выборки будет несколько меньше:

$$м_{\tilde{x}} = \sqrt{\frac{51,25}{1\,000}\left(1 - \frac{1\,000}{20\,000}\right)} = 0,22 \text{ кв. м.}$$

Соответственно уменьшится и предельная ошибка выборки, что вызовет сужение границ генеральной средней. Особенно ощутимо влияние поправки на бесповторность отбора при относительно большом проценте выборки.

Мы рассмотрели определение границ генеральной средней. Рассмотрим теперь, как определяются *границы генеральной доли,* т.е. границы доли единиц, обладающих тем или иным значением признака.

Воспользуемся еще раз данными табл. 8.4 для того, чтобы определить границы доли лиц, обеспеченность жильем которых составляет менее 10 кв. м. Согласно результатам обследования, численность таких лиц составила 103 чел. Определим выборочную долю и дисперсию:

$$w = \frac{103}{1\,000} = 0,103;$$

$$у_w^2 = w(1-w) = 0{,}103 \cdot 0{,}897 = 0{,}0924.$$

Рассчитаем среднюю ошибку выборки:

$$м_w = \sqrt{\frac{0{,}0924}{1\,000}\left(1-\frac{1\,000}{20\,000}\right)} = 0{,}0094.$$

Предельная ошибка выборки с заданной вероятностью составит:

$$\Delta_w = 2 \cdot 0{,}0094 = 0{,}0188 \approx 0{,}019.$$

Определим границы генеральной доли:

$$0{,}103 - 0{,}019 \le p \le 0{,}103 + 0{,}019,$$

или

$$0{,}084 \le p \le 0{,}122.$$

Следовательно, с вероятностью 0,954 можно утверждать, что доля лиц, имеющих менее 10 кв. м жилья на человека, в целом по данному городу находится в пределах от 8,4 до 12,2%.

Мы рассмотрели определение границ генеральной средней и генеральной доли по результатам уже проведенного выборочного наблюдения при известном объеме выборки или проценте отбора. На этапе же проектирования выборочного наблюдения именно объем выборочной совокупности и требует определения.

Чем больше объем выборки, тем меньше значения средней и предельной ошибок выборочного наблюдения и, следовательно, тем уже границы генеральной средней и генеральной доли. В то же время необходимо учитывать, что большой объем выборки приводит к удорожанию обследования, увеличению сроков сбора и обработки материалов, требует привлечения дополнительного персонала и соответствующего материально-технического обеспечения. Поэтому при подготовке выборочного наблюдения необходимо определить тот минимально необходимый объем выборки, который обеспечит требуемую точность полученных статистических характеристик при заданном уровне вероятности.

Представим формулу (8.4) следующим образом:

$$\Delta_{\tilde{x}} = t\frac{y}{\sqrt{n}}. \tag{8.7}$$

Отсюда можно вывести формулу для определения необходимого объема собственно-случайной повторной выборки:

$$n = \frac{t^2 y^2}{\Delta_{\tilde{x}}^2}. \tag{8.8}$$

Полученный на основе использования формулы (8.8) результат всегда округляется в большую сторону. Например, если необходимый объем выборки составляет 493,1 ед., то, обследовав 493 ед., мы не достигнем требуемой точности. Поэтому для достижения желаемого результата обследованием должны быть охвачены 494 ед. С другой стороны, рассчитанное значение необходимого объема выборки свободно может быть увеличено в большую сторону на несколько единиц. Если мы располагаем необходимыми ресурсами, если по причинам организационного порядка (компактность расположения единиц, фиксированная нагрузка на каждого регистратора и т.п.) мы вполне можем охватить больший объем, то включение в выборочную совокупность 500 или, например, 550 ед. только уменьшит значения полученных случайной и предельной ошибок.

Как видно из формулы (8.8), необходимый объем выборки будет тем больше, чем выше заданный уровень вероятности и чем сильнее варьирует наблюдаемый признак. В то же время повышение допустимой предельной ошибки выборки приводит к снижению необходимого ее объема.

Предположим, что в рассмотренном выше примере нас вполне устроят границы генеральной средней (средней площади, приходящейся на 1 чел.) с точностью 0,5 кв. м. Определим необходимый объем выборки:

$$n = \frac{2^2 \cdot 7{,}16^2}{0{,}5^2} = 820{,}2.$$

Мы получили, что для определения с заданной точностью границ средней площади, приходящейся на 1 жителя, достаточно в порядке собственно-случайной повторной выборки обследовать 821 чел.

Расчет необходимого объема выборки предполагает, что организаторы выборочного наблюдения уже на этапе его проектирования располагают по крайней мере косвенными данными о вариации изучаемых признаков. Источниками таких данных могут служить:

- результаты обследования данного объекта в предшествующие периоды;
- результаты обследования аналогичных объектов (жителей других населенных пунктов, предприятий других регионов и т.п.);
- специально проведенное небольшое по объему выборочное обследование данного объекта, ставящее целью лишь изучение вариации наблюдаемых признаков.

При определении необходимого объема выборки для определения границ генеральной доли задача оценки вариации решается значительно проще. Если дисперсия изучаемого альтернативного признака неизвестна, то можно использовать ее максимальное возможное значение:

$$\text{у}^2_{w\max} = w(1-w) = 0,5(1-0,5) = 0,25.$$

Пример. Предприятию связи с вероятностью 0,954 необходимо определить удельный вес телефонных разговоров продолжительностью менее 1 мин. с предельной ошибкой 2%. Сколько разговоров нужно обследовать в порядке собственно-случайного повторного отбора для решения этой задачи?

Для получения ответа на поставленный вопрос воспользуемся формулой (8.8) и будем ориентироваться на максимальную возможную дисперсию доли телефонных разговоров такой продолжительности. Расчет приводит к следующему результату:

$$n = \frac{2^2 \cdot 0,25^2}{0,02^2} = 2\,500.$$

Таким образом, обследованием должны быть охвачены не менее 2 500 разговоров на предмет их продолжительности.

Необходимый объем собственно-случайной бесповторной выборки может быть определен по следующей формуле:

$$n = \frac{t^2 \text{у}^2 N}{t^2 \text{у}^2 + \Delta_{\tilde{x}}^2 N}. \qquad (8.9)$$

Формула (8.9) выводится из формулы (8.4) при условии расчета средней ошибки выборки с поправкой на бесповторность отбора. При определении границ генеральной доли в знаменатель этой формулы подставляется допустимая предельная ошибка доли Δ_w.

Пример. В табл. 8.4 представлены результаты 5%-го бесповторного выборочного обследования (следовательно, объем генеральной совокупности составляет 20 000 чел.). Воспользовавшись значением дисперсии, полученным на основе расчетов, произведенных в табл. 8.5, определим необходимую численность собственно-случайной бесповторной выборки для определения жилищных условий с предельной ошибкой, не превышающей 0,5 кв. м, при уровне вероятности 0,954:

$$n = \frac{2^2 \cdot 51{,}25 \cdot 20\,000}{2^2 \cdot 51{,}25 + 0{,}5^2 \cdot 20\,000} = 787{,}7.$$

Выполненный расчет показал, что для достижения заданной точности необходимо обследовать не менее 788 чел. Как и следовало ожидать, необходимая численность бесповторной выборки оказалась несколько меньше необходимой численности выборки при повторном отборе, полученной выше (821 чел.).

Укажем на одну особенность формулы (8.9). При проведении вычислений объем генеральной совокупности должен быть выражен только в единицах, а не в тысячах или в миллионах единиц. Например, подставив в данную формулу общую численность населения региона, выраженную в тысячах человек, мы не получим правильного значения необходимой численности выборки, также выраженной в тысячах человек, как это иногда бывает в других расчетах. Результат вычислений будет неверен.

8.4
МЕХАНИЧЕСКАЯ (СИСТЕМАТИЧЕСКАЯ) ВЫБОРКА

Механическая выборка может быть применена в тех случаях, когда генеральная совокупность каким-либо образом упорядочена, т.е. имеется определенная последовательность в расположении единиц (табельные номера работников, списки избирателей, телефонные номера респондентов, номера домов и квартир и т.п.). Для проведения отбора желательно, чтобы все единицы также имели порядковые номера от 1 до N.

Для проведения механической выборки устанавливается *пропорция отбора*, которая определяется соотнесением объемов выборочной и генеральной совокупностей. Так, если из совокупности в 500 000 ед. предполагается отобрать 10 000 ед., то пропорция отбора составит 1 / 50 = 1 / (500 000 : 10 000). Отбор единиц осуществляется в соответствии с установленной пропорцией через равные интервалы. Например, при пропорции 1 : 50 (2%-я выборка) отбирается каждая 50-я единица, при пропорции 1 : 20 (5%-я выборка) – каждая 20-я единица и т.д.

Интервал отбора также можно определить как частное от деления 100% на установленный процент отбора. Так, при 2%-м отборе интервал составит 50 (100% : 2%), при 4%-м отборе – 25 (100% : 4%). В тех случаях, когда результат деления получается дробным, сформировать выборку механическим способом при строгом соблюдении процента отбора не представляется возможным. Например, по этой причине нельзя сформировать 3%-ю или 6%-ю выборки.

Генеральную совокупность при механическом отборе можно ранжировать или упорядочить по величине изучаемого или коррелирующего с ним признака, что позволит повысить репрезентативность выборки. Однако в этом случае возрастает опасность *систематической ошибки*, связанной с занижением значений изучаемого признака (если из каждого интервала регистрируется первое значение) или его завышением (если из каждого интервала регистрируется последнее значение). Поэтому целесообразно из каждого интервала отбирать центральную или одну из двух центральных единиц. При этом порядковый номер единицы, с которой начинается отбор, определяется следующим образом. Если интервал отбора обозначить как k, то номер первой отбираемой единицы будет равен $(k + 1) / 2$ при k-нечетном и $k / 2$ или $(k + 2) / 2$ при k-четном. Например, при 5%-й выборке интервал отбора составит 20 единиц, тогда номер единицы, являющейся началом отбора, будет равен 20 : 2 = 10 или (20 + 2) : 2 = 11, т.е. отбор можно начинать с 10-й или с 11-й единицы. В первом случае в выборку попадут 10, 30, 50, 70 и с таким же интервалом последующие единицы; во втором случае попадут единицы с номерами 11, 31, 51, 71 и т.д.

Опасность систематической ошибки при механической выборке также может появиться вследствие случайного совпадения выбранного интервала и циклических закономерностей в расположении единиц генеральной совокупности. Так, при переписи населения 1989 г. в ходе 25%-го выборочного обследования семей была опасность попадания в выборку квартир только одного типа (например, только однокомнатных или только трехкомнатных), так как на лестничных площадках многих типовых домов располагаются именно по 4 квартиры. Чтобы избежать систематической ошибки, в каждом новом подъезде счетчик менял начало отбора.

Для определения *средней ошибки механической выборки*, а также необходимой ее численности используются соответствующие формулы, применяемые при собственно-случайном бесповторном отборе, формулы (8.6) и (8.9). При этом, определив необходимую численность выборки и сопоставив ее с объемом генеральной совокупности, как правило, приходится производить соответствующее округление для получения целочисленного интервала отбора.

Пример. В области зарегистрировано 6000 малых предприятий. Определим, сколько из них нужно отобрать в порядке механического отбора для определения средней численности занятых с ошибкой ±2 чел. (P = 0,997). По результатам ранее проведенного обследования известно, что среднее квадратическое отклонение численности занятых составляет 9 чел. Произведем расчет, воспользовавшись формулой (8.9):

$$n = \frac{3^2 \cdot 9^2 \cdot 6000}{3^2 \cdot 9^2 + 2^2 \cdot 6000} = 176,9 \approx 177.$$

С учетом полученного необходимого объема выборки (177 предприятий) определим интервал отбора: 6 000:177=33,9. Определенный таким способом интервал всегда округляется в меньшую сторону, так как при округлении в большую сторону произведенная выборка не достигнет рассчитанного по формуле необходимого объема. Следовательно, в нашем примере из общего регистра малых предприятий необходимо отбирать каждое 33-е предприятие. При этом процент отбора составит 3,03% (100% : 33).

8.5 ТИПИЧЕСКАЯ (СТРАТИФИЦИРОВАННАЯ) ВЫБОРКА

Типический отбор целесообразно использовать в тех случаях, когда все единицы генеральной совокупности объединены в несколько крупных типических групп. Такие группы также называют *стратами,* или слоями, в связи с чем типический отбор также называют *стратифицированным,* или *расслоенным*. При обследовании населения в качестве типических групп могут быть выбраны области, районы, социальные, возрастные или образовательные группы, при обследовании предприятий – отрасли или подотрасли, формы собственности и т.п.

Рассматривать генеральную совокупность в разрезе нескольких крупных групп единиц имеет смысл только в том случае, если средние значения изучаемых признаков по группам существенно различаются. Например, с большой уверенностью можно предположить, что доходы населения крупного города будут в среднем выше доходов населения, проживающего в сельской местности; численность работников промышленного предприятия в среднем будет выше численности работников торгового или сельскохозяйственного предприятия; средний возраст студентов будет значительно ниже среднего возраста занятого населения, и тем более пенсионеров. В то же время нет никакого смысла при выделении типических групп ориентироваться на признак, не связанный или очень слабо связанный с изучаемым. Например, при изучении доходов населения вряд ли улучшению результатов выборочного обследования будет способствовать деление населения на группы на основе первой буквы фамилии, так как маловероятно, что доходы людей, чья фамилия начинается с букв в интервале от А до К, будут существенно выше или ниже доходов лиц, носящих фамилию, начинающуюся с букв в интервале Л–Я.

Отбор единиц в выборочную совокупность из каждой типической группы осуществляется собственно-случайным или механическим способом. Поскольку в выборочную совокупность в той или иной пропорции обязательно попадают представители всех групп, типизация генеральной совокупности позволяет исключить влияние межгрупповой дисперсии на среднюю ошибку выборки. В то же время в выделенных типических группах обследуются далеко не все единицы, а только включенные в выборку. Следовательно, на величину полученной ошибки будет влиять различие между единицами внутри этих групп, т.е. *внутригрупповая вариация*. Поэтому ошибка типической выборки будет определяться величиной не общей дисперсии, а только ее части – средней из внутригрупповых дисперсий.

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

При *типической выборке, пропорциональной объему типических групп*, число единиц, подлежащих отбору из каждой группы, определяется следующим образом:

$$n_i = n\frac{N_i}{N}, \qquad (8.10)$$

где N_i – объем i-й группы;
n_i – объем выборки из i-й группы.

Пример. Общая численность населения области составляет 1 млн чел., в том числе городского – 600 тыс. чел. и сельского – 400 тыс. чел. Если в ходе выборочного наблюдения планируется обследовать 50 тыс. жителей, то эта численность должна быть поделена пропорционально объему типических групп:

городское население – $n_\text{г} = 50\,000\frac{600\,000}{1\,000\,000} = 30\,000$ чел.;

сельское население – $n_\text{с} = 50\,000\frac{400\,000}{1000\,000} = 20\,000$ чел.

Процесс формирования данной выборки представлен на рис. 8.3.

Средняя ошибка типической выборки определяется по формулам:

$$\text{м} = \sqrt{\frac{\overline{\sigma^2}}{n}} \text{ (повторный отбор)}, \qquad (8.11)$$

$$\text{м} = \sqrt{\frac{\overline{\sigma^2}}{n}\left(1-\frac{n}{N}\right)} \text{ (бесповторный отбор)}, \qquad (8.12)$$

где $\overline{\sigma^2}$ – средняя из внутригрупповых дисперсий.

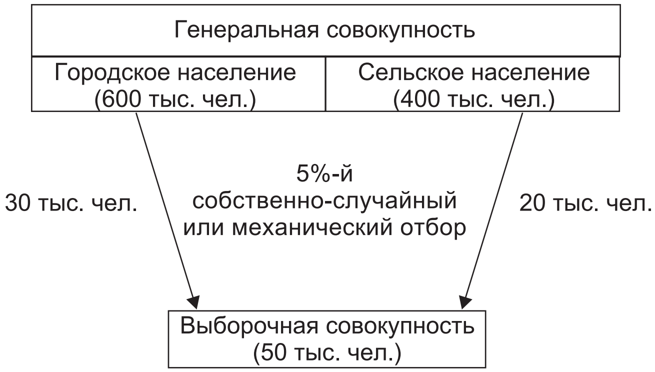

Рис. 8.3. Процесс формирования типической выборки, пропорциональной объему типических групп

Рассмотрим данный вариант типической выборки на условном примере.

Пример. 10%-й бесповторный типический отбор работников предприятия, пропорциональный размерам цехов, проведенный с целью оценки потерь из-за временной нетрудоспособности, привел к следующим результатам (табл. 8.6).

Таблица 8.6

Результаты обследования работников предприятия

Цех	Всего работников, чел	Обследовано, чел.	Число дней временной нетрудоспособности за год	
			средняя	дисперсия
1	1 000	100	18	49
2	1 400	140	12	25
3	800	80	15	16

Рассчитаем среднюю из внутригрупповых дисперсий:

$$\overline{\sigma^2} = \frac{\sum у_i^2 n_i}{\sum n_i} = \frac{49 \cdot 100 + 25 \cdot 140 + 16 \cdot 80}{100 + 140 + 80} = 30,25.$$

Определим среднюю и предельную ошибки выборки (с вероятностью 0,954):

$$м_{\tilde{x}} = \sqrt{\frac{30,25}{320}\left(1 - \frac{320}{3\,200}\right)} = 0,29;$$

$$\Delta_{\tilde{x}} = 2 \cdot 0,29 = 0,58.$$

Рассчитаем выборочную среднюю:

$$\tilde{x} = \frac{\sum x_i n_i}{\sum n_i} = \frac{18 \cdot 100 + 12 \cdot 140 + 15 \cdot 80}{100 + 140 + 80} = 14,6 \text{ дня.}$$

В результате проведенных расчетов с вероятностью 0,954 можно сделать вывод, что среднее число дней временной нетрудоспособности одного работника в целом по предприятию находится в пределах:

$$14,6 - 0,58 \le \bar{x} \le 14,6 + 0,58.$$

При определении *необходимого объема типической выборки* в рассмотренных выше формулах (8.11) и (8.12) общую дисперсию наблюдаемого признака необходимо заменить на среднюю из внутригрупповых дисперсий. Тогда данные формулы примут следующий вид:

$$n = \frac{t^2 \overline{\sigma^2}}{\Delta_{\tilde{x}}^2} \quad \text{(повторный отбор);} \qquad (8.13)$$

$$n = \frac{t^2 \overline{\sigma^2} N}{t^2 \overline{\sigma^2} + \Delta_{\tilde{x}}^2 N} \quad \text{(бесповторный отбор).} \qquad (8.14)$$

Предположим, в рассмотренном выше примере нам необходимо определить среднее число дней временной нетрудоспособности одного работника с предельной ошибкой 0,5 дня. Учитывая величину полученной ранее средней из внутригрупповых дисперсий определим необходимый объем типической выборки при условии бесповторного отбора:

$$n = \frac{2^2 \cdot 30,25 \cdot 3\,200}{2^2 \cdot 30,25 + 0,5^2 \cdot 3\,200} = 420,4.$$

Таким образом мы получили, что при заданных условиях для достижения требуемой точности необходимо обследовать выборочным методом не менее 421 чел. Распределим эту численность на три цеха рассматриваемого предприятия пропорционально их размерам:

$$n_1 = 421 \frac{1\,000}{3\,200} = 131,6;$$

$$n_2 = 421 \frac{1\,400}{3\,200} = 184,2;$$

$$n_3 = 421 \frac{800}{3\,200} = 105,3.$$

Расчеты показывают, что в 1-м цехе необходимо обследовать 132 чел., во 2-м цехе – 184 чел. и в 3-м цехе – 105 чел.

Мы рассмотрели типический отбор, пропорциональный объему типических групп. Второй вариант формирования типической выборки заключается в *отборе единиц, пропорциональном вариации признака* в типических группах. Логика такого отбора заключается в следующем: если внутри какой-либо типической группы наблюдаемый признак варьирует слабо, то для определения границ генеральных характеристик из данной группы достаточно обследовать относительно небольшое число единиц; при сильной же вариации признака объем выборки должен быть соответственно увеличен.

При выборке, пропорциональной вариации признака, число наблюдений по каждой группе рассчитывается по формуле:

$$n_i = n\frac{у_i N_i}{\sum у_i N_i}, \tag{8.15}$$

где σ_i – среднее квадратическое отклонение признака в i-й группе.

Средняя ошибка такого отбора определяется следующим образом:

$$м = \frac{1}{N}\sqrt{\sum\frac{у_i^2 N_i^2}{n_i}} \quad \text{(повторный отбор);} \tag{8.16}$$

$$м = \frac{1}{N}\sqrt{\sum\frac{у_i^2 N_i^2}{n_i}\left(1-\frac{n_i}{N_i}\right)} \quad \text{(бесповторный отбор).} \tag{8.17}$$

Отбор, пропорциональный вариации признака, дает лучшие результаты, однако на практике его применение затруднено из-за трудности получения сведений о вариации до проведения выборочного наблюдения.

Воспользуемся данными, приведенными в табл. 8.6, для иллюстрации этого способа выборочного наблюдения.

Пример. Используя имеющиеся значения внутригрупповых дисперсий, определим необходимый объем выборки по каждому цеху, пропорциональный вариации изучаемого признака, при условии, что в целом выборка составляет 10%, или 320 чел.:

$$\sum у_i N_i = \sqrt{49}\cdot 1\,000 + \sqrt{25}\cdot 1\,400 + \sqrt{16}\cdot 800 = 17\,200;$$

$$n_1 = 320 \cdot \frac{\sqrt{49} \cdot 1\,000}{17\,200} = 130 \text{ чел.};$$

$$n_2 = 320 \cdot \frac{\sqrt{25} \cdot 1400}{17200} = 130 \text{ чел.};$$

$$n_3 = 320 \cdot \frac{\sqrt{16} \cdot 800}{17\,200} = 60 \text{ чел.}$$

С учетом полученных значений рассчитаем среднюю ошибку выборки:

$$\text{м}_{\tilde{x}} = \frac{1}{3\,200}\sqrt{\frac{49 \cdot 1\,000^2}{130}\left(1 - \frac{130}{1000}\right) + \frac{25 \cdot 1400^2}{130}\left(1 - \frac{130}{1\,400}\right) +}$$

$$\overline{+\frac{16 \cdot 800^2}{60}\left(1 - \frac{60}{800}\right)} = 0,28.$$

В данном случае средняя, а следовательно, и предельная ошибки будут несколько меньше, что отразится и на границах генеральной средней.

8.6 СЕРИЙНАЯ ВЫБОРКА

Сущность серийной выборки заключается в собственно-случайном либо механическом отборе групп единиц (серий), внутри которых производится сплошное обследование. *Единицей отбора* при этой выборке является *группа* или *серия*, а не отдельная единица генеральной совокупности, как это имело место в рассматриваемых ранее выборках.

Данный способ отбора удобен в тех случаях, когда единицы генеральной совокупности изначально объединены в небольшие более или менее равновеликие группы или серии. В качестве таких серий могут выступать упаковки с определенным количеством готовой продукции, партии товара, студенческие группы, бригады и другие подобные объединения. Например, в Великобритании серийный отбор используется в обследованиях населения, когда серией являются домохозяйства, объединенные общим почтовым индексом. В случайном порядке производится выборка индексов, и под обследование попадают все домохозяйства, имеющие индекс попавших в выборочную совокупность почтовых отделений.

В отдельных случаях серийная выборка имеет не столько методологические, сколько организационные преимущества перед другими способами формирования выборочной совокупности. Например, Управление маркетинга и регионального развития Московского государственного университета экономики, статистики и информатики периодически проводит обследования школьников Москвы. С организационной точки зрения достаточно сложно опрашивать отдельных учеников из разных классов. Значительно проще из общего списка всех классов всех школ округа сформировать выборку классов, а внутри отобранных классов провести 100%-е обследование учащихся.

В связи с тем что при серийном отборе внутри отобранных групп обследуются все без исключения единицы, внутригрупповая вариация признака не отразится на ошибках выборочного наблюдения. В то же время обследуются не все группы, а только попавшие в выборку. Следовательно, на ошибках получаемых характеристик отразятся различия между группами, которые определяются межгрупповой дисперсией. Поэтому средняя ошибка серийной выборки рассчитывается по формулам

$$\text{м} = \sqrt{\frac{\text{д}^2}{r}} \qquad \text{(повторный отбор);} \qquad (8.18)$$

$$\text{м} = \sqrt{\frac{\text{д}^2}{r}\left(1 - \frac{r}{R}\right)} \qquad \text{(бесповторный отбор),} \qquad (8.19)$$

где r – число отобранных серий;
R – общее число серий.

Межгрупповую дисперсию при равновеликих группах вычисляют следующим образом:

$$д^2 = \frac{\sum(\tilde{x}_i - \tilde{x})^2}{r}, \qquad (8.20)$$

где $\tilde{x}_i$ – средняя i-й серии;
$\tilde{x}$ – общая средняя по всей выборочной совокупности.

Пример. Предположим, партия готовой продукции предприятия упакована в 200 коробок по 50 изделий в каждой. В целях контроля соблюдения параметров технологического процесса проведена 5%-я серийная выборка, в ходе которой отбиралась каждая 20-я коробка. Все изделия, находящиеся в отобранных упаковках, были подвергнуты сплошному обследованию, заключавшемуся в определении их точного веса. Полученные результаты представлены в табл. 8.7.

Таблица 8.7

Результаты выборочного обследования готовой продукции

Номер коробки	1	2	3	4	5	6	7	8	9	10
Средний вес изделия в коробке, г	997	1001	1003	998	1000	1000	998	999	1000	1002

С вероятностью 0,954 требуется определить границы среднего веса изделия во всей партии.

На основе приведенных в таблице внутригрупповых средних определим средний вес изделия по выборочной совокупности:

$$\tilde{x} = \frac{997 + 1\,001 + ... + 1\,002}{10} = 999{,}8 \text{ г.}$$

С учетом полученной средней рассчитаем межгрупповую дисперсию:

$$д^2 = \frac{(997-999,8)^2 + (1\,001-999,8)^2 + ... + (1\,002-999,8)^2}{10} = 3,16.$$

Рассчитаем среднюю и предельную ошибки выборки:

$$м = \sqrt{\frac{3,16}{10}\left(1-\frac{10}{200}\right)} = 0,55 \text{ г};$$

$$\Delta_{\tilde{x}} = 2 \cdot 0,55 = 1,1 \text{ г}.$$

Определим границы генеральной средней:

$$999,8 - 1,1 \le \bar{x} \le 999,8 + 1,1.$$

На основе результатов проведенных расчетов с вероятностью 0,954 можно утверждать, что средний вес изделия в целом по всей партии продукции находится в пределах от 998,7 г до 1000,9 г.

Для определения необходимого объема серийной выборки при заданной предельной ошибке используются следующие формулы:

$$r = \frac{t^2 д^2}{\Delta_{\tilde{x}}^2} \quad \text{(повторный отбор);} \qquad (8.21)$$

$$r = \frac{t^2 д^2 R}{t^2 д^2 + \Delta_{\tilde{x}}^2 R} \quad \text{(бесповторный отбор).} \qquad (8.22)$$

Предположим, в рассмотренном выше примере необходимо определить границы среднего веса изделия с предельной ошибкой ±0,5 г. Используя полученные выше данные о вариации веса, определим, сколько коробок с изделиями нужно обследовать в порядке бесповторной серийной выборки, чтобы получить результат с заданной точностью и при выбранном уровне вероятности:

$$r = \frac{2^2 \cdot 3,16 \cdot 200}{2^2 \cdot 3,16 + 0,8^2 \cdot 200} = 17,97.$$

Выполненный расчет позволяет заключить, что для получения границ генеральной средней с заданной точностью необходимо обследовать не менее 18 коробок с изделиями, отобранных собственно-случайным или механическим способом.

8.7 ПРАКТИКА ПРИМЕНЕНИЯ ВЫБОРОЧНОГО НАБЛЮДЕНИЯ В СОЦИАЛЬНО-ЭКОНОМИЧЕСКИХ ИССЛЕДОВАНИЯХ

В настоящее время выборочное наблюдение находит достаточно широкое применение в обследованиях промышленных и сельскохозяйственных предприятий, изучении цен на потребительском рынке, в обследованиях бюджетов и занятости населения[1]. Выборочный метод является важнейшим источником информации в маркетинговых и социологических исследованиях, в контроле качества продукции; разработаны методологические подходы к применению выборочного наблюдения в аудите. Остановимся на рассмотрении некоторых из указанных областей применения выборки.

При статистическом наблюдении за деятельностью предприятий в качестве основы выборки используются данные, содержащиеся в Едином государственном регистре предприятий и организаций, который ведется Государственным комитетом Российской Федерации по статистике, в регистрах ряда других ведомств, в частности, налоговых органов.

В большей степени методы выборочного наблюдения используются для изучения деятельности *малых предприятий,* при этом наблюдаются как стоимостные экономические показатели (объем производства, продаж, инвестиций), так и показатели деловой активности (оценка текущего состояния предприятия, спрос на продукцию, прогноз на ближайшую перспективу). Обследование проводится ежеквартально по форме федерального государственного статистического наблюдения № ПМ «Сведения об основных показателях деятельности малого предприятия». Объектом наблюдения является совокупность субъектов малого предпринимательства или малых предприятий, классификационные признаки которых определены в законодательном порядке.

---

[1] См.: Методологические положения по статистике. Вып. 3 / Госкомстат России. – М., 2000.

Обследование малых предприятий проводится на основе многомерной типической (расслоенной) выборки, при этом расслоение генеральной совокупности осуществляется по следующим признакам:

- по территории (79 слоев);
- по отраслям (63 слоя);
- по формам собственности (4 слоя);
- по объему выручки (5 слоев).

Общий объем выборочной совокупности не превышает 20% совокупности генеральной.

Одной из проблем при проведении выборочного наблюдения является проблема так называемых неответов, когда попавшая в выборку единица (респондент) по тем или иным причинам не отвечает на часть вопросов или даже на все вопросы, представленные в формуляре. Эта проблема проявляется и при наблюдении за деятельностью малых предприятий.

Вся совокупность неответивших респондентов делится на три группы:

- предприятия, прекратившие или приостановившие свою деятельность;
- предприятия, ведущие финансово-хозяйственную деятельность;
- предприятия, по которым нет объективной информации, функционируют они или нет.

Первая группа предприятий исключается из выборочной совокупности.

Для восстановления данных по предприятиям второй группы применяется *метод перевзвешивания:* неответившему предприятию присваивается значение изучаемого показателя, соответствующее его среднему значению по слою (группе), к которому это предприятие принадлежит.

Для восстановления данных по предприятиям третьей группы используется *метод заполнения случайным подбором* в классах замещения (random hot deck within classes): неответившему предприятию присваиваются значения наблюдаемых признаков, взятых у предприятия-донора. Предприятие-донор выбирается в случайном порядке из предприятий, входящих в соответствующий класс замещения (отрасль экономики).

Одной из конечных задач проведения выборочного наблюдения, в том числе и наблюдения за деятельностью малых предприятий, является получение (оценка) суммарных значений наблюдаемых признаков по всей генеральной совокупности – общей выручки, общего объема производства, общего объема инвестиций и т.д. Для решения этой задачи применяется *метод прямого пересчета.* Сущность этого метода заключается в умножении среднего значения признака, полученного в результате выборочного наблюдения, на объем генеральной совокупности. При выборочном наблюдении за деятельностью малых предприятий методом прямого пересчета получают суммарные значения наблюдаемых признаков по всем выделенным слоям.

Основные задачи выборочного обследования *бюджетов домашних хозяйств* состоят в получении статистических данных о распределении населения по уровню материального благосостояния, данные об уровне бедности и потребления продуктов питания. На основе результатов наблюдения определяют весовые коэффициенты для расчета индекса потребительских цен и получают необходимые данные для составления счетов сектора домашних хозяйств в системе национальных счетов.

Генеральная совокупность объединяет все типы домашних хозяйств, за исключением коллективных (больницы, дома-интернаты, школы-интернаты, монастыри и т.п.).

Обследование проводится в каждом регионе РФ на основе двухступенчатой случайной выборки с использованием процедуры расслоения на каждой из ступеней отбора.

На первой ступени проводится территориальное расслоение населения по месту проживания, т.е. население делится на городское и сельское.

На второй ступени каждый слой делится на несколько подслоев по следующим признакам:

- по размеру домохозяйства, т.е. числу его членов (7 слоев – 1 чел., 2 чел.,…, 7 и более чел.);
- по принадлежности жилого помещения (2 слоя – государственное и частное);
- по типу жилого помещения (3 слоя – отдельная квартира, коммунальная квартира или общежитие, съемное жилье);
- по наличию (отсутствию) в пользовании земельного участка (2 слоя).

При распространении итогов выборочного наблюдения бюджетов домохозяйств на генеральную совокупность применяется *метод взвешивания*, при этом используются данные микропереписи населения 1994 г. о распределении домохозяйств по составу.

Каждому обследованному домохозяйству k-го вида j-го слоя присваивается статистический вес, характеризующий распространение данного вида домохозяйств в генеральной совокупности. Веса определяются по следующей формуле:

$$G_{kj} = \frac{w_{kj} N_j}{n_{kj}},$$

где w_{kj} – удельный вес числа членов домашних хозяйств k-го вида j-го слоя в общей численности населения, обследованного в рамках микропереписи 1994 г.;

N_j – численность населения j-го слоя по данным текущей статистики;

n_{kj} – число членов домашних хозяйств k-го вида j-го слоя по данным выборочного обследования.

Предположим, по данным микропереписи населения удельный вес одиночек в городах данного региона составляет 10% (w_{kj} = 0,1). По данным текущего учета, численность городского населения региона – 600 тыс. чел.(N_j = 600 000). При проведении выборочного обследования в выборку попало 70 домохозяйств, состоящих из 1 человека (n_{kj} = 70). Рассчитаем весовой коэффициент:

$$G_{kj} = \frac{0{,}1 \cdot 600\,000}{70} = 857.$$

Числитель данного коэффициента представляет собой оценку общей численности членов домохозяйств данного вида по состоянию на текущий момент времени. Отношение же числителя к знаменателю показывает, во сколько раз численность членов этих домохозяйств в генеральной совокупности превышает их численность по выборке. Таким образом, при распространении результатов выборочного наблюдения на всю генеральную совокупность значения признаков каждого домохозяйства данного вида необходимо увеличить в 857 раз для получения суммарных значений по слою в целом.

ОСНОВНЫЕ ПОНЯТИЯ

Бесповторный отбор – процесс формирования выборочной совокупности, при котором попавшая в выборку единица в дальнейшей процедуре отбора не участвует.

Выборочная доля – доля единиц в выборочной совокупности, обладающих определенным вариантом или вариантами изучаемого признака.

Выборочная совокупность – совокупность отобранных для обследования единиц.

Выборочная средняя – среднее значение изучаемого признака по выборочной совокупности.

Выборочное наблюдение – несплошное наблюдение, при котором признаки регистрируются у отдельных единиц изучаемой статистической совокупности, отобранных с использованием специальных методов, а полученные в процессе обследования результаты с определенным уровнем вероятности распространяются на всю исходную совокупность.

Генеральная доля – доля единиц в генеральной совокупности, обладающих определенным вариантом или вариантами изучаемого признака.

Генеральная совокупность – исходная изучаемая статистическая совокупность, из которой на основе отбора единиц или групп единиц формируется совокупность выборочная.

Генеральная средняя – среднее значение изучаемого признака по генеральной совокупности.

Метод отбора – алгоритм извлечения единиц или групп единиц из генеральной совокупности, реализующий принцип случайности отбора и лежащий в основе того или иного способа формирования выборочной совокупности (вида выборки).

Объем выборочной совокупности – планируемое или фактическое число единиц генеральной совокупности, отбираемых для регистрации наблюдаемых признаков.

Ошибка репрезентативности – расхождение между статистическими характеристиками выборочной и генеральной совокупностей, обусловленное нарушением принципов формирования выборки или случайными факторами.

Повторный отбор – процесс формирования выборочной совокупности, при котором попавшая в выборку единица продолжает участвовать в дальнейшей процедуре отбора и может быть отобрана в выборочную совокупность повторно.

ТЕСТЫ

1. Какая категория шире:

а) несплошное наблюдение;

б) выборочное наблюдение.

2. Равная вероятность попадания единиц в выборочную совокупность:

а) основной принцип собственно-случайной выборки;

б) основной принцип серийной выборки при случайном отборе серий;

в) основной принцип любой случайной выборки.

3. Какая выборка может быть реализована только на основе бесповторного отбора:

а) собственно-случайная;

б) механическая;

в) типическая;

г) серийная.

4. Между ошибками выборки и объемом выборочной совокупности:

а) существует прямая зависимость;

б) имеет место обратная зависимость;

в) зависимость практически отсутствует.

5. Предельная ошибка выборки при уровне вероятности, превышающем 0,7:

а) меньше средней ошибки выборки;

б) равна средней ошибке выборки;

в) больше средней ошибки выборки.

6. Какой отбор при прочих равных условиях обеспечивает меньшую необходимую численность выборки:

а) повторный;

б) бесповторный.

7. Средняя ошибка типической выборки при обоснованной типизации генеральной совокупности:

а) меньше средней ошибки собственно-случайной выборки;

б) равна средней ошибке собственно-случайной выборки;

в) больше средней ошибки собственно-случайной выборки.

8. Какие единицы обследуются внутри групп при типическом отборе:

а) все единицы;

б) отобранные собственно-случайным способом;

в) отобранные собственно-случайным или механическим способом.

9. Какие единицы обследуются внутри каждой серии при серийном отборе:

а) все единицы;

б) отобранные собственно-случайным способом;

в) отобранные собственно-случайным или механическим способом.

10. Для каких способов формирования выборочной совокупности необходимый объем выборки определяется по одним и тем же формулам:

а) собственно-случайного и механического;
б) собственно-случайного и типического;
в) собственно-случайного и серийного;
г) типического и механического.

ЛИТЕРАТУРА

Гусынин А.Б. Теория выборочных обследований: Учеб. пособие для системы дистанционного образования. – М.: МЭСИ, 1999.

Джессен Р. Методы статистический обследований / Под ред. Е.М. Четыркина; Пер с англ. Ю.П. Лукашина и Я.Ш. Паппе. – М.: Финансы и статистика, 1985.

Елисеева И.И., Терехов А.А. Статистические методы в аудите. – М.: Финансы и статистика, 1998.

Йейтс Ф. Выборочный метод в переписях и обследованиях / Под ред. А.Г. Волкова; Пер. с англ. Е.И. Арона. – М.: Статистика, 1965.

Кокрен У. Методы выборочного исследования / Под ред. А.Г. Волкова; Пер. с англ. И.М. Сонина. – М.: Статистика, 1976.

Методологические положения по статистике. Вып. 3. / Госкомстат России. – М., 2000.

Шварц Г. Выборочный метод / Под ред. И.Г. Венецкого и В.М. Ивановой; Пер. с нем. Я.Ш. Паппе. – М.: Статистика, 1978.

ГЛАВА 9

СТАТИСТИЧЕСКОЕ ИЗУЧЕНИЕ ВЗАИМОСВЯЗИ СОЦИАЛЬНО-ЭКОНОМИЧЕСКИХ ЯВЛЕНИЙ

9.1 ПРИЧИННОСТЬ, РЕГРЕССИЯ, КОРРЕЛЯЦИЯ

Исследование объективно существующих связей между явлениями – важнейшая задача общей теории статистики. В процессе статистического исследования зависимостей вскрываются причинно-следственные отношения между явлениями, что позволяет выявлять факторы (признаки), оказывающие существенное влияние на вариацию изучаемых явлений и процессов. *Причинно-следственные отношения* – это связь явлений и процессов, при которой изменение одного из них – причины – ведет к изменению другого – следствия.

Причина – это совокупность условий, обстоятельств, действие которых приводит к появлению следствия. Если между явлениями действительно существуют причинно-следственные отношения, то эти условия должны обязательно реализовываться вместе с действием причин. Причинные связи носят всеобщий и многообразный характер, и для обнаружения причинно-следственных связей необходимо отбирать отдельные явления и изучать их изолированно.

Особое значение при исследовании причинно-следственных связей имеет выявление временнóй последовательности: причина всегда должна предшествовать следствию, однако не каждое предшествующее событие следует считать причиной, а последующее – следствием.

В реальной социально-экономической действительности причину и следствие необходимо рассматривать как смежные явления, появление которых обусловлено комплексом сопутствующих более простых причин и следствий. Между сложными группами причин и следствий возможны многозначные связи, в которых за одной причиной будет следовать то одно, то другое действие или одно действие будет иметь несколько различных причин. Чтобы установить однозначную причинную связь между явлениями или предсказать возможные следствия конкретной причины, необходима полная абстракция от всех прочих явлений в исследуемой временно́й или пространственной среде. Теоретически такая абстракция воспроизводится. Приемы абстракции часто применяются при изучении взаимосвязей между двумя признаками (парная корреляция). Но чем сложнее изучаемые явления, тем труднее выявить причинно-следственные связи между ними. Взаимное переплетение различных внутренних и внешних факторов неизбежно приводит к некоторым ошибкам в определении причины и следствия.

Особенностью причинно-следственных связей в социально-экономических явлениях является их транзитивность, т.е. причина X и следствие $У$ связаны соотношением $X \rightarrow X' \rightarrow X'' \rightarrow У$, а не непосредственно $X \rightarrow У$. Однако промежуточные факторы, как правило, при анализе опускаются.

Так, например, при использовании показателей международной методологии расчетов фактором валовой прибыли ($У$) считается валовое накопление основных и оборотных фондов (X), но при этом допускаются такие факторы, как валовой выпуск (X'), оплата труда (X'') и т. д. Правильно вскрытые причинно-следственные связи позволяют установить силу воздействия отдельных факторов на результаты хозяйственной деятельности.

Социально-экономические явления представляют собой результат одновременного воздействия большого числа причин. Следовательно, при изучении этих явлений необходимо, абстрагируясь от второстепенных, выявлять главные, основные причины.

На ***первом этапе*** статистического изучения связи осуществляется качественный анализ изучаемого явления методами экономической теории, социологии, конкретной экономики.

На ***втором этапе*** строится модель связи на основе методов статистики: группировок, средних величин, таблиц и т. д.

На ***третьем, последнем этапе*** интерпретируются результаты; анализ вновь связан с качественными особенностями изучаемого явления.

Статистика разработала множество методов изучения связей, выбор которых зависит от целей исследования и поставленных задач. Связи между признаками и явлениями, ввиду их большого разнообразия, классифицируются по ряду оснований. Признаки по значению для изучения взаимосвязи делятся на два класса. Признаки, обуславливающие изменения других, связанных с ними признаков, называются *факторными*, или просто *факторами*. Признаки, изменяющиеся под действием факторных признаков, являются *результативными*. Связи между явлениями и их признаками классифицируются по степени тесноты связи, направлению и аналитическому выражению.

В статистике различают функциональную связь и стохастическую зависимость. *Функциональной* называют такую связь, при которой определенному значению факторного признака соответствует одно и только одно значение результативного признака. Функциональная связь проявляется во всех случаях наблюдения и для каждой конкретной единицы исследуемой совокупности.

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

По степени тесноты связи различают количественные критерии оценки тесноты связи (табл. 9.1).

Таблица 9.1

Количественные критерии оценки тесноты связи

Величина коэффициента корреляции	Характер связи
До $\lvert\pm 0{,}3\rvert$	Практически отсутствует
$\lvert\pm 0{,}3\rvert - \lvert\pm 0{,}5\rvert$	Слабая
$\lvert\pm 0{,}5\rvert - \lvert\pm 0{,}7\rvert$	Σмеренная
$\lvert\pm 0{,}7\rvert - \lvert\pm 1{,}0\rvert$	Сильная

По направлению выделяют связь прямую и обратную. При *прямой связи* с увеличением или уменьшением значений факторного признака происходит увеличение или уменьшение значений результативного. Так, например, рост производительности труда способствует увеличению уровня рентабельности производства. В случае *обратной связи* значения результативного признака изменяются под воздействием факторного, но в противоположном направлении по сравнению с изменением факторного признака. Так, с увеличением уровня фондоотдачи снижается себестоимость единицы производимой продукции.

По аналитическому выражению выделяют связи прямолинейные (или просто линейные) и нелинейные. Если статистическая связь между явлениями может быть приближенно выражена уравнением прямой линии, то ее называют *линейной связью*; если же она выражается уравнением какой-либо кривой линии (параболы, гиперболы, степенной, показательной, экспоненциальной и т. д.), то такую связь называют *нелинейной*, или *криволинейной*.

В статистике не всегда требуются количественные оценки связи, часто важно определить лишь ее направление и характер, выявить форму воздействия одних факторов на другие. Для выявления наличия связи, ее характера и направления в статистике используются методы приведения параллельных данных; аналитических группировок; графический; корреляционный, регрессионный.

Метод приведения параллельных данных основан на сопоставлении двух или нескольких рядов статистических величин. Такое сопоставление позволяет установить наличие связи и получить представление о ее характере. Сравним изменения двух величин *X* и *У*. С увеличением величины X величина У также возрастает. Поэтому связь между ними прямая, и описать ее можно или уравнением прямой, или уравнением параболы второго порядка.

X	1	2	3	4	5	6	7	8	9
У	5	6	9	10	14	17	15	20	23

Взаимосвязь двух признаков изображается *графически* с помощью поля корреляции. В системе координат на *оси абсцисс* откладываются *значения факторного признака*, а на *оси ординат – результативного*. Каждое пересечение линий, проводимых через эти оси, обозначается точкой. При отсутствии тесных связей наблюдается беспорядочное расположение точек на графике. Чем сильнее связь между признаками, тем теснее будут группироваться точки вокруг определенной линии, выражающей форму связи (рис. 9.1).

Для социально-экономических явлений характерно, что наряду с существенными факторами, формирующими уровень результативного признака, на него оказывают воздействие многие другие неучтенные и случайные факторы. Это свидетельствует о том, что взаимосвязи явлений, которые изучает статистика, носят корреляционный характер и аналитически выражаются функцией вида $\bar{y}_x = f(x)$.

Корреляционный метод имеет своей задачей количественное определение тесноты связи между двумя признаками (при парной связи) и между результативным и множеством факторных признаков (при многофакторной связи).

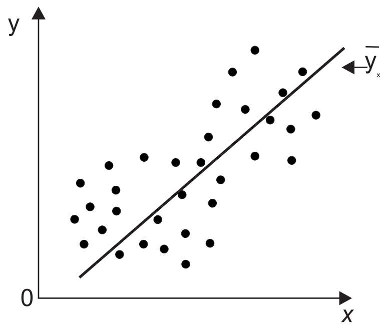

Рис. 9.1. График корреляционного поля

Корреляция – это статистическая зависимость между случайными величинами, не имеющими строго функционального характера, при которой изменение одной из случайных величин приводит к изменению математического ожидания другой.

В статистике различаются следующие варианты зависимостей:

- парная корреляция – связь между двумя признаками (результативным и факторным или двумя факторными);
- частная корреляция – зависимость между результативным и одним факторным признаками при фиксированном значении других факторных признаков;
- множественная корреляция – зависимость результативного и двух или более факторных признаков, включенных в исследование.

Теснота связи количественно выражается величиной коэффициентов корреляции. Коэффициенты корреляции, представляя количественную характеристику тесноты связи между признаками, дают возможность определить «полезность» факторных признаков при построении уравнений множественной регрессии. Величина коэффициента корреляции служит также оценкой соответствия уравнения регрессии выявленным причинно-следственным связям.

Первоначально исследования корреляции проводились в биологии, а позднее распространились и на другие области, в том числе на социально-экономическую. Одновременно с корреляцией начала использоваться и регрессия. Корреляция и регрессия тесно связаны между собой: корреляция оценивает силу (тесноту) статистической связи, регрессия исследует ее форму. Та и другая служат для установления соотношения между явлениями, для определения наличия или отсутствия связи.

Корреляционный и регрессионный анализ как общее понятие включает в себя измерение тесноты, направления связи и установление аналитического выражения (формы) связи (регрессионный анализ).

Регрессионный метод заключается в определении аналитического выражения связи, в котором изменение одной величины (называемой зависимой или результативным признаком) обусловлено влиянием одной или нескольких независимых величин (факторов), а множество всех прочих факторов, также оказывающих влияние на зависимую величину, принимается за постоянные и средние значения. Регрессия может быть однофакторной (парной) и многофакторной (множественной).

По *форме зависимости* различают:

- линейную регрессию, которая выражается уравнением прямой (линейной функцией) вида: $\overline{Y}_x = a_0 + a_1 x$;
- нелинейную регрессию, которая выражается уравнениями вида:

$$\overline{Y}_x = a_0 + a_1 x + a_2 x^2 \text{ – парабола;}$$

$$\overline{Y}_x = a_0 + \frac{a_1}{x} \text{ – гипербола и т.д.}$$

По *направлению связи* различают:

- прямую регрессию (положительную), возникающую при условии, если с увеличением или уменьшением независимой величины значения зависимой также соответственно увеличиваются или уменьшаются;
- обратную (отрицательную) регрессию, появляющуюся при условии, что с увеличением или уменьшением независимой величины зависимая соответственно уменьшается или увеличивается.

Положительную и отрицательную регрессии можно легче понять, если использовать их графическое изображение (рис. 9.2, 9.3).

Для простой (парной) регрессии в условиях, когда достаточно полно установлены причинно-следственные связи, приобретает практический смысл только последнее положение; при множественности причинных связей невозможно четко отграничить одни причинные явления от других.

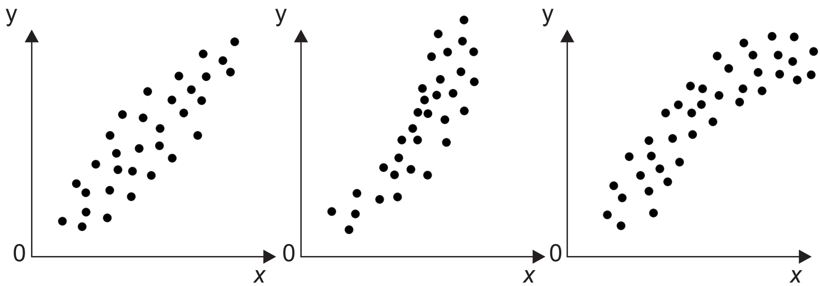

Рис. 9.2. Прямая (положительная) регрессия

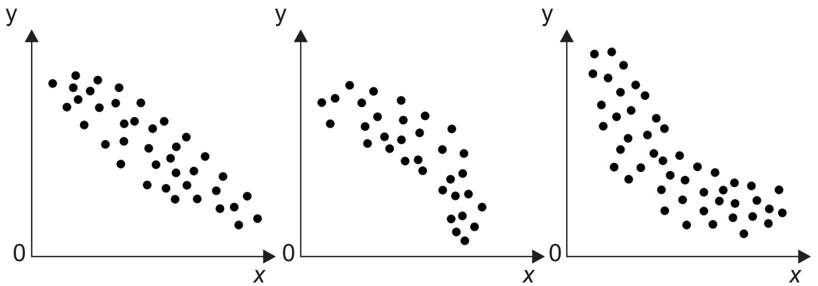

Рис. 9.3. Обратная (отрицательная) регрессия

9.2 ОСНОВНЫЕ ЗАДАЧИ И ПРЕДПОСЫЛКИ ПРИМЕНЕНИЯ КОРРЕЛЯЦИОННО-РЕГРЕССИОННОГО АНАЛИЗА

Все явления и процессы, характеризующие социально-экономическое развитие и составляющие единую систему национальных счетов, тесно взаимосвязаны и взаимозависимы между собой.

В статистике показатели, характеризующие эти явления, могут быть связаны либо корреляционной зависимостью, либо быть независимыми (см. табл. 9.1).

Корреляционная зависимость является частным случаем стохастической зависимости, при которой изменение значений факторных признаков (x_1, x_2, ..., x_k) влечет за собой изменение среднего значения результативного признака.

Корреляционная зависимость исследуется с помощью методов корреляционного и регрессионного анализов.

Корреляционный анализ изучает взаимосвязи показателей и позволяет оценить:

- тесноту связи между показателями с помощью парных, частных и множественных коэффициентов корреляции (раздел 9.6);
- уравнение регрессии.

Основной предпосылкой применения корреляционного анализа является необходимость подчинения совокупности значений всех факторных (x_1, x_2, ..., x_k) и результативного (Y) признаков k-мерному нормальному закону распределения или близость к нему. Если объем исследуемой совокупности достаточно большой ($n > 50$), то нормальность распределения может быть подтверждена на основе расчета и анализа критериев Пирсона, Ястремского, Боярского, Колмогорова, чисел Вастергарда и т. д. Если $n < 50$, то закон распределения исходных данных определяется на базе построения и визуального анализа поля корреляции. При этом если в расположении точек наблюдается линейная тенденция, то можно предположить, что совокупность исходных данных (Y, x_1, x_2, ..., x_k) подчиняется нормальному распределению.

Целью регрессионного анализа является оценка функциональной зависимости условного среднего значения результативного признака (Y) от факторных (x_1, x_2, ..., x_k).

Основной предпосылкой регрессионного анализа является то, что только результативный признак (Y) подчиняется нормальному закону распределения, а факторные признаки x_1, x_2, ..., x_k могут иметь произвольный закон распределения. В анализе динамических рядов в качестве факторного признака выступает время t. При этом в регрессионном анализе заранее подразумевается наличие причинно-следственных связей между результативным (Y) и факторными (x_1, x_2, ..., x_k) признаками.

Уравнение регрессии, или статистическая модель связи социально-экономических явлений, выражаемая функцией

$$\bar{Y}_x = f(x_1, x_2, ..., x_k),$$

является достаточно адекватной реальному моделируемому явлению или процессу, если выполняются следующие *требования к их построению*:

- совокупность исследуемых исходных данных должна быть однородной и математически описываться непрерывными функциями;
- моделируемые явления должны описываться одним или несколькими уравнениями причинно-следственных связей;
- все факторные признаки должны иметь количественное (цифровое) выражение;
- объем исследуемой выборочной совокупности должен быть достаточно большим;
- причинно-следственные связи между явлениями и процессами должны описываться линейной или приводимой к линейной формами зависимости;
- параметры модели связи не должны иметь количественных ограничений;
- территориальная и временная структура изучаемой совокупности должна быть постоянной.

Соблюдение данных требований позволяет исследователю построить статистическую модель связи, наилучшим образом аппроксимирующую моделируемые социально-экономические явления и процессы.

Теоретическая обоснованность моделей взаимосвязи, построенных на основе корреляционно-регрессионного анализа, обеспечивается соблюдением следующих *основных условий.*

- все признаки и их совместные распределения должны подчиняться нормальному закону распределения;
- дисперсия моделируемого признака (*У*) должна все время оставаться постоянной при изменении величины (*У*) и значений факторных признаков;
- отдельные наблюдения должны быть независимыми, т.е. результаты, полученные в i-м наблюдении, не должны быть связаны с предыдущими и содержать информацию о последующих наблюдениях, а также влиять на них.

Отступление от выполнения этих условий и предпосылок приводит к тому, что модель регрессии будет неадекватно отражать реально существующие связи между анализируемыми признаками.

Одной из проблем построения уравнения регрессии является ее *размерность***,** т.е. определение числа факторных признаков, включаемых в модель. Их число должно быть оптимальным.

Сокращение размерности за счет исключения второстепенных, несущественных факторов позволяет получить модель, реализуемую быстрее и качественнее. В то же время построение модели малой размерности может привести к тому, что она будет недостаточно полно описывать исследуемое явление или процесс в единой системе национального счетоводства.

Практика выработала определенный критерий, позволяющий установить оптимальное соотношение между числом факторных признаков, включаемых в модель, и объемом исследуемой совокупности. Согласно данному критерию, число факторных признаков (k) должно быть в 5–6 раз меньше объема изучаемой совокупности.

Общая блок-схема реализации корреляционного и регрессионного методов анализа представлена на рис. 9.4.

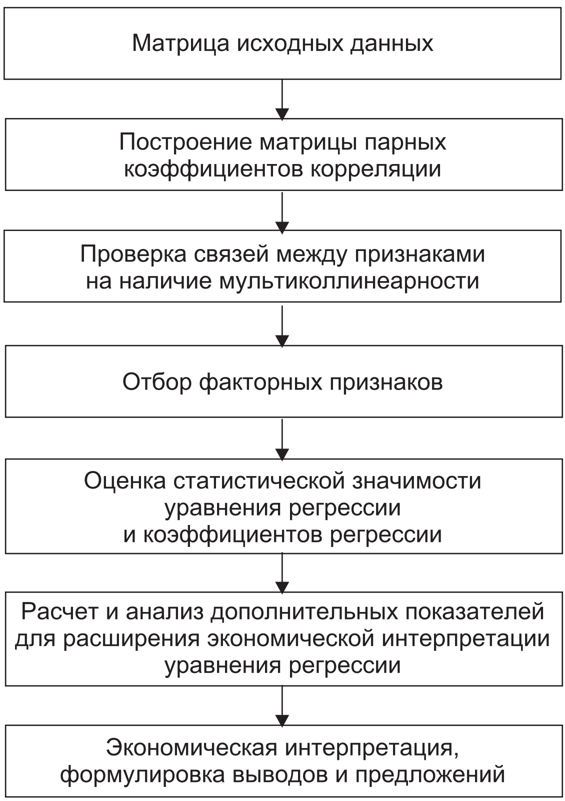

Рис. 9.4. Схема проведения корреляционно-регрессионного анализа

Приведенная последовательность реализации корреляционного и регрессионного методов анализа позволяет достаточно полно охарактеризовать и смоделировать реально существующие взаимосвязи и взаимозависимости между показателями, характеризующими развитие социально-экономических явлений и процессов.

Построение корреляционно-регрессионных моделей, какими бы сложными они ни были, само по себе не вскрывает полностью всех причинно-следственных связей. Основой их адекватности является предварительный качественный анализ, основанный на учете специфики и особенностей исследуемых социально-экономических явлений и процессов.

9.3 ПАРНАЯ РЕГРЕССИЯ НА ОСНОВЕ МЕТОДА НАИМЕНЬШИХ КВАДРАТОВ И МЕТОДА ГРУППИРОВОК

Парная регрессия характеризует связь между двумя признаками – результативным и факторным. Аналитическая связь между ними описывается следующими уравнениями:

- прямой – $\overline{Y}_x = a_0 + a_1 x$;
- гиперболы – $\overline{Y}_x = a_0 + a_1 / x$; (9.1)
- параболы – $\overline{Y}_x = a_0 + a_1 x + a_2 x^2$ и т.д.

Определить тип уравнения можно, исследуя зависимость графически. Однако существуют более общие указания, позволяющие выявить уравнение связи, не прибегая к графическому изображению. Если результативный и факторный признаки возрастают одинаково, примерно в арифметической прогрессии, то это свидетельствует о том, что связь между ними линейная, а при обратной связи – гиперболическая. Если факторный признак увеличивается в арифметической прогрессии, а результативный – значительно быстрее, то используется связь параболическая или степенная.

Оценка параметров уравнений регрессии (a_0, a_1 и a_2 в уравнении параболы второго порядка) осуществляется методом наименьших квадратов, в основе которого лежит предположение о независимости наблюдений исследуемой совокупности.

Основной принцип *метода наименьших квадратов* рассмотрим на следующем примере: будем считать, что две величины (два показателя) X и Y взаимосвязаны между собой, причем Y находится в некоторой зависимости от X. Следовательно, Y будет зависимой, а X – независимой величинами.

Сущность метода наименьших квадратов заключается в нахождении параметров модели (a_0, a_1), при которых минимизируется сумма квадратов отклонений эмпирических (фактических) значений результативного признака от теоретических, полученных по выбранному уравнению регрессии:

$$S = \sum (Y - \overline{Y}_x)^2 \to \min.$$

Для прямой зависимости:

$$S = \sum (y - a_0 - a_1 x)^2 \to \min.$$

Рассмотрим S в качестве функции параметров a_0 и a_1, проведем математические преобразования (дифференцирование) и получим:

$$\begin{cases} \dfrac{dS}{da_0} = \sum 2(a_0 + a_1 x - y) = 0; \\ \dfrac{dS}{da_1} = \sum 2(a_0 + a_1 x - y)x = 0. \end{cases}$$

Откуда система нормальных уравнений для нахождения параметров линейной парной регрессии методом наименьших квадратов примет следующий вид:

$$\begin{cases} na_0 + a_1 \sum x = \sum y; \\ a_0 \sum x + a_1 \sum x^2 = \sum xy, \end{cases} \qquad (9.2)$$

где n – объем исследуемой совокупности (число единиц наблюдений).

В уравнениях регрессии параметр a_0 показывает усредненное влияние на результативный признак неучтенных (не выделенных для исследования) факторов; параметр a_1 (а в уравнении параболы и a_2) – коэффициент регрессии показывает, насколько изменяется в среднем значение результативного признака при увеличении факторного на единицу собственного измерения.

Пример. Имеются следующие данные о показателях, характеризующих деятельность 10 аудиторско-консультационных фирм Москвы в 2001 г. (табл. 9.2).

Предположим наличие линейной зависимости между рассматриваемыми признаками.

Таблица 9.2

Расчет сумм для определения параметров парного линейного уравнения регрессии*

№ п/п	Общая численность профессионалов, чел. X	Совокупная выручка, млн руб. Y	x_2	xy	$\overline{Y}_x$
1	23	2,62	529	60,26	2,661
2	32	3,04	1 024	97,28	2,967
3	50	3,15	2 500	157,50	3,579
4	53	3,83	2 809	202,99	3,681
5	55	3,58	3 025	196,90	3,749
6	58	4,08	3 364	236,64	3,851
7	59	4,09	3 481	241,31	3,885
8	62	4,20	3 844	260,40	3,987
9	69	4,18	4 761	288,42	4,225
10	75	4,24	5 625	318,00	4,429
Итого	536	37,01	30 962	2 059,7	37,010

* Данные условные.

Система нормальных уравнений для данного примера имеет вид

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

$$\begin{cases} 10a_0 + 536a_1 \qquad = 37,01; \\ 536a_0 + 30\,962a = 2059,70. \end{cases}$$

Отсюда: $a_0 = 1,879$; $a_1 = 0,034$.

Следовательно, $\overline{Y}_x = 1,879 + 0,034x$. Таким образом, при увеличении числа профессионалов фирмы на одного человека ее совокупная выручка увеличится в среднем на 34 тыс. руб.

На практике часто исследования проводятся по большому числу наблюдений. В этом случае исходные данные удобнее представлять в сводной корреляционной таблице. При этом анализу подвергаются сгруппированные данные и по факторному X, и по результативному Y признакам, т.е. уравнение парной регрессии целесообразно строить на основе *сгруппированных данных***.**

Если значения признаков X и Y заданы в определенных интервалах ($a - в$), то для каждого интервала сначала определяют середину интервала $\left(\frac{x'}{y'} = \frac{a+b}{2} \right)$, а затем уже коррелируют значения x' и y' и строят уравнения регрессии между ними.

Пример. Определим зависимость между величиной капитала и объемом вложений в ценные бумаги коммерческих банков РФ на 01.01.2002 г. (табл. 9.3).

Система нормальных уравнений для определения коэффицитов уравнения регрессии примет вид:

$$\begin{cases} na_0 + a_1 \sum xf_x = \sum yf_x; \\ a_0 \sum xf_x + a_1 \sum x^2 f_x = \sum xyf_y, \end{cases} \qquad (9.3)$$

где $n = 30$ – число анализируемых коммерческих банков;
$f_x; f_y$ – число банков согласно распределению соответственно по факторному и результативному признакам;
yf_y ; xf_x – значение результативного и факторного признаков по конкретной группе коммерческих банков.

Таблица 9.3

Распределение коммерческих банков РФ по величине капитала и объему вложений в ценные бумаги на 01.01.2002 г. (данные условные)

Группы коммерческих банков по величине капитала, млн руб., Y	Группы коммерческих банков по объему вложений в ценные бумаги, млн руб., X					Число коммерческих банков, f_y	yf_y	xyf_y
	x_i / y_x	0–528	52,8–105,6	105,6–158,4	158,4–211,2			
		26,4	79,2	132,0	184,8			
0,1 – 91,8	45,95	5	1	2	–	8	367,6	21835,44
91,8 – 183,5	137,65	3	5	1	1	10	1376,5	109018,80
183,5 – 275,2	229,35	–	3	2	1	6	1376,1	157425,84
275,2 – 366,9	321,05	–	1	2	3	6	1926,3	288174,48
Число коммерческих банков, f_y	–	8	10	7	5	30	5046,5	576454,56
xf_x	–	211,2	792,0	924,0	924,0	2851,2	–	–
x^2f_x	–	5575,68	62726,4	121968,0	170755,2	361025,28	–	–

Так, для первой группы:

$$yf_y = 45{,}95 \cdot 8 = 367{,}6;$$

$$xf_x = 26{,}4 \cdot 8 = 211{,}2;$$

$$xyf_y = 45{,}95 \cdot 26{,}4 \cdot 5 + 45{,}95 \cdot 79{,}2 \cdot 1 + 45{,}95 \cdot 132{,}0 \cdot 2 = 21835{,}44;$$

$$x^2f_x = 26{,}4 \cdot 26{,}4 \cdot 8 = 5575{,}68.$$

Таким образом, подставив в систему суммарные значения, получим:

$$\begin{cases} 30a_0 + 2851{,}201a_1 = 5046{,}50 \\ 2851{,}2a_0 + 361025{,}28a_1 = 576454{,}56. \end{cases}$$

$$a_0 = 66{,}049; \ \ a_1 = 1{,}075.$$

Отсюда: $Y_x = 66{,}049 + 1{,}075x$.

Если связь между признаками y и x криволинейная и описывается уравнением параболы второго порядка, то

$$\overline{Y}_x = a_0 + a_1x + a_2x^2.$$

В данном случае задача сводится к определению неизвестных параметров: a_0, a_1, a_2.

Значения величин x и y представлены двумя рядами данных:

$$\begin{matrix} y_1 & y_2 & y_3 & \cdots & y_n \\ x_1 & x_2 & x_3 & \ldots & x_n \end{matrix}$$

Если бы все значения, полученные по данным наблюдения, лежали строго на кривой, описываемой уравнением параболы, или для каждой из точек было бы справедливо равенство:

$$Y_i - a_0 - a_1x_{1i} - a_2x_{1i}{}^2 = 0,$$

то не существовало бы никаких проблем. Однако на практике имеем другое:

$$Y_i - a_0 - a_1 x_{1i} - a_2 x_{1i}{}^2 = \Delta_i,$$

где Δ_i – разность между данными наблюдения и данными, полученными по уравнению связи.

Эта разность как раз и появляется из-за ошибок упрощения, поэтому возникает проблема нахождения таких коэффициентов уравнения (регрессии), при которых ошибка была бы минимальной. Можно минимизировать сумму абсолютных отклонений (ошибок), т.е.

$$S = \sum_{i=1}^{n} \left| \Delta_i \right| \to \min,$$

или минимизировать сумму кубических ошибок, и тогда получим метод наименьших кубов:

$$S = \sum_{i=1}^{n} \left| \Delta_i^3 \right| \to \min,$$

или, наконец, минимизировать наибольшую абсолютную ошибку:

$$\min \max_i \left| \Delta_i \right|.$$

Однако наиболее оптимальным вариантом является оценка ошибки по методу наименьших квадратов:

$$S = \sum_{i=1}^{n} \Delta_i^2 \to \min.$$

Метод наименьших квадратов обладает тем замечательным свойством, что делает число нормальных уравнений равным числу неизвестных коэффициентов. Приведенное уравнение параболы второго порядка имеет три неизвестных коэффициента: a_0, a_1, a_2.

Следовательно, применяя метод наименьших квадратов, мы получим уравнение:

$$S=\sum_{i=1}^{n}\Delta_i^2=\sum_{i=1}^{n}(y-a_0-a_1x-a_2x^2)^2\to\min.$$

Для нахождения значений неизвестных коэффициентов a_0, a_1, a_2, при которых функция $f(a_0, a_1, a_2)$ была бы минимальной, необходимо приравнять частные производные по этим величинам нулю, т.е.

$$\begin{cases}\dfrac{dS}{da_0}=2\sum(y-a_0-a_1x-a_2x^2)=0;\\ \dfrac{dS}{da_1}=2\sum(y-a_0-a_1x-a_2x^2)\cdot x=0;\\ \dfrac{dS}{da_2}=2\sum(y-a_0-a_1x-a_2x^2)\cdot x^2=0.\end{cases}$$

Проделав простейшие преобразования, получим систему нормальных уравнений:

$$\begin{cases}na_0+a_1\sum x+a_2\sum x^2=\sum y;\\ a_0\sum x+a_1\sum x^2+a_2\sum x^3=\sum yx;\\ a_0\sum x^2+a_1\sum x^3+a_2\sum x^4=\sum yx^2.\end{cases}\qquad(9.4)$$

Решив систему, найдем значения неизвестных коэффициентов a_0, a_1, a_2 и получим уравнение регрессии. Вычислим по уравнению регрессии теоретические значения $\overline{y}_x$ и сравним с данными наблюдения, т.е. рассчитаем так называемую остаточную сумму квадратов (табл. 9.4).

Остаточная сумма квадратов совпадает с минимальной возможной величиной по методу наименьших квадратов.

Оценка обратной зависимости между x и y, когда с увеличением (уменьшением) x уменьшается (увеличивается) результативный признак y, может быть осуществлена на основе уравнения гиперболы:

Таблица 9.4

Расчет остаточной суммы квадратов

Номер наблюдения	Значения по данным		$\Delta_i = y_i - \bar{y}_x$	Δ_i^2
	наблюдения	уравнения регрессии $\bar{y}_x$		
1	y_1	$\bar{y}_1$	Д$_1$	Δ_1^2
2	y_2	$\bar{y}_2$	Д$_2$	Δ_2^2
3	y_3	$\bar{y}_3$	Д$_3$	Δ_3^2
.	.	.	.	.
.	.	.	.	.
.				
.	y_i	$\bar{y}_i$	Д$_i$	Δ_i^{2*}
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
n	y_n	$\bar{y}_n$	Д$_n$	Δ_n^2

* $\sum\limits_{i=1}^{n} \Delta_i^2 \, (i = 1, 2, ..., n)$.

$$\bar{Y}_x = a_0 + \frac{a_1}{x}.$$

Систему нормальных уравнений для нахождения параметров гиперболы можно представить следующим образом:

$$\begin{cases} na_0 + a_1 \sum \frac{1}{x} = \sum y; \\ a_0 \sum \frac{1}{x} + a_1 \sum \frac{1}{x^2} = \sum \frac{y}{x}. \end{cases} \quad (9.5)$$

Применение метода наименьших квадратов объясняется неизбежным наличием случайных ошибок в результатах опыта.

Статистические данные обладают ошибками упрощения, которые возникают как следствие:

- неполноты охвата, потому что часть единиц совокупности, полученных в результате наблюдения, не может быть использована в исследовании;
- неполноты факторов, определяющих то или иное социально-экономическое явление, в силу того, что ни в одно уравнение, или модель, нельзя включить бесконечное число аргументов (во всех случаях отбирается только часть воздействующих факторов, причем отбор носит чисто субъективный характер);
- характера выбранного уравнения связи. Как бы хорошо оно ни было обосновано, как бы теоретически адекватно ни описывало исследуемое явление, оно не может быть его точным аналогом.

Решение вопроса о возможности использования метода наименьших квадратов для изучения связей между социально-экономическими явлениями зависит от свойства оценок, получаемых с помощью этого метода.

Даже при сравнительно небольшом числе наблюдений применение метода наименьших квадратов позволяет получить достоверные оценки.

Метод наименьших квадратов может быть также использован в анализе косвенных наблюдений, являющихся функциями многих неизвестных.

Обобщенная блок-схема построения уравнения парной регрессии представлена на рис. 9.5.

9.4 МНОЖЕСТВЕННАЯ (МНОГОФАКТОРНАЯ) РЕГРЕССИЯ

Изучение связи между тремя и более связанными между собой признаками носит название *множественной (многофакторной) регрессии.* При исследовании зависимостей методами множественной регрессии задача формулируется так же, как и при использовании

Рис. 9.5 . Схема построения уравнения парной регрессии

парной регрессии, т.е. требуется определить аналитическое выражение связи между результативным признаком (Y) и факторными признаками (x_1, x_2, x_3, ... x_k), найти функцию:

$$\overline{Y}_{1,2,\dots,k} = f(x_1, x_2, \dots, x_k)\,. \tag{9.6}$$

Построение моделей множественной регрессии включает несколько этапов:

- выбор формы связи (уравнения регрессии);
- отбор факторных признаков;

- обеспечение достаточного объема совокупности для получения несмещенных оценок.

Рассмотрим каждый из них.

Выбор формы связи затрудняется тем, что с использованием математического аппарата теоретически зависимость между признаками выражается большим числом различных функций.

Выбор типа уравнения осложнен тем, что для любой формы зависимости выбирается целый ряд уравнений, которые в определенной степени будут описывать эти связи. Некоторые предпосылки для выбора определенного уравнения регрессии получают на основе анализа предшествующих аналогичных исследований или на базе анализа подобных работ в смежных отраслях знаний. Поскольку уравнение регрессии строится главным образом для объяснения и количественного выражения взаимосвязей, оно должно хорошо отражать сложившиеся между исследуемыми факторами фактические связи.

Наиболее приемлемым способом определения вида исходного уравнения регрессии является метод *перебора различных уравнений.*

Сущность данного метода заключается в том, что большое число уравнений (моделей) регрессии, отобранных для описания связей какого-либо социально-экономического явления или процесса, реализуется на ЭВМ с помощью специально разработанного алгоритма перебора с последующей статистической проверкой, главным образом, на основе t-критерия Стьюдента и F-критерия Фишера–Снедекора.

Способ перебора является достаточно трудоемким и связан с большим объемом вычислительных работ.

Практика построения многофакторных моделей взаимосвязи показывает, что все реально существующие зависимости между социально-экономическими явлениями можно описать, используя *пять типов моделей*:

1) линейную:

$$\bar{Y}_{1,2,\ldots,k} = a_0 + a_1 x_1 + a_2 x_2 + \ldots + a_k x_k;$$

2) степенную:

$$\bar{Y}_{1,2,\ldots,k} = a_0 x_1^{a_1} \cdot x_2^{a_2}, \ldots, \ x_k^{a_k};$$

3) показательную:

$$\overline{Y}_{1,2,...,k} = e^{a_0 + a_1 x_1 + a_2 x_2 + ... + a_k x_k};$$

4) параболическую:

$$\overline{Y}_{1,2,...,k} = a_0 + a_1 x_1^2 + a_2 x_2^2 + ... + a_k x_k^2;$$

5) гиперболическую:

$$\overline{Y}_{1,2,...,k} = a_0 + \frac{a_1}{x_1} + \frac{a_2}{x_2} + ... + \frac{a_k}{x_k}.$$

Основное значение имеют линейные модели в силу простоты и логичности их экономической интерпретации. Нелинейные формы зависимости приводятся к линейным путем линеаризации.

Важным этапом построения уже выбранного уравнения множественной регрессии является отбор и последующее включение факторных признаков.

Сложность формирования уравнения множественной регрессии заключается в том, что почти все факторные признаки находятся в зависимости один от другого.

Определение размерности модели связи, т.е. определение оптимального числа факторных признаков, является одной из основных проблем построения множественного уравнения регрессии. В то же время чем больше факторных признаков включено в уравнение, тем оно лучше описывает явление. Однако модель размерностью 100 и более факторных признаков сложно реализуема и требует больших затрат машинного времени. Сокращение размерности модели за счет исключения второстепенных, экономически и статистически несущественных факторов способствует простоте и качеству ее реализации. Но построение модели регрессии малой размерности может привести к тому, что такая модель будет недостаточно адекватна исследуемым явлениям и процессам.

Проблема отбора факторных признаков для построения моделей взаимосвязи может быть решена на основе эвристических или многомерных статистических методов анализа.

Метод экспертных оценок как эвристический метод анализа основных макроэкономических показателей, формирующих единую международную систему расчетов, основан на интуитивно-логических предпосылках, содержательно-качественном анализе (подробнее данный метод рассмотрен в главе 13). Анализ экспертной информации проводится на базе расчета и анализа непараметрических показателей связи: ранговых коэффициентов корреляции Спирмена, Кендалла и конкордации (раздел 9.8).

Наиболее приемлемым способом отбора факторных признаков является *метод шаговой регрессии* (шаговый регрессионный анализ). Сущность метода шаговой регрессии заключается в последовательном включении факторов в уравнение регрессии и последующей проверке их значимости. Факторы поочередно вводятся в уравнение так называемым прямым методом. При проверке значимости введенного фактора определяется, насколько уменьшается сумма квадратов остатков и растет величина множественного коэффициента корреляции (R). Одновременно используется и обратный метод, т.е. исключение факторов, ставших незначимыми на основе t-критерия Стьюдента. Фактор является *незначимым*, если его включение в уравнение регрессии только изменяет значение коэффициентов регрессии, не уменьшая суммы квадратов остатков и не увеличивая их значения. Если при включении в модель соответствующего факторного признака величина множественного коэффициента корреляции увеличивается, а коэффициент регрессии не изменяется (или меняется несущественно), то данный признак *существенен*, и его включение в уравнение регрессии необходимо.

Если же при включении в модель факторного признака коэффициенты регрессии меняют не только величину, но и знаки, а множественный коэффициент корреляции не возрастает, то данный факторный признак признается *нецелесообразным* для включения в модель связи.

Сложность и взаимное переплетение отдельных факторов, обусловливающих исследуемое экономическое явление (процесс), могут проявляться в так называемой мультиколлинеарности. Под *мультиколлинеарностью* понимается тесная зависимость между факторными признаками, включенными в модель.

Наличие мультиколлинеарности между признаками приводит к следующему:

- искажению величины параметров модели, которые имеют тенденцию к завышению;

- изменению смысла экономической интерпретации коэффициентов регрессии;
- слабой обусловленности системы нормальных уравнений;
- осложнению процесса определения наиболее существенных факторных признаков.

В решении проблемы мультиколлинеарности можно выделить несколько этапов:

- установление наличия мультиколлинеарности;
- определение причин возникновения мультиколлинеарности;
- разработку мер по ее устранению.

Возникновение мультиколлинеарности между признаками вызвано следующими причинами:

- факторные признаки характеризуют одну и ту же сторону явления или процесса: например, показатели объема производимой продукции и среднегодовой стоимости основных фондов одновременно включать в модель не рекомендуется, так как они оба характеризуют размер предприятия;
- в качестве факторных признаков используются показатели, суммарное значение которых представляет собой постоянную величину;
- факторные признаки являются составными элементами друг друга;
- факторные признаки по экономическому смыслу дублируют друг друга.

Одним из индикаторов определения наличия мультиколлинеарности между признаками является превышение парным коэффициентом корреляции величины 0,8 ($r_{x_i x_j}$) и др.

Устранение мультиколлинеарности может реализовываться через исключение из корреляционной модели одного или нескольких линейно-связанных факторных признаков или преобразование исходных факторных признаков в новые, укрупненные факторы. Вопрос о том, какой из факторов следует отбросить, решается на основании качественного и логического анализов изучаемого явления.

Качество уравнения регрессии зависит от степени достоверности и надежности исходных данных и объема совокупности. Исследователь должен стремиться к увеличению числа наблюдений, так как большой объем наблюдений является одной из предпосылок построения адекватных статистических моделей.

Аналитическая форма выражения связи результативного признака и ряда факторных называется *многофакторным* (множественным) *уравнением регрессии,* или моделью *связи.*

Уравнение линейной множественной регрессии имеет вид:

$$\overline{Y}_{1,2,3,...,k} = a_0 + a_1 x_1 + a_2 x_2 + ... a_k x_k, \quad (9.7)$$

где $\overline{Y}_{1,2,3,...k}$ – теоретические значения результативного признака, полученные в результате подстановки соответствующих значений факторных признаков в уравнение регрессии;
$x_1, x_2, ..., x_k$ – факторные признаки;
$a_0, a_1, ..., a_k$ – параметры модели (коэффициенты регрессии).

Параметры уравнения могут быть определены графическим методом, методом наименьших квадратов и т.д.

Методом наименьших квадратов (см. раздел 9.3) минимизируем выражение:

$$S = \sum_{i=1}^{n} (Y - a_0 - a_1 x_1 - a_2 x_2 - ... - a_k x_k)^2 = \min.$$

$$\frac{ds}{da_0} = 0; \frac{ds}{da_1} = 0; \frac{ds}{da_2} = 0; ...; \frac{ds}{da_k} = 0.$$

Например, по параметру a_1:

$$\frac{ds}{da_1} = \sum 2(Y - a_0 - a_1 x_1 - a_2 x_2 - ... - a_k x_k) \cdot (-x_1) = 0.$$

Сделав соответствующие преобразования по всем значениям параметров a_1, получим:

$$-2\sum Yx_1 + 2a_0 \sum x_1 + 2a_1 \sum x_1^2 + 2a_2 \sum x_2 x_1 + ... + 2a_k \sum x_k x_1,$$

отсюда:

$$a_0 \sum x_1 + a_1 \sum x_1^2 + a_2 \sum x_2 x_1 + ... + a_k \sum x_k x_1 = \sum Y_x.$$

В результате таких преобразований система нормальных уравнений с k неизвестными (по числу параметров a_i,) имеет вид

$$\begin{cases} na_0 + a_1\sum x_1 + a_2\sum x_2 + \ldots + a_i\sum x_i + \ldots + a_k\sum x_k = \sum Y; \\ a_0\sum x_1 + a_1\sum x_1^2 \;\; + a_2\sum x_2 x_1 + \ldots + a_i\sum x_i x_1 + \ldots + a_k\sum x_k x_1 = \sum Y_{x_1}; \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ a_0\sum x_k + a_1\sum x_1 x_k \;\; + a_2\sum x_2 x_k + \ldots + a_i\sum x_i x_k + \ldots + a_k\sum x_k^2 = \sum Y_{x_k}. \end{cases} \qquad (9.8)$$

Таким образом, обобщенная блок-схема построения множественного уравнения регрессии может быть представлена в следующем виде (рис. 9.6).

Одним из способов построения множественных уравнений регрессии является построение *модели связи в стандартизованном масштабе.*

Оценка влияния каждого факторного признака, включенного в уравнение регрессии, на результативный признак может быть значительно затруднена, если факторные признаки различны по своей сущности и имеют различные единицы измерения. В этих случаях для более точной оценки влияния факторных признаков на результативный используют множественные модели регрессии в стандартизованном масштабе. Модель регрессии в стандартизованном масштабе предполагает, что все значения исследуемых признаков переводятся в стандарты по формуле

$$t_{x_i} = \frac{x_i - \bar{x}_i}{\text{у}_i}, \qquad (9.9)$$

где x_i – значение признака в натуральном масштабе.

Рис. 9.6. Схема построения множественного уравнения регрессии

Уравнение множественной регрессии в стандартизованном масштабе следующее:

$$t_{1,2,\ldots,k} = \text{в}_1 t_1 + \text{в}_2 t_2 + \ldots + \text{в}_k t_k, \tag{9.10}$$

где $t_1, t_2, t_3, \ldots, t_k$ – стандартизованные значения признаков $x_1, x_2, \ldots, x_k$;
$\overline{t}_{1,2,\ldots k}$ – среднее значение стандартизованной переменной соответствующего результативного признака, полученного по уравнению регрессии;
$\beta_1\ \beta_2, \ldots, \beta_k$ – стандартизованные коэффициенты регрессии.

Параметры многофакторной модели регрессии в стандартизованном масштабе определяются методом наименьших квадратов, рассмотренным выше.

Представим систему нормальных уравнений:

$$\begin{cases} \sum tt_1 = \text{в}_1 \sum t_1^2 + \text{в}_2 \sum t_1 t_2 + \ldots + \text{в}_k \sum t_1 t_k; \\ \sum tt_2 = \text{в}_1 \sum t_1 t_2 + \text{в}_2 \sum t_2^2 + \ldots + \text{в}_k \sum t_2 t_k; \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ \sum tt_k = \text{в}_1 \sum t_1 t_k + \text{в}_2 \sum t_2 t_k + \ldots + \text{в}_k \sum t_k^2, \end{cases}$$

где t – значение результативного признака в стандартизованном масштабе.

Коэффициенты $\beta_1\ \beta_2, \ldots, \beta_k$ дают возможность провести сравнительную оценку силы влияния изменения каждого факторного признака на изменение результативного (моделируемого) признака.

От уравнения в стандартизованном масштабе можно легко перейти к уравнению в натуральном масштабе. Коэффициенты a_i получают из соотношения

$$a_i = \text{в}_i \frac{\text{у}_i}{\text{у}_y}, \tag{9.11}$$

а свободный член a_0 – из выражения

$$a_0 = \overline{Y} - a_1 \overline{x}_1 - a_2 \overline{x}_2 - \ldots - a_k \overline{x}_k.$$

Таблица 9.5

Расчетные данные для определения параметров уравнения регрессии*

№ п/п	Численность работающих, чел., x_1	Стоимость основных фондов, млн руб., x_2	Прибыль, тыс. руб. Y	x_1^2	x_1x_2	yx_1	x_2^2	yx_2	$\overline{Y}_x$
1	77	5,9	1 070	5 929	454,3	82 390	34,81	6313,0	1012,8
2	77	5,9	1 001	5 929	454,3	77 077	34,81	5905,9	1012,8
3	81	4,9	789	6 561	396,3	63 909	24,01	3866,1	854,7
4	82	4,3	779	6 724	352,6	63 878	18,49	3349,7	817,8
5	89	3,9	606	7 921	347,1	53 934	15,21	2363,4	530,8
6	96	4,3	221	9 216	412,8	21 216	18,49	950,3	237,1
Итого	502	29,2	4 466	42 280	2418,0	362 404	145,82	22 748,4	4466,0

* Данные условные.

Пример. По следующим данным о прибыли (y), численности работающих (x_1) и стоимости основных фондов (x_2) АОЗТ «Скат» определим зависимость между признаками (табл. 9.5).

Система нормальных уравнений имеет вид

$$\begin{cases} na_0 + a_1\sum x_1 + a_2\sum x_2 = \sum Y; \\ a_0\sum x_1 + a_1\sum x_1^2 + a_2\sum x_1x_2 = \sum x_1Y; \\ a_0\sum x_2 + a_1\sum x_1x_2 + a_2\sum x_2^2 = \sum x_2Y. \end{cases}$$

$$\begin{cases} 6a_0 + 502a_1 + 29,2a_2 = 4466; \\ 502a_0 + 42280a_1 + 2418a_2 = 362404; \\ 29,2a_0 + 2418a_1 + 145,82a_2 = 22748,4. \end{cases}$$

Таким образом,

$$\overline{Y}_x = 4247,79 - 41,43x_1 - 7,60x_2.$$

9.5 ОЦЕНКА СУЩЕСТВЕННОСТИ СВЯЗИ. ПРИНЯТИЕ РЕШЕНИЙ НА ОСНОВЕ УРАВНЕНИЯ РЕГРЕССИИ

Проверка *адекватности* моделей, построенных на основе уравнений регрессии, начинается с проверки значимости каждого коэффициента регрессии.

Значимость коэффициентов регрессии осуществляется с помощью t-критерия Стьюдента:

$$t_p = \frac{|a_i|}{\sqrt{у_{a_i}^2}}, \tag{9.12}$$

где $\sigma_{a_i}^2$ – дисперсия коэффициента регрессии.

Параметр модели признается статистически значимым, если

$$t_p > t_{kp}\ (\alpha;\ n = n - k - 1),$$

где a – уровень значимости критерия проверки гипотезы о равенстве нулю параметров, измеряющих связь, т.е. статистическая существенность связи утверждается при отклонении нулевой гипотезы об отсутствии связи;

$n = n–k–1$ – число степеней свободы, которое характеризует число свободно варьирующих элементов совокупности.

Наиболее сложным в этом выражении является определение дисперсии, которая может быть рассчитана двояким способом.

Наиболее простой способ, выработанный методикой экспериментирования, заключается в том, что величина дисперсии коэффициента регрессии может быть приближенно определена по выражению

$$у_{a_i}^2 = \frac{у_y^2}{k}, \tag{9.13}$$

где σ^2_y – дисперсия результативного признака;

k – число факторных признаков в уравнении.

Более точную оценку величины дисперсии можно получить по формуле:

$$у_{a_i} = \frac{\sigma_y \cdot \sqrt{1-R^2}}{у_{x_i} \cdot \sqrt{n} \cdot \sqrt{1-R_i}}, \tag{9.14}$$

где R_i – величина множественного коэффициента корреляции по фактору x_i с остальными факторами.

Проверка адекватности всей модели осуществляется с помощью расчета F-критерия и величины средней ошибки аппроксимации $\bar{\varepsilon}$.

Если $F_p > F_\alpha$ при $\alpha = 0{,}05$ или $\alpha = 0{,}01$, то H_0 – гипотеза о несоответствии заложенных в уравнении регрессии связей реально существующим отвергается. Величина F_α определяется по специальным таблицам на основании величины $\alpha = 0{,}05$ или $\alpha = 0{,}01$ и числа степеней свободы:

$$v_1 = k–1,\ v_2 = n - k,$$

где n – число наблюдений;
k – число факторных признаков в уравнении.

Значение средней ошибки аппроксимации не должно превышать 12–15%.

$$\bar{\text{e}} = \frac{1}{n}\sum\frac{\left|Y - \bar{Y}_{1,2,\ldots,k}\right|}{y}\cdot 100,$$

где $\bar{\varepsilon}$ – средняя ошибка аппроксимации.

Наиболее сложным этапом, завершающим регрессионный анализ, является *интерпретация* уравнения, т.е. перевод его с языка статистики и математики на язык экономики.

Интерпретация моделей регрессии (рис. 9.7) осуществляется методами той отрасли знаний, к которой относятся исследуемые явления. Но всякая *интерпретация начинается со статистической оценки уравнения регрессии в целом* и *оценки значимости входящих в модель факторных признаков*, т.е. с выяснения, как они влияют на величину результативного признака. Чем больше величина коэффициента регрессии, тем значительнее влияние данного признака на моделируемый. Особое значение при этом имеет знак перед коэффициентом регрессии. Знаки коэффициентов регрессии говорят о характере влияния на результативный признак. Если факторный признак имеет знак плюс, то с увеличением данного фактора результативный признак возрастает; если факторный признак со знаком минус, то с его увеличением результативный признак уменьшается. Интерпретация этих знаков полностью определяется социально-экономическим содержанием моделируемого (результативного) признака. Если его величина изменяется в сторону увеличения, то плюсовые знаки факторных признаков имеют положительное влияние. При изменении результативного признака в сторону снижения положительное значение имеют минусовые знаки факторных признаков. Если экономическая теория подсказывает, что факторный признак должен иметь положительное значение, а он со знаком минус, то необходимо проверить расчеты параметров уравнения регрессии. Такое явление чаще всего бывает из-за допущенных ошибок при решении. Однако при анализе совокупного влияния факторов при наличии взаимосвязей между ними характер их влияния может меняться. Для того чтобы быть уверенным, что факторный признак изменил знак влияния, необходима тщательная проверка решения данной модели, так как часто знаки могут меняться в силу допущенных ошибок при сборе или обработке информации.

При анализе адекватности уравнения регрессии исследуемому процессу возможны следующие варианты.

- построенная модель на основе ее проверки по F-критерию Фишера в целом адекватна, и все коэффициенты регрессии значимы. Такая модель может быть использована для принятия решений и осуществления прогнозов;
- модель по F-критерию Фишера адекватна, но часть коэффициентов регрессии незначима. В этом случае модель пригодна для принятия некоторых решений, но не для производства прогнозов;
- модель по F-критерию Фишера адекватна, но все коэффициенты регрессии незначимы. В этом случае модель полностью считается неадекватной. На ее основе не принимаются решения и не осуществляются прогнозы.

С целью расширения возможностей экономического анализа используются *частные коэффициенты эластичности*, определяемые по формуле

$$Э_{x_i} = a_i \cdot \frac{\bar{x}_i}{\bar{y}}, \tag{9.15}$$

где $Э_x$ – коэффициент эластичности;
$\bar{x}_i$ – среднее значение соответствующего факторного признака;
$\bar{y}$ – среднее значение результативного признака;
a_i – коэффициент регрессии при соответствующем факторном признаке.

Коэффициент эластичности показывает, на сколько процентов в среднем изменится значение результативного признака при изменении факторного признака на 1%.

Пример. Рассчитаем коэффициент эластичности ($Э_{x_i}$) по исходным данным зависимости между прибылью АОЗТ «Скат» (y), численностью работающих (x_1) и стоимостью основных фондов (x_2) по данным табл. 9.5.

$$a_1 = -41{,}43;\ \ a_2 = -7{,}6;$$

$$\bar{y} = \frac{4466}{6} = 744{,}3;$$

$$\bar{x}_1 = \frac{502}{6} = 83{,}67;\ \ \bar{x}_2 = \frac{29{,}2}{6} = 4{,}9;$$

$$Э_{x_1} = a_i \frac{\bar{x}_1}{\bar{\text{у}}} = -41{,}43 \cdot \frac{83{,}67}{744{,}3} = -4{,}7;$$

$$Э_{x_2} = a_2 \frac{\bar{x}_2}{\bar{y}} = -7{,}6 \cdot \frac{4{,}9}{744{,}3} = -0{,}05.$$

Это значит, что при увеличении численности работающих на 1% прибыль АОЗТ «Скат» снизится на 4,7%, а при увеличении стоимости основных фондов на 1% прибыль снизится на 0,05%.

Другим показателем экономического анализа является *частный коэффициент детерминации:*

$$d_{x_i} = r_{yx_i} \cdot \text{в}_{x_i} \tag{9.16}$$

где d_{x_i} – частный коэффициент детерминации:
r_{yx_i} – парный коэффициент корреляции между результативным и i-м факторным признаками;
β_{x_i} – соответствующий коэффициент уравнения множественной регрессии в стандартизованном масштабе.

Частный коэффициент детерминации показывает, на сколько процентов вариация результативного признака объясняется вариацией i-го признака, входящего в множественное уравнение регрессии.

Пример. По данным, приведенным в табл. 9.5, рассчитаем частный коэффициент детерминации для фактора x_1 – численности работающих:

$$d_{x_1} = r_{yx_1} \cdot \text{в}_{x_1};$$

$$r_{yx_1} = \frac{\overline{yx_1} - \bar{y} \cdot \bar{x}_1}{\text{у}_y \text{у}_{x_1}};$$

$$\overline{yx_1} = \frac{\sum yx_1}{n} = \frac{362\,404}{6} = 60400{,}7;$$

$$\bar{y} = \frac{\sum y}{n} = \frac{4466}{6} = 744{,}3;$$

$$\bar{x}_1 = \frac{\sum x_1}{n} = \frac{502}{6} = 83,67;$$

$$\text{у}_y^2 = \overline{y^2} - (\bar{y})^2 = \frac{3\,792\,340}{6} - (744,3)^2 = 78074,2;$$

$$\text{у}_y = \sqrt{\text{у}_y^2} = \sqrt{78074,2} = 279,4;$$

$$\text{у}_{x_1}^2 = \overline{x_1^2} - (\bar{x}_1)^2 = \frac{42280}{6} - (83,67)^2 = 46;$$

$$\text{у}_{x_1} = \sqrt{\text{у}_{x_1}^2} = \sqrt{46} = 6,78;$$

$$r_{yx_1} = \frac{60400,7 - 744,3 \cdot 83,67}{279,4 \cdot 6,78} = -0,98;$$

$$\text{в}_{x_1} = a_1 \cdot \frac{\text{у}_{x_1}}{\sigma_y} = -41,43\,\frac{6,78}{279,4} = -1,006;$$

$$d_{x_1} = -0,98 \cdot (-1,006) = 0,99.$$

Рассчитаем d_{x_2} – частный коэффициент детерминации для фактора x_2 – стоимости основных фондов:

$$d_{x_2} = r_{yx_2} \cdot \text{в}_{x_2};$$

$$\overline{yx_2} = \frac{22748,4}{6} = 3791,4;$$

$$\bar{x}_2 = \frac{29,2}{6} = 4,9;$$

$$\text{у}_{x_2}^2 = \frac{145,82}{6} - (4,9)^2 = 0,29;$$

$$\text{у}_{x_2} = 0,54;$$

$$r_{yx_2} = \frac{3791,4 - 744,3 \cdot 4,9}{279,4 \cdot 0,54} = -0,96;$$

$$\text{в}_{x_2} = 7,6\,\frac{0,54}{279,4} = 0,01;$$

$$d_{x_2} = 0,96 \cdot 0,01 = 0,01.$$

Это свидетельствует о том, что на 99% вариация прибыли АОЗТ «Скат» объясняется изменением численности работающих.

Множественный коэффициент детерминации (R^2), представляющий собой множественный коэффициент корреляции в квадрате; он характеризует, какая доля вариации результативного признака обусловлена изменением факторных признаков, входящих в многофакторную регрессионную модель.

Для более точной оценки влияния каждого факторного признака на моделируемый используют Q-коэффициент, определяемый по формуле

$$Q_{x_i} = Э_{x_i} \cdot V_{x_i}, \tag{9.17}$$

где V_{x_1} – коэффициент вариации соответствующего факторного признака;
Q_{x_1} – для фактора x_1 – численности работающих:

$$Q_{x_1} = Э_{x_1} \cdot Vx_1;$$
$$Э_{x_1} = 4,7;$$
$$Vx_1 = \frac{у_{x_1}}{\bar{x}_1} \cdot 100\% = \frac{6,79}{83,67} \cdot 100\% = 8,1\%;$$
$$Qx_1 = -4,7 \cdot 0,081 = -0,38.$$

Q_{x_2} – для фактора x_2 – стоимости основных фондов:

$$Q_{x_2} = Э_{x_2} \cdot V_{x_2};$$
$$Э_{x_2} = -0,05;$$
$$Vx_2 = \frac{у_{x_2}}{\bar{x}_2} \cdot 100\% = \frac{0,54}{4,9} \cdot 100\% = 11,0\%;$$
$$Qx_2 = -0,05 \cdot 11,0 : 100 = -0,006.$$

В целом, оценивая положительно значение уравнений регрессии как адекватных моделей связи, необходимо отметить их отрицательные свойства. Хорошую аппроксимацию эти модели имеют только для тех значений результативного признака, которые находятся в середине ранжированного ряда индивидуальных значений признака.

Ошибка аппроксимации для данных значений не превышает 1–2%. На концах же исходного ряда величина ошибки аппроксимации может достигать до 50%. На основе уравнений регрессии невозможно получить оптимальное значение моделируемого показателя. Модели на основе уравнений регрессии обладают слабыми экстраполяционными свойствами, так как не отражают тенденции развития социально-экономических явлений и процессов и годны для построения лишь краткосрочных прогнозов, носящих вероятностный характер.

Таким образом, обобщенная блок-схема интерпретации моделей регрессии имеет вид (рис. 9.7).

Наиболее полная экономическая интерпретация моделей регрессии позволяет выявить резервы развития и повышения деловой активности субъектов экономики.

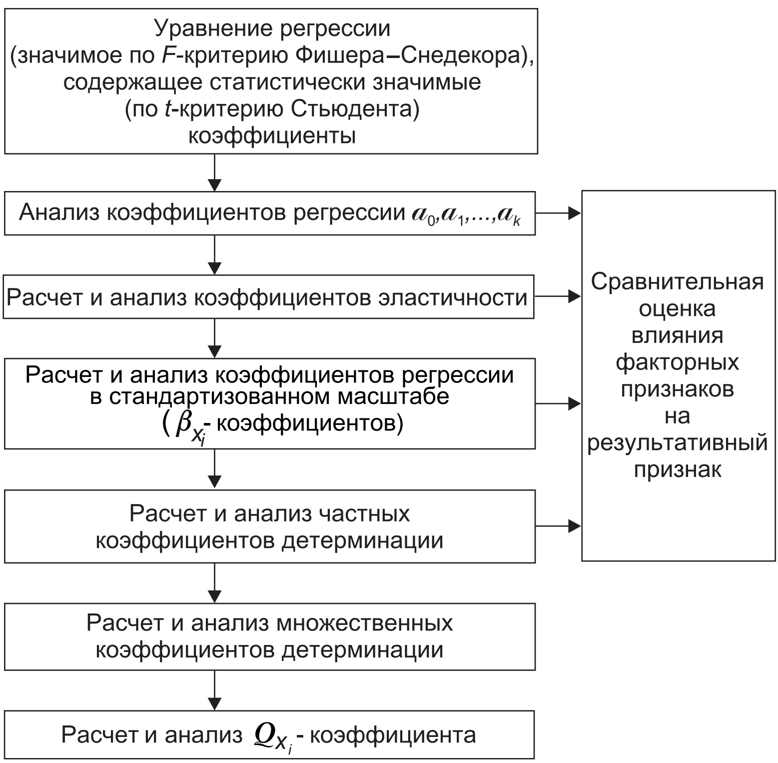

Рис. 9.7. Схема интерпретации моделей регрессии

9.6 СОБСТВЕННО-КОРРЕЛЯЦИОННЫЕ ПАРАМЕТРИЧЕСКИЕ МЕТОДЫ ИЗУЧЕНИЯ СВЯЗИ. ОЦЕНКА СУЩЕСТВЕННОСТИ КОРРЕЛЯЦИИ

Измерение тесноты и направления связи является важной задачей изучения и количественного измерения взаимосвязи социально-экономических явлений. Оценка тесноты связи между признаками предполагает определение меры соответствия вариации результативного признака от одного (при изучении парных зависимостей) или нескольких (множественных) факторов.

Линейный коэффициент корреляции был впервые введен в начале 1890-х гг. Пирсоном, Эджвортом и Велдоном и характеризует тесноту и направление связи между двумя коррелируемыми признаками в случае наличия между ними линейной зависимости.

В теории разработаны и на практике применяются различные модификации формул для расчета данного коэффициента:

$$r = \frac{\overline{(x-\bar{x})\cdot(y-\bar{y})}}{\sigma_x \cdot \sigma_y}. \qquad (9.18)$$

где r_{xy} – линейный коэффициент корреляции.

Используя математические свойства средней и формулу (9.18), получим:

$$r_{xy} = \frac{\overline{(x-\bar{x})\cdot(y-\bar{y})}}{\sigma_x \cdot \sigma_y} = \frac{\overline{xy} - \overline{\bar{x}y} - \overline{x\bar{y}} + \overline{\bar{x}\bar{y}}}{\sigma_x \cdot \sigma_y} =$$

$$= \frac{\overline{xy} - \bar{x}\bar{y} - \bar{x}\bar{y} + \bar{x}\bar{y}}{\sigma_x \cdot \sigma_y} = \frac{\overline{xy} - \bar{x}\bar{y}}{\sigma_x \cdot \sigma_y}. \qquad (9.19)$$

Дальнейшие преобразования позволяют получить следующую формулу линейного коэффициента корреляции:

$$r_{xy} = \frac{\sum (x-\bar{x})\cdot(y-\bar{y})}{n\cdot \text{у}_x \cdot \text{у}_y}, \text{ или } r_{xy} = \frac{\sum (x-\bar{x})\cdot(y-\bar{y})}{\sqrt{\sum (x-\bar{x})^2 \cdot \sum (y-\bar{y})^2}}, \qquad (9.20)$$

где n – число наблюдений.

Произведя расчет по итоговым значениям исходных переменных, линейный коэффициент корреляции можно вычислить по формуле:

$$r_{xy} = \frac{n\cdot\sum xy - \sum x\cdot\sum y}{\sqrt{\left[n\sum x^2 - \left(\sum x\right)^2\right]\cdot\left[n\sum y^2 - \left(\sum y\right)^2\right]}} \qquad (9.21)$$

или

$$r_{xy} = \frac{\sum xy - \sum x\cdot\frac{\sum y}{n}}{\sqrt{\left[\sum x^2 - \frac{\left(\sum x\right)^2}{n}\right]\cdot\left[\sum y^2 - \frac{\left(\sum y\right)^2}{n}\right]}}. \qquad (9.22)$$

Коэффициент корреляции может быть выражен через дисперсии слагаемых:

$$r_{xy} = \frac{\text{у}_x^2 + \text{у}_y^2 - \text{у}_{x-y}^2}{2\text{у}_x\cdot\text{у}_y}. \qquad (9.23)$$

Формулы (9.21), (9.22), (9.23) применяются при изучении совокупностей малого объема ($n \le 20 \div 30$).

Между линейным коэффициентом корреляции и коэффициентом регрессии существует определенная зависимость, выражаемая формулой

$$r_{xy} = a_i\,\frac{\text{у}_{x_i}}{\text{у}_y}, \qquad (9.24)$$

где a_i – коэффициент регрессии в уравнении связи;
σ_{x_i} – среднее квадратическое отклонение соответствующего статистически существенного факторного признака.

Линейный коэффициент корреляции имеет большое значение при исследовании социально-экономических явлений и процессов, распределение которых близко к нормальному. Легко доказывается, что условие $r_{xy} = 0$ является необходимым и достаточным для того, чтобы величины x и y были независимы. При этом условии коэффициенты регрессии a_{yx}, a_{xy} также обращаются в нуль, а прямые регрессии y по x и x по y оказываются взаимно перпендикулярными (параллельными: одна оси абсцисс, а вторая оси ординат).

Если же $r_{yx} = 1$, то это означает, что все точки (x, y) находятся на прямой и зависимость между x и y является функциональной. Прямые регрессии в этом случае совпадают. Указанное положение распространяется также на случай нормального распределения трех и более величин.

Линейный коэффициент корреляции изменяется в пределах от – 1 до 1: $-1 \le r \le 1$. Знаки коэффициентов регрессии и корреляции совпадают. При этом интерпретацию выходных значений коэффициента корреляции можно представить в табл. 9.6.

Значимость линейного коэффициента корреляции проверяется на основе t-критерия Стьюдента. При этом выдвигается и проверяется гипотеза (H_0) о равенстве коэффициента корреляции нулю [$H_0 : r_{yx} = 0$]. При проверке этой гипотезы используется t-статистика:

$$t_p = \sqrt{\frac{r_{xy}{}^2}{1-r_{xy}{}^2}\cdot(n-2)} = \frac{|r_{xy}|}{\sqrt{1-r_{xy}{}^2}}\sqrt{n-2}. \qquad (9.25)$$

Таблица 9.6

Оценка линейного коэффициента корреляции

Значение линейного коэффициента связи	Характер связи	Интерпретация связи
$r = 0$	Отсутствует	–
$0 < r < 1$	Прямая	С увеличением x увеличивается y
$-1 < r < 0$	Обратная	С увеличением x уменьшается y, и наоборот
$r = 1$	Функциональная	Каждому значению факторного признака строго соответствует одно значение результативного признака

При выполнении H_0 t-статистика имеет распределение Стьюдента с входными параметрами: $\{\alpha, k = n - 2\}$.

Если расчетное значение $t_р > t_{кр}$ (табличное), то гипотеза H_0 отвергается, что свидетельствует о значимости линейного коэффициента корреляции, а следовательно, и о статистической существенности зависимости между x и y.

Данный критерий оценки значимости применяется для совокупностей $n < 50$.

При большом числе наблюдений ($n > 100$) используется следующая формула t-статистики:

$$t_p = \frac{|r|}{\sqrt{1-r^2}} \cdot \sqrt{n}. \qquad (9.26)$$

Для статистически значимого линейного коэффициента корреляции можно построить интервальные оценки с помощью Z-распределения Фишера:

$$Z = \frac{1}{2} \cdot \ln \frac{1+r}{1-r}.$$

Первоначально определяется интервальная оценка для Z по выражению

$$Z \in \left[Z' \pm t_{\Gamma} \cdot \sqrt{\frac{1}{n-3}} \right], \qquad (9.27)$$

где t_γ – табулированные значения для нормального распределения, зависимые от $\gamma = 1 - \alpha$ (α – уровень вероятности);

Z' – табличные значения, $Z = f(r)$ – распределения. Функция Z' –нечетная, т.е. $Z' = f(-r) = -f(r)$.

Пример. На основе выборочных данных о деятельности аудиторско-консультационных фирм Москвы в 2001 г. оценим тесноту связи между совокупной выручкой этих фирм и общей численностью профессионалов (табл.9.7).

Таблица 9.7

Расчетные данные для определения коэффициентов корреляции

№ п/п	Совокупная выручка, млн руб. y	Общая численность профессионалов, чел., x	xy	y^2	x^2
1	2,62	23	60,26	6,86	529
2	3,04	32	97,28	9,24	1024
3	3,15	50	157,50	9,92	2500
4	3,83	53	202,99	14,67	2809
5	3,58	55	196,90	12,82	3025
6	4,08	58	236,64	16,65	3364
7	4,09	59	241,31	16,73	3481
8	4,20	62	260,40	17,64	3844
9	4,18	69	288,42	17,47	4761
10	4,24	75	318,00	17,98	5625
Сумма	37,01	536	2059,7	139,98	30962
Средняя	3,701	53,6	205,97	13,998	3096,2

1. Используя формулу (9.19), получаем:

$$r_{xy} = \frac{\overline{xy} - \bar{x}\,\bar{y}}{у_x \cdot у_y};$$

$$у_y^2 = \overline{y}^2 - (\bar{y})^2 = 13,998 - (3,701)^2 = 0,301;$$

$$у_x^2 = \overline{x}^2 - (\bar{x})^2 = 3096,2 - (53,6)^2 = 223,24;$$

$$r_{xy} = \frac{205,97 - 53,6 \cdot 3,701}{\sqrt{223,24 \cdot 0,301}} = 0,93.$$

2. По формуле (9.21) значение коэффициента корреляции составило:

$$r_{xy}=\frac{10\cdot 2059{,}70-536\cdot 37{,}01}{\sqrt{\left[10\cdot 30962-536^2\right]\cdot\left[10\cdot 139{,}98-(37{,}01)^2\right]}}=$$
$$=\frac{20597-19837{,}36}{\sqrt{(309620-287296)\cdot(1399{,}8-1369{,}74)}}=$$
$$=\frac{759{,}64}{\sqrt{22324\cdot 30{,}06}}=\frac{759{,}64}{819{,}18}=0{,}93.$$

Результат тот же.

По формуле (9.22):

$$r_{xy}=\frac{2059{,}7-536\cdot\dfrac{37{,}01}{10}}{\sqrt{\left[30962-\dfrac{536^2}{10}\right]\cdot\left[139{,}98-\dfrac{37{,}01^2}{10}\right]}}=$$
$$=\frac{2059{,}7-1983{,}736}{\sqrt{(30962-28729{,}6)\cdot(139{,}98-136{,}97)}}=$$
$$=\frac{75{,}964}{\sqrt{2232{,}4\cdot 3{,}01}}=\frac{75{,}964}{81{,}973}=0{,}93.$$

Таким образом, результат по всем формулам одинаков и свидетельствует о сильной прямой зависимости между изучаемыми признаками.

Проверка значимости коэффициента корреляции:

$$t_p=\frac{|r_{xy}|}{\sqrt{1-{r_{xy}}^2}}\cdot\sqrt{n-2}=\frac{0{,}93}{\sqrt{1-0{,}93^2}}\cdot\sqrt{10-2}=7{,}156.$$

Гипотеза H_0 отвергается при уровне значимости $\alpha = 0{,}05$ и числе степеней свободы $k = 10 - 2 = 8$, так как $|t_р| > t_{kp} = 2{,}306$, что свидетельствует о значимости данного коэффициента корреляции.

Доверительные интервалы линейного коэффициента корреляции между суммарной выручкой и численностью профессионалов аудиторско-консультационных фирм Москвы получились: $\alpha = 0{,}05$; $\gamma = 1 - \alpha = 1 - 0{,}05 = 0{,}95$. Тогда $t_\gamma = 1{,}96$ – для нормального закона распределения (приложение 1) $r_{xy} = 0{,}93$; $Z' = 1{,}6584$.

По таблице Z-распределения Фишера (приложение 6):

$$Z' - t_г \cdot \sqrt{\frac{1}{n-3}} \le Z \le Z' + t_г \cdot \sqrt{\frac{1}{n-3}};$$
$$1{,}6584 - 1{,}96\sqrt{\frac{1}{10-3}} \le Z \le 1{,}6584 + 1{,}96\sqrt{\frac{1}{10-3}};$$
$$0{,}9176 \le Z \le 2{,}3992.$$

По таблице Z-распределения Фишера (приложение 6) определим r $(0{,}725 \le r \le 0{,}983)$.

В случае наличия линейной и нелинейной зависимости между двумя признаками для измерения тесноты связи применяют так называемое *корреляционное отношение*. Различают эмпирическое и теоретическое корреляционное отношение.

Эмпирическое корреляционное отношение рассчитывается по данным группировки, когда δ^2 характеризует отклонения групповых средних результативного показателя от общей средней:

$$з = \sqrt{\frac{y^2 - \overline{y}^2}{y^2}} = \sqrt{1 - \frac{\overline{y}^2}{y^2}} = \sqrt{\frac{\overline{\delta^2}}{y^2}}, \qquad (9.28)$$

где η – корреляционное отношение;
σ^2 – общая дисперсия;
$\overline{\sigma}^2$ – средняя из частных (групповых) дисперсий;
$\overline{\delta}^2$ – межгрупповая дисперсия (дисперсия групповых средних).

Все эти дисперсии являются дисперсиями результативного признака.

Теоретическое корреляционное отношение определяется по формуле:

$$\text{з} = \sqrt{\frac{\text{д}^2}{\text{у}^2}}, \tag{9.29}$$

где δ^2 – дисперсия выравненных значений результативного признака, т.е. рассчитанных по уравнению регрессии;

σ^2 – дисперсия эмпирических (фактических) значений результативного признака.

$$\text{д} = \sqrt{\frac{\sum(\bar{y}_x - \bar{y})^2}{n}} = \sqrt{\text{д}^2_{\bar{y}_x}};$$

$$\text{у} = \sqrt{\frac{\sum(y - \bar{y})^2}{n}} = \sqrt{\text{у}\,\frac{2}{y}}.$$

Тогда

$$\text{з} = \sqrt{\frac{\sum(\bar{y}_x - \bar{y})^2}{\sum(y - \bar{y})^2}} \tag{9.30}$$

объясняется влиянием факторного признака.

В основе расчета корреляционного отношения лежит правило сложения дисперсий (см. главу 7), т.е.

$$\text{у}^2 = \text{д}^2 + \overline{\text{у}_i^2}, \tag{9.31}$$

$\overline{\sigma}_i^2$ – при изучении степени коррелированности факторов отражает вариацию результативного признака (Y) под влиянием всех не учтенных при анализе факторов, т.е. носит остаточный характер:

$$\sigma_i^2 = \sigma^2_{\text{ост}} = \frac{\sum(y - \overline{\text{у}_x})^2}{n}.$$

Отсюда формула корреляционного отношения принимает вид

$$\text{з} = \sqrt{\frac{\text{д}^2}{\text{у}^2}} = \sqrt{\frac{\text{у}^2 - \text{у}^2_{\text{ост}}}{\text{у}^2}} = \sqrt{1 - \frac{\text{у}^2_{\text{ост}}}{\text{у}^2}}. \tag{9.32}$$

Корреляционное отношение изменяется в пределах от 0 до 1 ($0 \leq \eta \leq 1$), и анализ степени тесноты связи полностью соответствует линейному коэффициенту корреляции (см. табл. 9.6).

Теоретическое корреляционное отношение также может вычисляться по формуле

$$\text{з} = \sqrt{1 - \frac{\sum (y - \overline{y_x})^2}{\sum (y - y)^2}}.$$

Корреляционное отношение является более универсальным показателем тесноты связи по сравнению с линейным коэффициентом корреляции.

Для измерения тесноты связи при множественной корреляционной зависимости, т.е. при исследовании трех и более признаков одновременно, вычисляются множественный, или совокупный, и частные коэффициенты корреляции.

Множественный коэффициент корреляции рассчитывается при наличии линейной связи между результативным и несколькими факторными признаками, а также между каждой парой факторных признаков.

Он вычисляется по формуле

$$R_{y / x_1 x_2 \ldots} = \sqrt{\frac{\text{д}^2}{\text{у}^2}} = \sqrt{1 - \frac{\text{у}^2_{\text{ост}}}{\text{у}^2}},$$

где R – множественный коэффициент корреляции;
δ^2 – дисперсия теоретических значений результативного признака, рассчитанная по уравнению множественной регрессии;
$\sigma^2_{\text{ост}}$ – остаточная дисперсия;
σ^2 – общая дисперсия результативного признака.

В случае оценки связи между результативными (Y) и двумя факторными признаками (x_1) и (x_2) множественный коэффициент корреляции можно определить по формуле

$$R_{y/x_1x_2} = \sqrt{\frac{r^2_{yx_1} + r^2_{yx_2} - 2r_{yx_1} \cdot r_{yx_2} \cdot r_{x_1x_2}}{1 - r^2_{x_1x_2}}}, \qquad (9.33)$$

где r – парные коэффициенты корреляции между признаками.

Множественный коэффициент корреляции можно рассчитать, используя парные коэффициенты r_{ij} и коэффициенты регрессии в стандартизованном масштабе (β_i):

$$R_{x_1x_2,...,x_k} = \sqrt{в_1 \cdot r_{yx_1} + в_2 \cdot r_{yx_2} + ... + в_k \cdot r_{yx_k}},$$

где β_i – коэффициенты в стандартизованном масштабе.

Множественный коэффициент корреляции изменяется в пределах от 0 до 1 и по определению положителен: $0 \le R \le 1$.

Приближение R к единице свидетельствует о сильной зависимости между признаками. При небольшом числе наблюдений величина коэффициента множественной корреляции, как правило, завышается.

Чтобы оценить общую вариацию результативного (моделируемого) признака в зависимости от факторных признаков, величина коэффициента множественной корреляции корректируется на основании следующего выражения:

$$\hat{R}_{y/x_1x_2,...,x_k} = \sqrt{1 - \left(1 - R^2\right)\frac{n-1}{n-k-1}}, \qquad (9.34)$$

где $\hat{R}$ – скорректированное значение;
n – число наблюдений;
k – число факторных признаков.

Корррективровка R не производится при условии, если

$$\frac{n-k}{k} \geq 20.$$

Проверка значимости коэффициента множественной корреляции осуществляется на основе F-критерия Фишера–Снедекора:

$$F_p = \frac{\frac{1}{2} \cdot R^2_{y/x_1x_2}}{\frac{1}{n-3} \cdot (1-R^2_{y/x_1x_2})}. \quad (9.35)$$

Гипотеза H_0 о незначимости коэффициента множественной корреляции ($H_0 : R = 0$) отвергается, если $F_р > F_{kр}$ (α; $\nu_1 = 2$; $\nu_2 = n - 3$).

Оценка доверительных границ R производится следующим образом: величина R приравнивается к гиперболическому тангенсу величины Z, т.е. $R = thr$, где

$$Z = \frac{1}{2} \cdot \ln \frac{1+R}{1-R}.$$

Плотность распределения Z является почти нормальной со средним значением

$$\overline{Z} = \frac{1}{2} \cdot \ln \frac{1+R}{1-R} + \frac{R}{2(N-1)} \quad (9.36)$$

и дисперсией

$$у^2_z = \frac{1}{N-3}.$$

Следовательно,

$$P\left\{-t\text{у}_z < Z - Z_0 < t\text{у}_z\right\} = \text{ф}(t),$$

отсюда:

$$\begin{aligned} &Z - t\text{у}_R < Z_0 < Z + t\text{у}_R; \\ &Z_1 = Z - t\text{у}_Z; \\ &Z_2 = Z + t\text{у}_Z. \end{aligned} \tag{9.37}$$

По таблицам Z-преобразования Фишера находят R_1 и R_2, т.е. $R_1 < \mathrm{R} < R_Z$ – верхняя и нижняя границы значений R.

На основе данных табл. 9.5 рассчитаем коэффициент множественной корреляции и его ошибку:

$$r_{yx_1} = \frac{\overline{yx_1} - \overline{y}\,\overline{x_1}}{\text{у}_y \cdot \text{у}_{x_1}} = -0,98;$$

$$r_{yx_2} = \frac{\overline{yx_2} - \overline{y}\,\overline{x_2}}{\text{у}_y \cdot \text{у}_{x_2}} = -0,78;$$

$$r_{x_1x_2} = \frac{\overline{x_1x_2} - \overline{x_1}\,\overline{x_2}}{\text{у}_{x_1} \cdot \text{у}_{x_2}} = -0,86;$$

Матрица линейных коэффициентов корреляции имеет вид:

$$\begin{bmatrix} 1 & -0,98 & 0,78 \\ & 1 & -0,86 \\ & & 1 \end{bmatrix}.$$

Множественный коэффициент корреляции составит:

$$R_{y/x_1x_2} = \sqrt{\frac{-0,98^2 + 0,78^2 - 2(-0,98)\cdot 0,78 \cdot (-0,86)}{1-(-0,86)^2}} = 0,99.$$

Проверка условия (n–k) / k = (6–2) / 2 = 2 ≤ 20 подтверждает возможность проведения корректировки данного коэффициента:

$$\tilde{R}_{y/x_1x_2} = \sqrt{1-\left(1-R^2_{y/x_1x_2}\right)\cdot\frac{n-1}{n-k-1}} = \sqrt{1-(1-0,99)^2\cdot\frac{6-1}{6-3}} = 0,98.$$

Проверка значимости коэффициента множественной корреляции показала:

$$F_{\text{р}} = \frac{1/2\cdot 0,98^2}{1/3\cdot(1-0,98)^2} = \frac{0,4802}{0,0132} = 36,3.$$

Гипотеза о незначимости коэффициента корреляции отвергается, так как

$$F_{\text{kp}} = 9,55\ (\alpha = 0,05;\ \text{н}_1 = 2;\ \text{н}_2 = n-3 = 3),\ \left[F_{\text{р}} = 36,35\right] > \left[F_{\text{kp}} = 9,55\right].$$

Определим доверительные границы, в которых находится R в генеральной совокупности. При этом доверительную вероятность примем равной $\gamma = 0,95$ ($\alpha = 0,05$; $\gamma = 1 - \alpha$), $t_\gamma = 1,96$ (приложение 1).

$$R = 0,98;\ z = 2,2976;\ \sigma_R = 1 / (6–1) = 0,2;$$

$$–1,96 \cdot 0,2 < Z – Z_0 \leq 1,96 \cdot 0,2;$$

$$–0,392 \leq Z – Z_0 \leq 0,392;$$

$$Z_1 = 2,2976 – 0,392 = 1,9056;$$

$$Z_2 = 2,2976 + 0,392 = 2,6896;$$

$$1,9056 \leq Z \leq 2,6896.$$

Отсюда по таблице Z-преобразования Фишера (приложение 6) получаем: $0,96 \leq R \leq 0,941$.

Частные коэффициенты корреляции характеризуют степень тесноты связи между двумя признаками x_1 и x_2 при фиксированном значении других (k – 2) факторных признаков, т.е. когда влияние x_3 исключается и оценивается связь между x_1 и x_2 в «чистом виде».

Коэффициент, в котором исключается влияние только одного факторного признака, называется *коэффициентом частной корреляции первого порядка*. В общем виде коэффициент корреляции первого порядка выражается так:

$$r_{1,2,3,4,...,k} = \frac{r_{1,2,3,4,...,k-1} - r_{1,k,3,...,k-1} \cdot r_{2,k,3,...,k-1}}{\sqrt{(1-r^2_{1,k,3,...,k-1}) \cdot (1-r^2_{2,k,3,...,k-1})}}.$$

В случае зависимости Y от двух факторных признаков x_1 и x_2 коэффициент частной корреляции следующий:

$$r_{yx_1/x_2} = \frac{r_{yx_1} - r_{x_1x_2} \cdot r_{yx_2}}{\sqrt{(1-r^2_{x_2y}) \cdot (1-r^2_{x_1x_2})}};$$
$$r_{yx_2/x_1} = \frac{r_{yx_2} - r_{x_1y} \cdot r_{x_1x_2}}{\sqrt{(1-r^2_{x_1y}) \cdot (1-r^2_{x_1x_2})}}, \qquad (9.38)$$

где r – парные коэффициенты корреляции между указанными в индексе переменными.

В первом случае исключено влияние факторного признака x_2, во втором – x_1. Значения парного и частного коэффициентов корреляции отличаются друг от друга, так как парный коэффициент характеризует связь между двумя признаками без учета влияния других признаков, а частный – учитывает наличие и влияние других факторов.

Проверка значимости и расчет доверительных интервалов для частных коэффициентов корреляции аналогичны расчетам для парных коэффициентов с тем лишь отличием, что число степеней свободы ν определяется как

$$\nu = n - k,$$

где k – порядок коэффициента частной корреляции.

На основании приведенных выше данных о зависимости трех факторов деятельности предприятий вычислим частные коэффициенты корреляции первого порядка:

$$r_{yx_1/x_2} = \frac{-0,99-0,78\cdot(-0,86)}{\sqrt{(1-0,78)^2}\cdot\sqrt{1-(-0,86)^2)}} = -0,999;$$

$$r_{yx_2/x_1} = \frac{0,78-(-0,99)\cdot(-0,86)}{\sqrt{(1-(-0,99)^2}\cdot\sqrt{(1-(0,78)^2)}} = -0,992;$$

$$r_{x_1x_2/y} = \frac{-0,86-(-0,99)\cdot(0,78)}{\sqrt{(1-(-0,99)^2)}\cdot\sqrt{(1-(0,78)^2)}} = -0,994.$$

Проверка значимости частных коэффициентов корреляции показала:

$$t_{p(yx_1/x_2)} = \frac{r_{yx_1/x_2}}{\sqrt{1-r^2_{yx_1/x_2}}}\cdot\sqrt{n-3} = \frac{-0,999\cdot\sqrt{3}}{\sqrt{1-(-0,999)^2}} = -38,7;$$

$$t_{p(yx_2/x_1)} = \frac{r_{yx_2/x_1}}{\sqrt{1-r^2_{yx_2/x_1}}}\cdot\sqrt{n-3} = \frac{-0,994\cdot\sqrt{3}}{\sqrt{1-(-0,994)^2}} = -15,74;$$

$$t_{p(x_1x_2/y)} = \frac{r_{x_1x_2/y}}{\sqrt{1-r^2_{x_1x_2/y}}}\cdot\sqrt{n-3} = \frac{-0,992\cdot\sqrt{3}}{\sqrt{1-(-0,992)^2}} = -13,61;$$

$$\left\{\left|t_{p(yx_1/x_2)}\right|; \left|t_{p(yx_2/x_1)}\right|; \left|t_{p(x_1x_2/y)}\right| > t_{kp}\ (\alpha = 0,05,\ v = n-3 = 3) = 3,181\right\}.$$

Следовательно, все приведенные коэффициенты корреляции значимы.

Для значимых частных коэффициентов корреляции доверительные интервалы составят:

$$Z' - t_y\cdot\sqrt{\frac{1}{n-4}} \le Z \le Z' + t_\gamma\cdot\sqrt{\frac{1}{n-4}};$$

$$\left|r_{yx_1/x_2}\right| = 0,999;$$

$$Z = 3,8002;$$

$$3,8002 - 1,96\cdot\sqrt{1/2} \le Z \le 3,8002 + 1,96\cdot\sqrt{1/2};$$

$$2,414 \le Z \le 5,186.$$

Используя таблицу Z-преобразований Фишера, получаем коэффициент корреляции в доверительных интервалах $0{,}98 \le r \le 1$ и т.д.

Обобщенную методику корреляционного метода анализа экономических явлений и процессов можно представить блок-схемой (рис. 9.8).

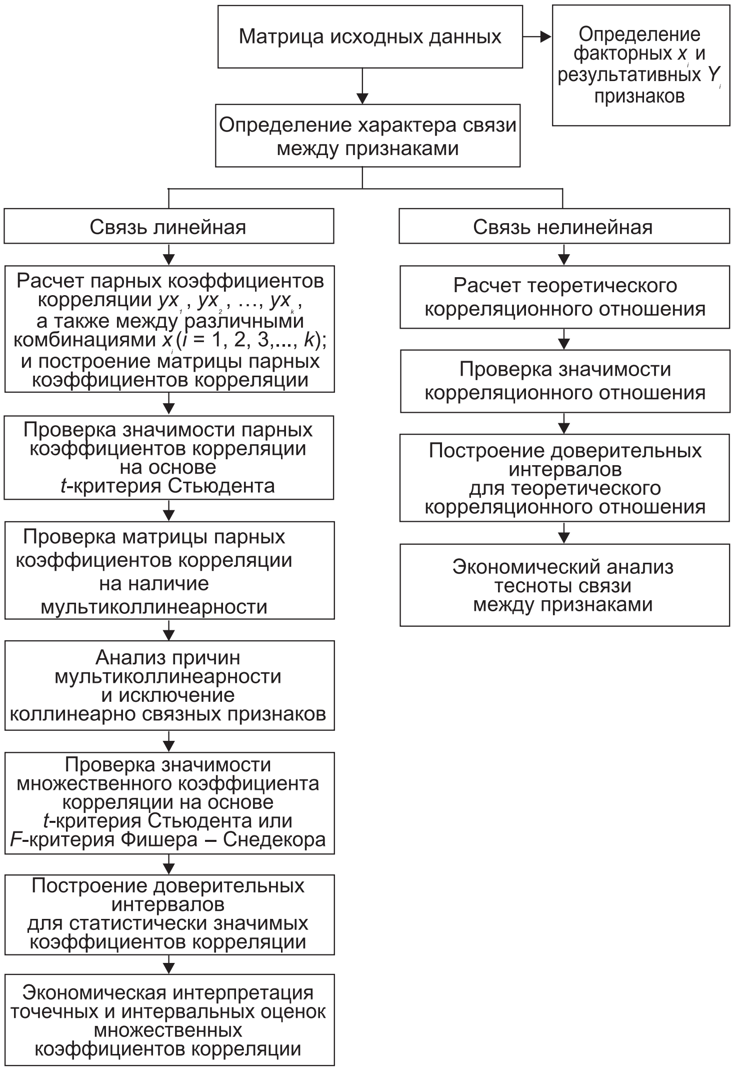

Рис. 9.8. Схема реализации корреляционного метода анализа

9.7
МЕТОДЫ ИЗУЧЕНИЯ СВЯЗИ СОЦИАЛЬНЫХ ЯВЛЕНИЙ

Важной задачей статистики является разработка методики статистической оценки социальных явлений, которая осложняется тем, что многие социальные явления не имеют количественной оценки.

Как правило, анализ социальных явлений, их связей и зависимостей должен начинаться с построения графиков связей. В настоящее время используются графики, характеризующие связь социальных явлений (рис. 9.9).

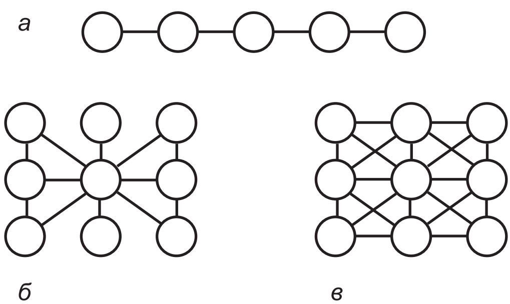

Рис. 9.9. Графики, характеризующие связь социальных явлений

С помощью графика (рис. 9.9а) «цепь» изображаются связи между социальными признаками, которые одинаково существенны и значимы.

График (рис. 9.9б) «звезда» изображает зависимость социальных явлений, которые тяготеют к одному наиболее значимому. Исключение данного признака нарушает взаимосвязи между оставшимися признаками.

На графике (рис. 9.9в) «сетка» выделяется несколько значимых признаков, которые тесно зависимы друг от друга.

Для количественной характеристики многомерных (многофакторных) связей социальных явлений используется *метод корреляционных плеяд*, основанный на расчете коэффициентов связи, которые носят общее название *информативных коэффициентов*. Метод корреляционных плеяд позволяет сгруппировать взаимосвязанные признаки в так называемые плеяды. Плеяды выделяются на основе матрицы информационных коэффициентов корреляции, коэффициентов сопряженности и т.д. Алгоритм построения корреляционных плеяд базируется на выделении максимальных значений информационных коэффициентов в исходной матрице значений. На базе отдельно выделенных плеяд строится «дерево плеяд», особенность которого в том, что внутриплеядные связи между факторными признаками тесные, а межплеядные – слабые.

Информационной основой для такого анализа служат данные различных социологических обследований на базе анкетирования.

Вычисление информационных коэффициентов служит основой для дальнейшего углубленного анализа связей между социальными явлениями. В настоящее время такой углубленный статистический анализ проводится при разработке корреляционных плеяд с дальнейшим переходом к факторному анализу или методу главных компонент.

Количественная оценка связей социальных явлений осуществляется на базе расчета и анализа целого ряда коэффициентов.

При наличии соотношения между вариацией качественных признаков говорят об их ассоциации, взаимосвязанности. Для оценки связи в этом случае используют ряд показателей.

Коэффициенты оценки связи качественных признаков, представленных двумя градациями. Для определения тесноты связи двух качественных признаков, каждый из которых состоит только из двух групп, применяются *коэффициенты ассоциации и контингенции*. При исследовании связи числовой материал располагают в виде таблиц сопряженности, например табл. 9.8. Для вычисления строится таблица, которая показывает связь между двумя явлениями, каждое из которых должно быть альтернативным, т.е. состоящим из двух качественно отличных друг от друга значений признака (например, хороший, плохой).

Таблица 9.8

Таблица для вычисления коэффициентов ассоциации и контингенции

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

Коэффициенты определяются по формулам:
коэффициент ассоциации ($K_а$)

$$K_а = \frac{ad - bc}{ad + bc}; \qquad (9.39)$$

коэффициент контингенции ($K_к$)

$$K_к = \frac{ad - bc}{\sqrt{(a+b)\cdot(b+d)\cdot(a+c)\cdot(c+d)}}. \qquad (9.40)$$

Коэффициент контингенции всегда меньше коэффициента ассоциации. Связь считается подтвержденной, если $K_а \geq 0{,}5$ или $K_к \geq 0{,}3$.

Пример. В одном из отделений Сбербанка России исследовалась связь между наличием вклада и семейным положением потенциальных вкладчиков на 01.01.2002 г. Результаты обследования характеризуются следующими данными (табл. 9.9).

Таблица 9.9

Зависимость наличия вкладов от семейного положения вкладчиков в одном из отделений Сбербанка России на 01.01.2002 г.

Семейное положение	Число вкладчиков, чел.	Из них	
		имеющие сбережения	не имеющие сбережения
Одинокие	400	250	150
Семейные	1250	800	450
Итого	1650	1050	600

$$K_а = \frac{250\cdot 450 - 150\cdot 800}{250\cdot 450 + 150\cdot 800} = \frac{112\,500 - 120\,000}{112\,500 + 120\,000} = \frac{7500}{232\,500} = -0{,}03;$$

$$K_к = \frac{250\cdot 450 - 150\cdot 800}{(250+150)\cdot(150+450)\cdot(450+800)\cdot(800+250)}$$

$$= \frac{112500 - 120000}{\sqrt{400 \cdot 600 \cdot 1250 \cdot 1050}} = \frac{7500}{561248,7} = -0,01.$$

Таким образом, наличие или отсутствие сбережений в обследуемом отделении Сбербанка России не зависит от семейного положения потенциальных вкладчиков.

Коэффициенты оценки связи качественных признаков, представленных несколькими градациями.

Когда каждый из качественных признаков состоит более чем из двух групп, то для определения тесноты связи возможно применение *коэффициента взаимной сопряженности Пирсона–Чупрова* (табл. 9.10).

Этот коэффициент вычисляется по следующей формуле:

Таблица 9.10

Вспомогательная таблица для расчета коэффициента взаимной сопряженности

x \ y	I	II	III	Всего
I II III			n_{xy}	n_x n_x n_x
Итого	n_y	n_y	n_y	n

$$K_{\text{п}} = \sqrt{\frac{\varphi^2}{1+\varphi^2}}; \quad K_{\text{ч}} = \sqrt{\frac{\varphi^2}{(K_1 - 1)(K_2 - 1)}}, \tag{9.41}$$

где φ^2 – показатель взаимной сопряженности;

φ – определяется как сумма отношений квадратов частот каждой клетки таблицы к произведению итоговых частот соответствующего столбца и строки. Вычитая из этой суммы 1, получим величину φ^2.

$$\varphi^2 = \sum \frac{n_{xy}^2}{n_x n_y} - 1;$$

K_1 – число значений (групп) первого признака;
K_2 – число значений (групп) второго признака.

Чем ближе величины $K_ч$ и $K_п$ к 1, тем связь теснее.

$$1+\varphi^2 = \sum \frac{\sum \frac{n_{xy}^2}{n_x}}{n_y} = \sum \frac{\sum \frac{n_{xy}^2}{n_y}}{n_x}.$$

Пример. С помощью коэффициента взаимной сопряженности проанализируем зависимость распределения сотрудников строительной фирмы ООО «Скат» по категориям от уровня их образования (табл. 9.11).

Таблица 9.11

Зависимость распределения сотрудников ООО «Скат» от уровня их образования*

Образование	Категория сотрудников			Итого
	руководители	служащие	рабочие	
Высшее	10	30	5	45
Неполное высшее	7	25	10	42
Среднее специальное	2	15	50	67
Среднее общее	1	10	25	36
Итого	20	80	90	190

* Данные условные.

$$1+\varphi^2 = \frac{\frac{10^2}{20}+\frac{30^2}{80}+\frac{5^2}{90}}{45} + \frac{\frac{7^2}{20}+\frac{25^2}{80}+\frac{10^2}{90}}{42} + \frac{\frac{2^2}{20}+\frac{15^2}{80}+\frac{50^2}{90}}{67} + \frac{\frac{1^2}{20}+\frac{10^2}{80}+\frac{25^2}{90}}{36} =$$

$$= \frac{5,00+11,25+0,28}{45} + \frac{2,45+7,81+1,11}{42} + \frac{0,20+2,81+27,78}{67} + \frac{0,05+1,25+6,94}{36} =$$

$$= 0,367+0,271+0,459+0,229 = 1,326;$$

$$1+\varphi^2 = 1{,}326;$$

$$\varphi^2 = 0{,}326;$$

$$K_{\text{п}} = \sqrt{\frac{0{,}326}{1{,}326}} = \sqrt{0{,}246} = 0{,}496$$

$$K_{\text{ч}} = \sqrt{\frac{0{,}326}{(4-1)\cdot(3-1)}} = \sqrt{\frac{0{,}326}{\sqrt{3\cdot 2}}} = \sqrt{\frac{0{,}326}{2{,}449}} = 0{,}365$$

Связь близка к умеренной.

В статистике существуют модификации *коэффициента Чупрова,* например, *через расчет* χ^2*-критерия Пирсона.* Коэффициент взаимной сопряженности ($K_{\text{ч}}$) вычисляется по формуле

$$K_{\text{ч}} = \sqrt{\frac{\chi^2}{n+\chi^2}}, \tag{9.42}$$

где $\text{ч}^2 = n\left\{\sum_{xy} n_{xy}^2 / n_x \cdot n_y - 1\right\}$ – наиболее распространенный критерий согласия, используемый для проверки статистической гипотезы о виде распределения. Коэффициент Чупрова изменяется в пределах $0 \le K_{\text{ч}} \le 1$.

По данным табл. 9.11 получили следующие результаты:

$$\text{ч}^2 = 190\cdot\left\{\begin{array}{l}\frac{10^2}{20\cdot 45}+\frac{30^2}{80\cdot 45}+\frac{5^2}{90\cdot 45}+\frac{7^2}{20\cdot 42}+\\ +\frac{25^2}{80\cdot 42}+\frac{10^2}{90\cdot 42}+\frac{2^2}{20\cdot 67}+\frac{15^2}{80\cdot 67}+\\ +\frac{50^2}{90\cdot 67}+\frac{1^2}{20\cdot 36}+\frac{10^2}{80\cdot 36}+\frac{25^2}{90\cdot 36}-1\end{array}\right\}=$$

$$=190\cdot\left\{\begin{array}{l}0{,}111+0{,}250+0{,}006+\\ +0{,}058+0{,}186+0{,}026+0{,}003+\\ +0{,}042+0{,}415+0{,}001+\\ +0{,}035+0{,}193-1\end{array}\right\}=190\cdot\{1{,}326-1\}=61{,}94;$$

$$K_ч = \sqrt{\frac{61{,}94}{190+61{,}94}} = \sqrt{\frac{61{,}94}{251{,}94}} = 0{,}496.$$

Другой модификацией *коэффициента взаимной сопряженности Чупрова* является:

$$K_ч = \sqrt{\frac{\chi^2}{n \cdot \sqrt{(k_1-1)\cdot(k_2-1)}}}, \qquad (9.43)$$

где K_1 – число строк в таблице;
K_2 – число граф в таблице;
n – число наблюдений.

Вычислим величину $K_ч$ для приведенного примера:

$$K_ч = \sqrt{\frac{61{,}94}{190\sqrt{(4-1)\cdot(3-1)}}} = \sqrt{\frac{61{,}94}{190\cdot\sqrt{3\cdot 2}}} = \sqrt{\frac{61{,}94}{190\cdot 2{,}449}} = \sqrt{0{,}133} = 0{,}365.$$

Связь средняя.

Особое значение для оценки связи имеет *бисcериальный коэффициент корреляции,* который дает возможность оценить связь между качественным альтернативным и количественным варьирующим признаками. Данный коэффициент вычисляется по формуле

$$r = \frac{\left|\bar{Y}_2 - \bar{Y}_1\right|}{у_2} \cdot \frac{pq}{Z}, \qquad (9.44)$$

где $\bar{Y}_2$ и $\bar{Y}_1$ – средние в группах;
$\sigma_у$ – среднее квадратическое отклонение фактических значений признака от среднего уровня;
p – доля первой группы;
q – доля второй группы;
z – табулированные (табличные) значения Z-распределения в зависимости от p.

Пример. Уровень доходов сотрудников одной коммерческой структуры характеризуется следующими данными (табл. 9.12).

Таблица 9.12

Зависимость уровня доходов строительной фирмы от формы собственности

Форма собственности фирмы	Σровень годовых доходов, тыс. руб.				Всего человек
	200–300	300–400	400–500	500–600	
	250	350	450	550	
Частная	5	7	6	4	22
Федеральная	9	4	2	1	16
Итого	14	11	8	5	38

[*] Цифры условные

$$\overline{Y}_1 = \frac{250\cdot 5 + 350\cdot 7 + 450\cdot 6 + 550\cdot 4}{22} = \frac{8600}{22} = 390{,}9;$$

$$\overline{Y}_2 = \frac{250\cdot 9 + 350\cdot 4 + 450\cdot 2 + 550\cdot 1}{16} = \frac{5100}{16} = 318{,}8;$$

$$\overline{Y}_{\text{общ}} = \frac{250\cdot 14 + 350\cdot 11 + 450\cdot 8 + 550\cdot 5}{38} = \frac{13700}{38} = 360{,}5;$$

$$\sigma_{\text{у}} = 104{,}7;$$

$$p = \frac{22}{38} = 0{,}58; \quad q = 0{,}42;$$

$$Z_{\text{табл}} = 0{,}3975;$$

$$p\cdot\frac{q}{Z} = 0{,}58\cdot\frac{0{,}42}{0{,}3975} = 0{,}61;$$

$$r = \frac{|318{,}8 - 390{,}9|}{104{,}7}\cdot 0{,}61 = 0{,}42.$$

Величина биссериального коэффициента корреляции также подтверждает умеренную тесноту связи между изучаемыми признаками.

9.8
НЕПАРАМЕТРИЧЕСКИЕ ПОКАЗАТЕЛИ СВЯЗИ. РАНГОВЫЕ КОЭФФИЦИЕНТЫ СВЯЗИ

В анализе социально-экономических явлений часто приходится прибегать к различным условным оценкам, например знакам отклонений рангов, а взаимосвязь между отдельными признаками измерять с помощью непараметрических коэффициентов связи. Данные коэффициенты исчисляются при условии, что исследуемые признаки подчиняются различным законам распределения.

Простейшим непараметрическим показателем тесноты связи между двумя признаками x и y является *коэффициент Фехнера*. В основе его расчета лежит принцип сопоставления не абсолютных значений признаков x и y, а их отклонений от среднего уровня.

Применение коэффициента Фехнера в практических расчетах основано на предположении, что отклонения эмпирических значений признака (x_i) от его средней величины ($x_i-\bar{x}$) носят случайный характер и должны случайным образом сочетаться с отклонениями эмпирических значений признака (y) от его среднего уровня ($y_i-\bar{y}$).

Соотношение пар совпадений или несовпадений знаков отклонений ($x_i-\bar{x}$) и ($y_i-\bar{y}$) позволяет судить о наличии и степени тесноты связи между x и y.

Коэффициент Фехнера ($K_ф$) определяется по формуле следующего вида:

$$K_{Ф} = \frac{С - H}{C + H},$$

где C – число совпадений знаков отклонений;
H – число несовпадений знаков отклонений.

Коэффициент Фехнера может принимать как положительные, так и отрицательные значения в пределах от (–1) до (+1), т.е. $-1 \le K_ф \le +1$.

При $K_ф = \pm 1$ связь между признаками x и y функциональная.

При $K_ф = 0$ связь отсутствует.

Промежуточные значения коэффициента Фехнера характеризуют степень тесноты связи между двумя признаками. Знак коэффициента Фехнера свидетельствует о направлении связи между двумя признаками.

Если $K_ф \in [-1; 0]$, то связь обратная, т.е. с увеличением или снижением x снижается или увеличивается y. Если $K_ф \in [0; +1]$, то связь прямая, т.е. с увеличением или снижением x увеличивается или снижается y.

Рассчитаем коэффициент Фехнера по данным табл. 9.7 о деятельности аудиторско-консультационных фирм Москвы в 2001 г., построив для этого табл. 9.13.

Таблица 9.13

Расчетные данные для определения коэффициента Фехнера

№ п/п	Совокупная выручка, млн руб., y	Общая численность профессионалов, чел., x	Знаки отклонений	
			$y_i - \bar{y}$	$x_i - \bar{x}$
1	2,62	23	–	–
2	3,04	32	–	–
3	3,15	50	–	–
4	3,83	53	+	–
5	3,58	55	–	+
6	4,08	58	+	+
7	4,09	59	+	+
8	4,20	62	+	+
9	4,18	69	+	+
10	4,24	75	+	+
Итого	37,01	536	–	–
Средняя	3,701	53,6	–	–

$$K_ф = \frac{8-2}{8+2} = 0,6.$$

Таким образом, связь между совокупной выручкой и численностью профессионалов аудиторско-консультационных фирм Москвы прямая и умеренная.

Недостатком коэффициента Фехнера, что значительно сокращает возможности его практической реализации, является равенство весов различных по абсолютной величине отклонений фактических значений признаков от их среднего уровня.

Ранговые коэффициенты связи. В качестве условных обозначений значений признаков и оценки связей между ними также используются ранги и ранговые коэффициенты связи.

Ранжирование – это процедура упорядочения объектов изучения, которая выполняется на основе предпочтения.

Ранг – это порядковый номер значений признака, расположенных в порядке возрастания или убывания их величин. Если значения признака имеют одинаковую количественную оценку, то ранг всех этих значений принимается равным средней арифметической от соответствующих номеров мест, которые определяют. Данные ранги называются *связными*.

Пример. Проранжируем предприятия автомобильной промышленности одного из регионов по величине балансовой прибыли (табл. 9.14).

Таблица 9.14

Балансовая прибыль предприятий автомобильной промышленности одного из регионов в 1998 г.*

Предприятие	Балансовая прибыль, млн руб.	Ранжирование (ранги)
1	10	6,5
2	12	4
3	10	6,5
4	12	4
5	12	4
6	15	2
7	17	1

* Цифры условные.

Наиболее предпочтительному предприятию, величина балансовой прибыли которого наибольшая, присваивается ранг «1»; затем в порядке уменьшения величины балансовой прибыли были проранжированы все рассматриваемые предприятия автомобильной промышленности.

Принцип нумерации значений исследуемых признаков является основой непараметрических методов изучения взаимосвязи между социально-экономическими явлениями и процессами.

Среди непараметрических методов оценки тесноты связи наибольшее значение имеют ранговые коэффициенты Спирмена (ρ) и Кендалла (τ). Эти коэффициенты могут быть использованы для определения тесноты связи как между количественными, так и между качественными признаками при условии, если их значения упорядочить или проранжировать по степени убывания или возрастания признака.

Коэффициент корреляции рангов (коэффициент Спирмена) рассчитывается по формуле (для случая, когда нет связных рангов)

$$\mathrm{c}_{x/y} = 1 - \frac{6\sum d_i^2}{n(n^2-1)}, \tag{9.45}$$

где d_i^2 – квадрат разности рангов;
n – число наблюдений (число пар рангов).

Коэффициент Спирмена принимает любые значения в интервале [–1; 1]. Значимость коэффициента корреляции рангов Спирмена проверяется на основе t-критерия Стьюдента. Расчетное значение критерия определяется по формуле

$$t_p = \mathrm{c}_{x/y} \cdot \sqrt{\frac{n-2}{1-\mathrm{c}_{x/y}^2}}. \tag{9.46}$$

Значение коэффициента корреляции считается статистически существенным, если $t_p > t_{kp}$ (α; $k = n–2$).

Пример. По данным группы предприятий одной из отраслей промышленности (табл. 9.15) определим с помощью коэффициента Спирмена зависимость между величиной балансовой прибыли и объемом реализованной продукции.

Таблица 9.15

Расчет коэффициента Спирмена[*]

Предприятие	Объем реализованной продукции, млрд руб. x	Балансовая прибыль, млн руб. y	Ранжирование			Сравнение рангов			Разность рангов $d_i = R_x - R_y$	d_i^2
			x	R_x	y	R_y	R_x	R_y		
1	1,8	20	1,3	1	20	1	2	1	1	1
2	2,3	75	1,8	2	42	2	3	3	0	0
3	8,6	42	2,3	3	75	3	10	2	8	64
4	1,3	80	3,5	4	80	4	1	4	–3	9
5	3,5	107	3,7	5	107	5	4	5	–1	1
6	3,8	125	3,8	6	125	6	6	6	0	0
7	4,5	140	4,5	7	140	7	7	7	0	0
8	5,8	175	5,8	8	175	8	8	8	0	0
9	3,7	200	6,5	9	200	9	5	9	–4	16
10	6,5	210	8,6	10	210	10	9	10	–1	1
Итого		–	–	–	–	–	–	–	–	92

[*] Данные условные

$$с_{xy} = 1 - \frac{6 \cdot 92}{10 \cdot 99} = 1 - \frac{552}{990} = 0,44.$$

Связь, близкая к умеренной.

Если совокупность значений по исследуемому признаку содержит связные ранги, то коэффициент корреляции Спирмена вычисляется по формуле

$$с_{xy} = \frac{\frac{1}{6} \cdot (n^3 - n) - \sum_{i=1}^{n} d_i^2 - T_x - T_y}{\sqrt{\left[\frac{1}{6} \cdot (n^3 - n) - 2Tx\right] \cdot \left[\frac{1}{6} \cdot (n^3 - n) - 2T_y\right]}}, \quad (9.47)$$

где $T_{xy} = \frac{1}{12} \cdot \sum_{j=1}^{k} (t_j^3 - t_j);$

t_j – число одинаковых рангов в j-м ряду.

На практике, если величины T_x и T_y несущественно отличаются относительно значения $\left[\frac{1}{6} \cdot (n^3 - n)\right]$, пользуются формулой

$$с_{xy} = 1 - \frac{\sum_{i=1}^{n} d_i^2}{\frac{1}{6} \cdot (n^3 - n) - (T_x + T_y)}. \quad (9.48)$$

Пример. По данным итогов торгов на биржевом рынке с 03.03.2002 г. – 09.03.2002 г. определим зависимость средней цены сделки от номинальной стоимости акции с помощью коэффициента Спирмена (табл. 9.16).

$$T_x = 1 / 12 \cdot [(5^3 - 5) + (3^3 - 3)] = 12;$$

$$T_y = 1 / 12 \cdot (2^3 - 2) = 0,5.$$

Применив формулу (9.47), получим:

$$\text{с}_{xy} = \frac{1/6\cdot(100-10)-23,5-12-0,5}{\sqrt{\left[1/6\cdot(1000-10)-2\cdot 12\right]\cdot\left[1/6\cdot(1000-10)-2\cdot 0,5\right]}}$$

$$= \frac{129}{\sqrt{141\cdot 164}} = 0,85.$$

Опыт проведения статистического анализа показывает, что формула 9.47 несущественно учитывает связные ранги при расчетах по формуле 9.45. Так, используя данные примера табл. 9.16, получаем, что

$$\text{с}_{xy} = 1-\frac{6\cdot 23,5}{10\cdot 99} = 0,85.$$

Но при этом вычисления по формуле (9.47) более громоздки. Поэтому на практике для расчета коэффициента Спирмена используют формулу (9.45) как для случая отсутствия, так и наличия связных рангов.

Результаты расчетов по формуле (9.48) подтвердили предыдущие:

$$\text{с}_{xy} = 1-\frac{23,5}{1/6\cdot(1000-10)-(12+0,5)} = 0,85,$$

что свидетельствует о сильной связи между рассматриваемыми признаками.

Ранговый коэффициент корреляции Кендалла (τ) может также использоваться для измерения взаимосвязи между качественными и количественными признаками, характеризующими однородные объекты, ранжированные по одному принципу. Расчет рангового коэффициента Кендалла осуществляется по формуле:

$$\text{ф}_{xy} = \frac{2S}{n(n-1)}, \tag{9.49}$$

где n – число наблюдений;

S – сумма разностей между числом последовательностей и числом инверсий по второму признаку.

Таблица 9.16

Расчет коэффициента Спирмена

Эмитент	Номинал, тыс. руб. x	Средняя цена сделки, тыс. руб. y	Ранжирование				Сравнение рангов		Разность рангов $d_i = R_x - R_y$	d_i^2
			x	R_x	y	R_y	R_x	R_y		
1	1,0	2,0	1,0	3	2,0	1	3	1	2	4
2	1,0	6,0	1,0	3	4,0	2,5	3	5	–2	4
3	1,0	4,0	1,0	3	4,0	2,5	3	2,5	0,5	0,25
4	1,0	4,0	1,0	3	5,7	4	3	2,5	0,5	0,25
5	2,5	7,8	1,0	3	6,0	5	6	6	0	0
6	10,0	16,0	2,5	6	7,8	6	9	8	1	1
7	10,0	10,8	5,0	7	10,8	7	9	7	2	4
8	5,0	20,0	10,0	9	16,0	8	7	10	–3	9
9	10,0	16,4	10,0	9	16,4	9	9	9	0	0
10	1,0	5,7	10,0	9	20,0	10	3	4	–1	1
Итого	–	–	–	–	–	–	–	–	–	23,5

Расчет данного коэффициента выполняется в следующей последовательности:

1) значения x ранжируются в порядке возрастания или убывания;

2) значения y располагаются в порядке, соответствующем значениям x;

3) для каждого ранга $У$ определяется число следующих за ним значений рангов, превышающих его величину. Суммируя таким образом числа, определяют величину P как меру соответствия последовательностей рангов по x и y и учитывают со знаком (+);

4) для каждого ранга y определяется число следующих за ним рангов, меньших его величины. Суммарная величина обозначается через Q и фиксируется со знаком (–);

5) определяется сумма баллов по всем членам ряда.

В приведенном примере (табл. 9.15):

$$P = 6 + 8 + 6 + 5 + 1 + 3 + 2 + 1 + 0 = 32;$$

$$Q = (-3) + 0 + (-1) + (-1) + (-4) + (-1) + (-1) + (-1) + (-1) = -13.$$

Таким образом,

$$ф_{xy} = \frac{2(32-13)}{10(10-1)} = 0,42,$$

что свидетельствует о наличии близкой к умеренной связи между рассматриваемыми признаками.

Если в изучаемой совокупности есть связные ранги, то расчеты необходимо проводить по следующей формуле:

$$ф_{xy} = \frac{S}{\sqrt{\left[n(n-1)/2 - V_x\right]\cdot\left[n(n-1)/2 - V_y\right]}}, \qquad (9.50)$$

где $V_{x/y} = 1/2\sum_{j=1}^{k} t_j \cdot (t_j - 1).$

По данным табл. 9.16 рассмотрим расчет коэффициента корреляции рангов Кендалла для случая наличия связных рангов:

$$P = 9+5+6+6+5+4+0+1+1 = 37;$$

$$Q = 0 + (-3) + 0 + 0 + 0 + 0 + (-3) + (-1) + 0 = -7;$$

$$S = 37 + (-7) = 30;$$

$$V_x = 1/2 \cdot [5 \cdot (5-1) + 3 \cdot (3-1)] = 13;$$

$$V_y = 1 / 2 \cdot [2 \cdot (2 - 1)] = 1;$$

$$ф_{xy} = \frac{30}{\sqrt{[10 \cdot (10-1)/2-13] \cdot [10 \cdot (10-1)/2-1]}} = \frac{30}{\sqrt{(45-13) \cdot (45-1)}} = 0,8,$$

что свидетельствует о существенной связи между номинальной стоимостью акций и средней ценой сделки на биржах. Как правило, коэффициент Кендалла меньше коэффициента Спирмена. При достаточно большом объеме совокупности значения данных коэффициентов имеют следующую зависимость:

$$ф_{xy} = \frac{2}{3} с_{xy}.$$

Связь между признаками можно признать статистически значимой, если значения коэффициентов ранговой корреляции Спирмена и Кендалла больше 0,5.

Для определения тесноты связи между произвольным числом ранжированных признаков применяется *множественный коэффициент ранговой корреляции (коэффициент конкордации) (W),* который вычисляется по формуле

$$W = \frac{12S}{m^2 \cdot (n^3 - n)},$$

где m – количество факторов;
n – число наблюдений;
S – отклонение суммы квадратов рангов от средней квадратов рангов.

Пример. Определим тесноту связи между уставным капиталом, числом выставленных акций и числом занятых на предприятиях, выставивших акции на аукционы в 2001 г. (табл. 9.17).

Таблица 9.17

Расчет коэффициента конкордации*

Номер предприятия	Σ ставный капитал, тыс. руб. x	Число выставленных акций y	Число занятых на предприятии z	R_x	R_y	R_z	Сумма строк	Квадраты сумм
1	29 540	856	119	9	7	1	17	289
2	16 050	930	125	1	9	2	12	144
3	41 020	1563	132	10	10	3	23	529
4	23 500	682	141	6	5	4	15	225
5	26 250	616	150	7	3	5	15	225
6	17 950	495	165	4	2	6	12	144
7	28 130	815	178	8	6	7	21	441
8	17 510	858	181	3	8	8	19	361
9	17 000	467	201	2	1	9	12	144
10	22 640	661	204	5	4	10	19	361
	–	–	–	–	–	–	165	2863

* Данные условные

$$S = 2863 - \frac{(165)^2}{10} = 2863 - 2722,5 = 140,5;$$

$$W = \frac{12S}{m^2(n^3 - n)} = \frac{12 \cdot 140,5}{(91\,000 - 10) = 0,19.}$$

Значимость коэффициента конкордации проверяется на основе c²-критерия Пирсона:

$$ч^2 = \frac{12S}{m \cdot n(n-1)}. \qquad (9.52)$$

Для нашего примера:

$$ч_\rho^2 = \frac{12 \cdot 140,5}{3 \cdot 10(10-1)} = 6,24.$$

Расчетное значение x_p^2 = 6,24 больше x_{kp}^2 = 16,919 (α = 0,05, $v = n - 1 = 9$), что подтверждает незначимость коэффициента конкордации и свидетельствует о слабой связи между рассматриваемыми признаками.

В случае наличия связных рангов коэффициент конкордации определяется по формуле

$$W = \frac{S}{1/12m^2 \cdot (n^3 - n) - m \cdot \sum_{j=1}^{m} Tj}, \qquad (9.53)$$

где $T_j = 1/2 \sum_{j=1}^{m} \left(t_j^3 - t_j\right)$;

t_j – количество связных рангов по отдельным показателям.

Проверка значимости осуществляется по формуле

$$ч^2 = \frac{S}{1/12m \cdot n(n-1) - \frac{1}{(n-1)} \cdot \sum_{j=1}^{m} T_j}. \qquad (9.54)$$

Коэффициент конкордации принимает любые значения в интервале [–1; 1].

Пример. По данным предприятий нефтеперерабатывающей промышленности определиv зависимость прибыли от реализации, от среднегодовой стоимости основных производственных фондов и объема валовой продукции (табл. 9.17).

$$S = 1229{,}5 - \frac{(84)^2}{7} = 1229{,}5 - 1008 = 221{,}5;$$

$$T_x = \frac{1}{12}\left[(2^3-2)+(2^3-2)\right] = 1;$$

$$T_y = \frac{1}{12}\left[(2^3-2)+(2^3-2)\right] = 1;$$

$$T_z = \frac{1}{12}\left[(2^3-2)+(2^3-2)\right] = 1;$$

$$\sum T_j = T_x + T_y + T_z = 1+1+1 = 3.$$

$$W = \frac{S}{1/12m^2(n^2-n) - m\sum T_j} = \frac{221{,}5}{1/12\cdot 3^2(7^3-7)-3\cdot 3} = \frac{221{,}5}{243{,}0} = 0{,}91.$$

Расчетное значение χ^2-критерия Пирсона для проверки значимости коэффициента конкордации по данным нашего примера составило:

$$\chi_p^2 = \frac{221{,}5}{1/12\cdot 3\cdot 7\cdot(7-1) - \frac{1}{7-1}\cdot 3} = 22{,}15.$$

Расчетное значение χ_p^2 = 22,15 больше χ_{kp}^2 = 12,592, (α = 0,05, $v = n - 1 = 6$), что подтверждает значимость коэффициента конкордации и свидетельствует о сильной связи между рассматриваемыми признаками.

Преимуществом ранговых коэффициентов корреляции Спирмена, Кендалла и конкордации является то, что с их помощью можно измерять и оценивать связи как между количественными, так и между атрибутивными признаками, которые поддаются ранжированию.

Таблица 9.18

Расчет коэффициента конкордации*

Предприятие	Прибыль от реализации, млн руб. x	Объем валовой продукции, млрд руб. y	Среднегодовая стоимость основных производственных фондов, млрд руб. z	R_x	R_y	R_z	Сумма строк	Квадраты сумм
1	40	1,7	0,27	1,5	1	1,5	4	16
2	75	3,2	0,55	3	5	4	12	144
3	82	2,9	0,97	4,5	3,5	5,5	13,5	182,25
4	40	1,8	0,27	1,5	2	1,5	5	25
5	106	11,8	0,98	6	6,5	7	19,5	380,25
6	82	2,9	0,35	4,5	3,5	3	11	121
7	109	11,8	0,97	7	6,5	5,5	19	361
Итого	–	–	–	28	28	28	84	1229,5

* Данные условные

ОСНОВНЫЕ ПОНЯТИЯ

Биссериальный коэффициент корреляции – оценивание связи между качественным альтернативным и количественным варьирующим признаками.

Корреляционная связь – изменение среднего значения результативного признака, которое обусловливается изменением факторных признаков.

Корреляционное отношение показывает связь между двумя признаками.

Корреляция – статистическая зависимость между случайными величинами, которая не имеет строго функционального характера, но изменение одной из случайных величин приводит к изменению математического ожидания другой.

Коэффициент взаимной сопряженности Пирсона–Чупрова определение тесноты связи двух качественных признаков, каждый из которых состоит более чем из двух групп.

Коэффициент детерминации показывает, на сколько процентов вариация результативного признака объясняется вариацией i-го признака (частный) или всех вошедших в модель факторных признаков (множественный).

Коэффициент конкордации определяет тесноту связи между произвольным числом ранжированных признаков.

Коэффициент регрессии a_i показывает, на сколько в среднем изменяется значение результативного признака при изменении факторного на единицу собственного измерения.

Коэффициенты ассоциации и контингенции определяют тесноту связи двух качественных признаков, каждый из которых состоит только из двух групп.

Коэффициенты корреляции Спирмена и Кендалла определяют тесноту связи между двумя количественными или качественными признаками после предварительного ранжирования их по возрастанию или убыванию.

Коэффициент эластичности показывает, на сколько процентов в среднем изменится значение результативного признака при изменении факторного признака на 1%.

Линейная связь – статистическая связь между явлениями, выраженная уравнением прямой линии.

Линейный коэффициент корреляции определяет тесноту и направление связи между двумя коррелируемыми признаками.

Множественная регрессия – модель связи трех и более признаков.

Множественный коэффициент корреляции отражает связь между результативным и несколькими факторными признаками.

Мультиколлинеарность — наличие тесной зависимости между факторными признаками.

Нелинейная связь – статистическая связь между социально-экономическими явлениями, аналитически выраженная уравнением кривой линии (параболы, гиперболы и т. д.).

Обратная связь – с увеличением или уменьшением значений факторного признака уменьшается или увеличивается значение результативного.

Парная регрессия – аналитическое выражение связи двух признаков.

Признак – основная отличительная черта, особенность изучаемого явления или процесса.

Причинно-следственные отношения – связь явлений и процессов, когда изменение одного из них – причины – ведет к изменению другого – следствия. Социально-экономические явления – это результат одновременного воздействия большого числа причин.

Прямая связь – с увеличением или уменьшением значений факторного признака увеличивается или уменьшается значение результативного.

Ранг – порядковый номер значения признака, расположенного в порядке возрастания или убывания величин.

Ранжирование – процедура упорядочения объектов изучения, которая выполняется на основе предпочтения значений признака в порядке возрастания или убывания.

Регрессионный анализ – аналитическое выражение связи, в котором изменение одной величины – результативного признака – обусловлено влиянием одной или нескольких независимых величин (факторов), а множество всех прочих факторов, также оказывающих влияние на зависимую величину, принимается за постоянные и средние значения.

Результативный признак – признак, изменяющийся под действием факторных признаков.

Стохастическая связь – связь, которая проявляется не в каждом отдельном случае, а в общем, среднем или большом числе наблюдений.

Факторный признак – признак, оказывающий влияние на изменение результативного.

Функциональная связь – связь, при которой определенному значению факторного признака соответствует одно и только одно значение результативного признака.

Частный коэффициент корреляции показывает степень тесноты связи между двумя признаками при фиксированном значении остальных факторных признаков.

Экономическая интерпретация модели – основные выводы и заключения на основе расчета и анализа частных коэффициентов эластичности, частных и множественного коэффициентов детерминации, Q-коэффициента.

ТЕСТЫ

1. По направлению связи бывают:

а) умеренные;

б) прямые;

в) прямолинейные.

2. По аналитическому выражению связи различаются:

а) обратные;

б) тесные;

в) криволинейные.

3. Функциональной является связь:

а) между двумя признаками;

б) при которой определенному значению факторного признака соответствует несколько значений результативного признака;

в) при которой определенному значению факторного признака соответствует одно значение результативного признака.

4. Аналитическое выражение связи определяется с помощью метода анализа:

а) корреляционного;

б) регрессионного;

в) группировок.

5. Анализ тесноты и направления связей двух признаков осуществляется на основе:

а) парного коэффициента корреляции;

б) частного коэффициента корреляции;

в) множественного коэффициента корреляции.

6. Мультиколлинеарность – это связь между:

а) признаками;

б) уровнями;

в) явлениями.

7. Оценка значимости параметров модели регрессии осуществляется на основе:

а) коэффициента корреляции;

б) средней ошибки аппроксимации;

в) t-критерия Стьюдента.

8. Оценка значимости уравнения регрессии осуществляется на основе:

а) коэффициента детерминации;

б) средней квадратической ошибки;

в) F-критерия Фишера.

9. Оценка связей социальных явлений производится на основе:

а) коэффициента ассоциации;

б) коэффициента контингенции;

в) коэффициента эластичности.

10. Коэффициент корреляции рангов Спирмена можно применять для оценки тесноты связи между:

а) количественными признаками;

б) качественными признаками, значения которых могут быть упорядочены;

в) любыми качественными признаками.

ЛИТЕРАТУРА

Айвазян С.А., Енюков И.С., Мешалкин Л.Д. Прикладная статистика. Основы моделирования и первичной обработки данных. – М.: Финансы и статистика, 1983. – 471 с.

Айвазян С.А., Енюков И.С., Мешалкин Л.Д. Прикладная статистика. Исследование зависимостей. – М.: Финансы и статистика, 1985. – 485 с.

Дрейпер Н., Смит Г. Прикладной регрессионный анализ. Кн. 1. – М.: Финансы и статистика, 1986. – 366 с.

Дрейпер Н., Смит Г. Прикладной регрессионный анализ. Кн. 2. – М.: Финансы и статистика, 1987. – 351 с.

Езекиэл М., Фокс К.А. Методы анализа корреляций и регрессий линейных и криволинейных. – М.: Статистика, 1966. – 558 с.

Ефимова М.Р., Петрова Е.В., Румянцев В.П. Общая теория статистики: Учебник. – М.: Инфра-М, 1997.

*Королев Ю.*Г. Метод наименьших квадратов в социально-экономических исследованиях. – М.: Статистика, 1980. –112 с.

Общая теория статистики: Учебник для вузов по специальности «Статистика» / Под ред Г.С. Кильдишева, В.Е. Овсиенко, П.М. Рабиновича, Т.В. Рябушкина. – М.: Статистика, 1980. –423 с.

Общая теория статистики / Под ред. А.Я. Боярского, Г.Л. Громыко. – 2-е изд. – М.: Изд-во МГУ, 1985. – 376 с.

Сиськов В.И. Корреляционный анализ в экономических исследованиях. – М.: Статистика, 1975. – 168 с.

Фестер Э., Ренц Б. Методы корреляционного и регресионного анализа. – М.: Финансы и статистика, 1983. – 302 с.

Френкель А.А., Адамова Е.В. Корреляционный и регресионный анализ в экономических приложениях: Учебное пособие. – М.: МЭСИ, 1987. – 96 с.

ГЛАВА 10

СТАТИСТИЧЕСКОЕ ИЗУЧЕНИЕ ДИНАМИКИ СОЦИАЛЬНО-ЭКОНОМИЧЕСКИХ ЯВЛЕНИЙ

10.1 ПОНЯТИЕ И КЛАССИФИКАЦИЯ РЯДОВ ДИНАМИКИ

Процесс развития социально-экономических явлений во времени в статистике принято называть динамикой. Для отображения динамики строят *ряды динамики* (хронологические, временные), которые представляют собой ряды изменяющихся во времени значений статистического показателя, расположенных в хронологическом порядке. В динамическом ряду процесс экономического развития изображается в виде совокупности перерывов непрерывного, позволяющих детально проанализировать особенности развития при помощи характеристик, которые отражают изменение параметров экономической системы во времени.

Составными элементами ряда динамики являются показатели уровней ряда и периоды времени (годы, кварталы, месяцы, сутки) или моменты (даты) времени.

Уровни ряда обычно обозначаются через y, моменты или периоды времени, к которым относятся уровни, – через t.

Существуют различные виды рядов динамики. Их можно классифицировать по следующим признакам.

1. *В зависимости от способа выражения уровней, ряды динамики подразделяются на ряды абсолютных, относительных и средних величин.*

В табл. 10.1 рядом динамики абсолютных величин являются данные первой строки; рядом средних величин – второй строки; рядом относительных величин – третьей строки.

2. В зависимости от того, как выражают уровни ряда состояние явления на определенные моменты времени (начало месяца, квартала, года и т.п.) или его величину за определенные интервалы времени (например, за сутки, месяц, год и т.п.), различают соответственно моментные и интервальные ряды динамики.

Таблица 10.1

Число квартир, построенных предприятиями и организациями всех форм собственности и их средний размер в РФ

Показатели	1980	1985	1992	1995	2000
1. Число квартир, тыс.	1190	1151	682	602	373
2. Средний размер квартир, $м^2$ общей площади	49,9	54,4	60,8	68,2	81,1
3. Удельный вес однокомнатных квартир от общего ввода, %	18	18	18	18	20

Пример. Моментный ряд – это ряд динамики, показывающий число кредитных организаций, зарегистрированных на территории РФ (на начало года):

1998	1999	2000	2001
2555	2483	2378	2126

Уровни этого ряда – обобщающие итоги статистики числа зарегистрированных кредитных организаций по состоянию на определенную дату (на начало каждого года).

Пример. Интервальный ряд динамики приведен в табл. 10.1.

Из различного характера интервальных и моментных рядов динамики вытекают некоторые особенности уровней соответствующих рядов.

Уровни интервального ряда динамики абсолютных величин характеризуют собой суммарный итог какого-либо явления за определенный отрезок времени. Они зависят от продолжительности этого периода времени, и поэтому их *можно суммировать как не содержащие повторного счета.*

Отдельные же уровни *моментного ряда динамики* абсолютных величин *содержат элементы повторного счета*, например, число вкладов населения, учитываемых за январь, существуют в настоящее время и являются единицами совокупности и в июне. Все это делает бессмысленным суммирование уровней моментных рядов динамики.

3. *В зависимости от расстояния между уровнями ряды динамики подразделяются на ряды динамики с равноотстоящими уровнями* и *неравноотстоящими уровнями во времени.*

Ряды динамики следующих друг за другом периодов или следующих через определенные промежутки дат называются *равноотстоящими* (см. пример о числе кредитных организаций, зарегистрированных на территории РФ. Если же в рядах даются прерывающиеся периоды или неравномерные промежутки между датами, то ряды называются *неравноотстоящими* (см. табл. 10.1).

4. *В зависимости от наличия основной тенденции изучаемого процесса ряды динамики подразделяются на стационарные* и *нестационарные*.

Если математическое ожидание значения признака и дисперсия (основные характеристики случайного процесса) постоянны, не зависят от времени, то процесс считается *стационарным* и ряды динамики также называются *стационарными.* Экономические процессы во времени обычно не являются стационарными, так как содержат основную тенденцию развития, но их можно преобразовать в стационарные путем исключения тенденций.

5. *По числу показателей можно выделить изолированные и комплексные (многомерные) ряды динамики.* Если ведется анализ во времени одного показателя, то ряд динамики *изолированный* (табл. 10.1). В *многомерном ряду* представлена динамика нескольких показателей, характеризующих одно явление (табл. 10.2).

Таблица 10.2

Динамика продукции сельского хозяйства РФ за 1997–2000 гг., млн руб.*

Продукция сельского хозяйства	1997	1998	1999	2000
Всего	309 217	307 583	611 993	781 576
в том числе:				
растениеводства	171 486	152 289	327 992	426 581
животноводства	137 731	155 294	284 001	354 995

*До 1998 г. – млрд руб.

На рис. 10.1 показана рассмотренная выше классификация рядов динамики.

Рис. 10.1. Классификация видов рядов динамики

10.2 СОПОСТАВИМОСТЬ УРОВНЕЙ И СМЫКАНИЕ РЯДОВ ДИНАМИКИ

Важнейшим условием правильного построения ряда динамики является сопоставимость всех входящих в него уровней. Данное условие решается либо в процессе сбора и обработки данных, либо путем их пересчета.

Проблема сопоставимости данных особенно остро стоит в рядах динамики, потому что они охватывают значительные периоды времени, за которые могли произойти изменения и привести к несопоставимости статистических данных. Рассмотрим основные причины несопоставимости уровней ряда динамики.

Несопоставимость уровней ряда может возникнуть *вследствие изменения единиц измерения или единиц счета.* Нельзя, например, сравнивать и анализировать цифры о производстве тканей, если за одни годы цифры даны в погонных метрах, а за другие – в квадратных метрах.

На сопоставимость уровней ряда динамики непосредственно влияет *методология учета* или *расчета показателей.* Например, если в одни годы среднюю урожайность считали с засеянной площади, а в другие – с убранной, то такие уровни будут несопоставимы.

Условием сопоставимости уровней ряда динамики является *периодизация динамики.* В процессе развития во времени прежде всего происходят количественные изменения явлений, а затем на определенных ступенях совершаются качественные скачки, приводящие к изменению закономерности развития явления. Поэтому научный подход к изучению рядов динамики заключается в том, чтобы ряды, охватывающие большие периоды времени, разбивать на такие, которые бы объединяли лишь однокачественные периоды развития совокупности, характеризующейся одной закономерностью развития.

Процесс выделения однородных этапов развития рядов динамики носит название *периодизации динамики.* Вопрос о том, какие этапы развития прошло то или иное явление за определенный исторический отрезок времени, решается теорией той науки, к области которой относится изучаемая совокупность явления.

Необходимость формировать ряды динамики по строго однородным периодам, или этапам, не означает отрицания возможности построения и изучения рядов динамики, охватывающих длительные исторические отрезки времени, включающие различные этапы развития явления. Нужно помнить, что само понятие однородности периодов весьма относительно, оно зависит от уровня абстракции, принятой в исследовании. Например, весь советский период развития России, несомненно, является особым однородным периодом, кардинально отличающимся от предыдущего развития нашей страны. Внутри советского периода, в свою очередь, можно выделить более короткие, однородные в определенном отношении интервалы времени – довоенные годы, годы Великой Отечественной войны, послевоенные годы восстановления народного хозяйства и т.д. Если ряды динамики не готовятся непосредственно для анализа, а играют чисто информационную роль, они могут быть непериодизированы.

Важно также, чтобы в ряду динамики *интервалы,* или *моменты*, по которым определены уровни, имели *одинаковый экономический смысл.* Скажем, при изучении роста поголовья скота бессмысленно сравнивать цифры поголовья по состоянию на 1 октября с 1 января, так как первая цифра включает не только скот, оставшийся на зимовку, но и предназначенный к убою, а вторая цифра включает только скот, оставленный на зимовку.

Условием сравнимости уровней интервального ряда является *наличие равных интервалов*, по которым даны уровни. Совершенно очевидно, что нельзя сравнивать квартальную продукцию с годовой.

Уровни ряда динамики могут оказаться *несопоставимыми по кругу охватываемых объектов* вследствие перехода ряда объектов из одного подчинения в другое.

Несопоставимость уровней ряда может возникнуть вследствие *изменения территориальных границ* областей, районов и т.д. При этом, говоря об изменении территории, к которой относятся уровни ряда за разное время, следует иметь в виду, что вопрос о сопоставимости или несопоставимости при изменении территории решается по-разному, в зависимости от цели исследования. Если, например, ставится задача показать изменение численности населения или объема промышленного производства в связи с изменением административно-территориальных границ области или района, то не только можно, но и нужно сопоставлять данные в фактических границах этой области или района. Если же изучаются показатели темпов естественного прироста населения или темпов развития промышленности, то, очевидно, сравниваемые показатели должны относиться к одним и тем же территориальным границам. Аналогичные проблемы возникают при анализе отдельных городов и даже государства в целом, если меняются административно-территориальные границы.

Следовательно, прежде чем анализировать динамический ряд, надо, исходя из цели исследования, убедиться в сопоставимости уровней ряда и при отсутствии последней добиваться ее, пользуясь дополнительными расчетами.

Приведение уровней ряда к сопоставимому виду. Данный прием осуществляется методом *смыкания рядов динамики*. Под смыканием понимают объединение в один ряд (более длинный) двух или нескольких рядов динамики, уровни которых исчислены по разной методологии или разным территориальным границам. Для осуществления смыкания необходимо, чтобы для одного из периодов (переходного) имелись данные, исчисленные по разной методологии (или в разных границах).

Пример. По одному из промышленных объединений имеются данные о произведенной продукции, методика получения которых в течение рассматриваемого периода претерпела некоторые изменения (табл. 10.3). Для анализа динамики объема продукции за 1994–2001 гг. необходимо сомкнуть (объединить) исследуемые два ряда в один. А чтобы уровни нового ряда были сопоставимы, следует пересчитать данные 1994–1997 гг. по новой методике. Для этого на основе данных об объеме продукции за 1997 г. в новой и старой методике находим соотношение между ними: 22,8 : 21,2 = 1,1. Умножая на полученный коэффициент данные за 1994–1996 гг., приводим их таким образом в сопоставимый вид с последующими уровнями. Сомкнутый (сопоставимый) ряд динамики показан в предпоследней строке табл. 10.3.

Таблица 10.3

Динамика объема продукции*

Показатели	1994	1995	1996	1997	1998	1999	2000	2001
Объем продукции, млн руб.:								
по старой методике	19,1	19,7	20,0	21,2	–	–	–	–
по новой методике	–	–	–	22,8	23,6	24,5	26,2	28,1
Сомкнутый (сопоставимый) ряд абсолютных величин, млн руб.	21,0	21,7	22,0	22,8	23,6	24,5	26,2	28,1
Сопоставимый ряд относительных величин, % к 1997 г.	90,1	92,9	94,3	100,0	103,5	107,5	114,9	123,2

* Цифры условные.

Другой способ смыкания рядов динамики заключается в том, что уровни года, в котором произошли изменения (в нашем примере уровни 1997 г.), как до изменений, так и после изменений (в старой и новой методике, т.е. 21,2 и 22,8) принимаются за 100%, а остальные пересчитываются в процентах по отношению к этим уровням соответственно (в старых ценах – по отношению к 21,2, в новых ценах – к 22,8). В результате получаем сомкнутый ряд динамики, который показан в последней строке табл. 10.3.

Приведение рядов динамики к одному основанию. Та же проблема приведения к сопоставимому виду возникает и при параллельном анализе развития во времени экономических показателей отдельных стран, административных и территориальных районов. Это, во-первых, вопрос о сопоставимости цен сравниваемых стран, во-вторых, о сопоставимости методики расчета сравниваемых показателей. В таких случаях ряды динамики приводятся к *одному основанию,* т.е. к одному и тому же периоду или моменту времени, уровень которого принимается за базу сравнения, а все остальные уровни выражаются в виде коэффициентов или в процентах по отношению к нему.

Пример. Имеются данные о числе построенных квартир в двух странах (табл. 10.4).

Таблица 10.4

Число построенных квартир за 1996–2000 гг., тыс.

Страна	1996	1997	1998	1999	2000
Россия	482	430	388	340	373
Беларусь	38	46	48	34	39

Различные значения абсолютных уровней приведенных рядов динамики затрудняют выявление особенностей роста числа квартир в двух странах. Поэтому приведем абсолютные уровни рядов динамики *к общему основанию,* приняв за постоянную базу сравнения уровни 1996 г.; получим следующие данные (табл. 10.5).

В относительных величинах, выраженных в базисных темпах роста по каждой стране, несопоставимость уровней рядов динамики нивелируется. Различный характер развития выступает более наглядно.

Таблица 10.5

Число построенных квартир за 1996–2000 гг., % к 1996 г.

Страна	1996	1997	1998	1999	2000
Россия	100,0	92,1	80,5	70,5	77,4
Беларусь	100,0	121,1	126,3	89,5	102,3

Коэффициент опережения (замедления). Чтобы ответить на вопрос, во сколько раз число построенных квартир больше в Беларуси по сравнению с Россией, необходимо сравнить базисные коэффициенты роста за изучаемый период, т.е. вычислить *коэффициент опережения* (*замедления*) – K_o:

$$K_o = \frac{T_2}{T_1},\ \text{где}\ (T_2 > T_1),$$

или

$$K_o = \frac{T_1}{T_2},\ \text{где}\ (T_1 > T_2).$$

Эту формулу удобнее применять в том случае, когда ряд представляет постоянное повышение. Для рядов, где нет ярко выраженной тенденции к росту, удобнее за основание к приведению рядов брать средние показатели рядов динамики, в частности, средние темпы роста:

$$K_o = \frac{\overline{T}_2}{\overline{T}_1},\ \text{или}\ K_o = \frac{\overline{T}_1}{\overline{T}_2}.$$

Для нашего примера (см. табл. 10.4) средние темпы роста для Беларуси – 100,33%, для России – 96,41%. Таким образом, коэффициент опережения будет равен 1,04 раза (1,0033 : 0,9641), т.е. в 2000 г. по сравнению с 1996 г. число построенных квартир в Беларуси было в 1,04 раза больше, чем в России.

10.3 АНАЛИТИЧЕСКИЕ ПОКАЗАТЕЛИ ИЗМЕНЕНИЯ УРОВНЕЙ РЯДА ДИНАМИКИ

Анализ скорости и интенсивности развития явления во времени осуществляется с помощью аналитических показателей, которые получаются в результате сравнения уровней ряда динамики между собой. К таким показателям относятся: абсолютный прирост, темп роста и прироста, абсолютное значение 1% прироста. При этом принято сравниваемый уровень называть *отчетным*, а уровень, с которым производят сравнение, – *базисным*.

Аналитические показатели ряда динамики. *Абсолютный прирост* (Δy) характеризует размер увеличения (или уменьшения) уровня ряда за определенный промежуток времени. Он равен разности двух сравниваемых уровней и выражает абсолютную скорость роста:

$$\Delta_{yi} = y_i - y_{i-k}, \tag{10.1}$$

где $i = 1, 2, 3, \ldots, n$.

Если $k = 1$, то уровень y_{i-1} является предыдущим для данного ряда, а абсолютные приросты изменения уровня будут цепными. Если же k постоянно для данного ряда, то абсолютные приросты будут базисными.

Интенсивность изменения уровня оценивается отношением отчетного уровня к базисному, которое всегда представляет собой положительное число.

Показатель интенсивности изменения уровня ряда в зависимости от того, выражается ли он в виде коэффициента или в процентах, принято называть *коэффициентом роста* или *темпом роста.* Иными словами, коэффициент роста и темп роста представляют собой две формы выражения интенсивности изменения уровня. Однако необходимо отметить, что не нужно пользоваться одновременно двумя формами, которые по существу идентичны. Разница между ними заключается только в единице измерения.

Коэффициент роста показывает, во сколько раз данный уровень ряда больше базисного уровня (если этот коэффициент больше единицы) или какую часть базисного уровня составляет уровень текущего периода за некоторый промежуток времени (если он меньше единицы). В качестве базисного уровня в зависимости от цели исследования может приниматься какой-то постоянный для всех уровень (часто начальный уровень ряда) либо для каждого последующего предшествующий ему:

$$Tp_{i/1} = \frac{y_i}{y_1} \cdot 100 \text{ или } Tp_{i/i-1} = \frac{y_i}{y_{i-1}} \cdot 100. \quad (10.2)$$

В первом случае говорят о базисных темпах роста, во втором – о цепных темпах роста.

Наряду с темпом роста можно рассчитать показатель темпа прироста, характеризующий относительную скорость изменения уровня ряда в единицу времени. *Темп прироста* показывает, на какую долю (или процент) уровень данного периода или момента времени больше (или меньше) базисного уровня.

Темп прироста есть отношение абсолютного прироста к уровню ряда, принятого за базу:

$$T_{\text{пр}_{i/i-1}} = \frac{\Delta_{i/i-1}}{y_{i-1}} = \frac{y_i - y_{i-1}}{y_{i-1}} \cdot 100 = (Kp_{i/i-1} - 1) \cdot 100 = Tp_{i/1-1} - 100. \quad (10.3)$$

Если темп роста всегда положительное число, то темп прироста может быть положительным, отрицательным и равным нулю.

В статистической практике часто вместо расчета и анализа темпов роста и прироста рассматривают *абсолютное значение одного процента прироста.* Оно представляет собой одну сотую часть базисного уровня и в то же время – отношение абсолютного прироста к соответствующему темпу прироста:

$$|\%| = \frac{\Delta_{i/i-1}}{T_{\text{пр}_{i/i-1}}} = \frac{y_i - y_{i-1}}{\frac{y_i - y_{i-1}}{y_i} \cdot 100} = \frac{y_{i-1}}{100} = 0{,}01 \cdot y_{i-1}, \quad (10.4)$$

где |%| – обозначение абсолютного значения 1% прироста.

Абсолютное значение 1% прироста служит косвенной мерой базисного уровня и вместе с темпом прироста позволяет рассчитать абсолютный прирост уровня за рассматриваемый период, т.е. он показывает, сколько абсолютных единиц приходится на 1% прироста (уменьшения).

Абсолютным ускорением в статистике называется разность между последующим и предыдущим абсолютными приростами ($\Delta' = \Delta_{yi} - \Delta_{yi-1}$). Ускорение показывает, насколько данная скорость больше (меньше) предыдущей.

Таким образом, абсолютное ускорение есть скорость изменения скорости. Оно может быть положительным и отрицательным числом.

Относительным ускорением называется отношение абсолютного ускорения к абсолютному приросту, принятому за базу (Δ' / Δ_{yi}), т.е. относительное ускорение есть темп прироста абсолютного прироста. Оно вычисляется лишь в том случае, если абсолютный прирост, принятый за базу сравнения, число положительное.

Пример. Для ряда 30, 33, 35, 39, 44 абсолютные приросты составят 3, 2, 4, 5; абсолютные ускорения – 1, 2, 1; относительные ускорения $(-1/3) \cdot 100 = -33{,}3\%$; $2/2 \cdot 100 = 100\%$; $1/4 \cdot 100 = 25\%$.

Пример. Для иллюстрации расчетов представленных аналитических показателей приведем следующий ряд динамики в табл. 10.6.

Средние обобщающие показатели ряда динамики. *Средний уровень* ряда динамики ($\bar{y}$) рассчитывается по средней хронологической. *Средней хронологической* называется средняя, исчисленная из значений, изменяющихся во времени. Такие средние обобщают хронологическую вариацию. В хронологической средней отражается совокупность тех условий, в которых развивалось изучаемое явление в данном промежутке времени.

Методы расчета среднего уровня интервального и моментного рядов динамики различны. Для интервальных рядов с равноотстоящими уровнями средний уровень находится по формуле средней арифметической простой, а для неравноотстоящих уровней – по средней арифметической взвешенной:

$$\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n}; \qquad (10.5)$$

$$\bar{y} = \frac{\sum_{i=1}^{n} y_i \cdot t_i}{\sum t_i}, \qquad (10.6)$$

где y_i – уровень ряда динамики;
n – число уровней;
t_i – длительность интервала времени между уровнями.

Таблица 10.6

Динамика производства легковых автомобилей за 1995–2001 гг. по стране

Год	Произведено, тыс. шт.	Абсолютный прирост, тыс. шт.		Темп роста, %		Темп прироста, %		Абсолютное значение 1% прироста, тыс. шт.
		по сравнению с предыдущим годом	по сравнению с 1995 г.	по сравнению с предыдущим годом	по сравнению с 1995 г.	по сравнению с предыдущим годом	по сравнению с 1995 г.	
1995	835,1	–	–	–	100,0	–	0,0	–
1996	867,4	32,3	32,3	103,87	103,87	3,87	3,87	8,35
1997	986,2	118,8	151,1	113,7	118,09	13,7	18,09	8,67
1998	836,0	–152,2	0,9	84,77	100,1	–28,9	0,1	5,27
1999	955,5	119,5	120,4	114,29	114,42	14,29	14,42	8,36
2000	969,0	13,5	133,9	101,41	116,03	1,41	16,03	9,57
2001	1000,0	31,0	164,9	103,2	119,75	3,2	19,75	9,69
Итого	6449,2	164,9	–	–	–	–	–	–

Пример. В табл. 10.6 приведен *интервальный ряд динамики с равноотстоящими уровнями*. По этим данным можно рассчитать среднегодовой уровень производства автомобилей за 1995–2001 гг. Он будет равен 921,3 тыс. шт., т.е. $\bar{y} = \frac{6449,2}{7}$. Это означает, что за период 1995 – 2001 гг. ежегодно производство автомобилей составляло 921,3 тыс. шт.

Средний уровень моментного ряда динамики так исчислять нельзя, так как отдельные уровни содержат элементы повторного счета. *Средний уровень моментного равноотстоящего ряда динамики* находится по формуле средней хронологической простой:

$$\bar{y} = \frac{\frac{y_1+y_2}{2}+\frac{y_2+y_3}{2}+\frac{y_3+y_4}{2}+\ldots+\frac{y_{n-1}+y_n}{2}}{n-1} =$$

$$= \frac{\frac{y_1}{2}+y_2+y_3+\ldots+y_{n-1}+\frac{y_n}{2}}{n-1} \quad \text{или} \quad \bar{y} = \frac{\frac{y_1+y_n}{2}+\sum_{i=2}^{n-1} y_i}{n-1}. \tag{10.7}$$

Средний уровень моментных рядов динамики с неравноотстоящими уровнями определяется по формуле средней хронологической взвешенной:

$$\bar{y} = \frac{(y_1+y_2)\cdot t_1+(y_2+y_3)\cdot t_2+\ldots+(y_{n-1}+y_n)\cdot t_{n-i}}{2\cdot(t_1+t_2+t_3+\ldots+t_{n-1})} =$$

$$= \frac{\sum(y_i+y_{i+1})\cdot t_1}{2\cdot\sum t_i}, \tag{10.8}$$

где y_i, y_n – уровни рядов динамики;
t_i – длительность интервала времени между уровнями.

Покажем расчет среднего уровня моментного ряда динамики.

Пример. Известны товарные остатки магазина на 1-е число каждого месяца, тыс. руб.:

1.I	1.II	1.III	1.IV
18	14	16	20

Среднемесячный товарный остаток за I квартал по формуле (10.7) составит:

$$\bar{y} = \frac{18/2 + 14 + 16 + 20/2}{3} = \frac{49}{3} = 16,3 \text{ тыс.руб.}$$

Этот же показатель можно получить иначе. При вычислении среднего уровня моментного ряда условно предполагается непрерывное, равномерное изменение уровня в промежутках между двумя датами. Основываясь на этом предположении, определим средние товарные остатки магазина на каждый месяц как полусумму остатков на начало и конец месяца. Средние товарные остатки за месяц, тыс. руб.:

$$\text{январь} = \frac{18+14}{2} = 16,$$

$$\text{февраль} = \frac{14+16}{2} = 15,$$

$$\text{март} = \frac{16+20}{2} = 18.$$

Среднемесячный товарный остаток магазина за I квартал в данном случае определяется как простая средняя арифметическая:

$$\bar{y} = \frac{16+15+18}{3} = 16,3 \text{ тыс. руб.}$$

Пример. Известна списочная численность рабочих организации на некоторые даты 2001 г., чел.:

1.I	1.III	1.VI	1.IX	1.I.2002 г.
1200	1100	1250	1500	1350

Среднегодовая численность работников за 2001 г. по формуле (10.8) составит:

$$\bar{y} = \frac{(1200+1100)2+(1100+1250)3+(1250+1500)3+(1500+1350)4}{2(2+3+3+4)} =$$

$$= \frac{31300}{24} \approx 1304 \text{ чел.}$$

Обобщающим показателем скорости изменения явления во времени является *средний абсолютный прирост* ($\bar{\Delta}$).

Этот показатель дает возможность установить, насколько в среднем за единицу времени должен увеличиваться уровень ряда (в абсолютном выражении), чтобы, отправляясь от начального уровня за данное число периодов (например, лет), достигнуть конечного уровня. Определяющим свойством интересующего нас показателя среднего абсолютного прироста при такой постановке задачи является общий абсолютный прирост за весь период, ограничивающий ряд динамики. Для его определения воспользуемся формулой средней арифметической простой:

$$\bar{\Delta}_y = \frac{\sum_{i=1}^{n-1} \Delta_{i/i-1}}{n-1}$$

или

$$\bar{\Delta}_y = \frac{y_n - y_1}{n-1}. \qquad (10.9)$$

Пример. Средний абсолютный прирост рассчитаем по данным табл. 10.6. Он равен $\bar{\Delta}_y = \frac{164,9}{6}$ = 27,48 тыс. шт., т.е. за период с 1995–2001 гг. в среднем ежегодно абсолютный прирост производства автомобилей составлял 27,48 тыс. шт.

Возможен и другой способ расчета среднего абсолютного прироста исходя из кумулятивных данных:

$$\bar{\Delta}_y = \frac{2\left(\sum_{i=1}^{n} y_i - ny_1\right)}{n(n+1)}. \qquad (10.10)$$

Обе формулы применяются в зависимости от цели исследования.

Сводной обобщающей характеристикой интенсивности изменения уровней ряда динамики служит *средний темп роста,* показывающий, во сколько раз в среднем за единицу времени изменился уровень динамического ряда.

Необходимость исчисления среднего темпа роста возникает вследствие того, что темпы роста из года в год колеблются. Кроме того, средний темп роста следует определить в тех случаях, когда имеются данные об уровне в начале какого-либо периода и в конце его, а промежуточные данные отсутствуют. Такого рода средний темп роста можно исчислить, если положить в основу расчетов рост не в арифметической прогрессии, которая характеризуется постоянной разностью, а геометрической (a, aq, aq^2, ..., aq^n), характеризующейся постоянным отношением, называемым знаменателем прогрессии (q). Следовательно, вопрос состоит в том, чтобы найти этот знаменатель. *Знаменатель геометрической прогрессии* (q) определяется делением последующего уровня прогрессии на его предыдущий. При делении n-го уровня на первый получаем:

$$\frac{aq^{n-1}}{a} = q^{n-1},$$

отсюда следует:

$$q^{n-1} = \sqrt{\frac{aq^{n-1}}{a}} = \sqrt[n-1]{\frac{B_n}{B_1}}, \tag{10.11}$$

где $B_1 = a$ – первый член прогрессии.

Зная q, мы точно можем определить, какую тенденцию развития явления имеет геометрическая последовательность. Формула (10.11) является средней геометрической и применяется в случае, когда определяющий показатель является не суммой значений, а их произведением. Следовательно, если варианты связаны между собой не знаком сложения, а знаком произведения, нужно вычислить среднюю геометрическую. Обычно средний темп роста вычисляется по формуле средней геометрической из цепных коэффициентов роста:

$$\overline{Tp}_y = \sqrt[m]{K_{2/1} \cdot K_{3/2} \dots \cdot K_{n/n-1}} = \sqrt[m]{\Pi K_{pi/i-1}}. \qquad (10.12)$$

Пример. Средний темп роста производства автомобилей за 1995–2001 гг. рассчитаем по данным табл. 10.6.

$$\overline{T}p = \sqrt[6]{1,0387 \cdot 1,137 \cdot 0,8477 \cdot 1,1429 \cdot 1,0141 \cdot 1,032} =$$
$$= \sqrt[6]{1,1975} = 1,023, \text{ или } 102,3\%.$$

Это говорит о том , что за период 1995–2001 гг. в среднем ежегодно темп роста производства автомобилей составлял 102,3%.

Так как всякий темп роста является отношением уровней ряда динамики, то в формуле $Tp_{2/1} = \frac{y_2}{y_1}100;\ Tp_{3/2} = \frac{y_3}{y_2}100...,$ средней геометрической темпы роста заменяются соответствующим отношением уровней. Заменив темпы роста выражающими их отношениями и, приняв во внимание, что эти величины перемножаются, найдем подкоренное выражение:

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

Следовательно, средний темп роста может быть выражен формулой:

$$\overline{Tp}_y = \sqrt[n-1]{\frac{y_n}{y_1}}. \qquad (10.13)$$

Пример. Продолжим расчеты (см. табл. 10.6). Средний темп роста производства автомобилей за 1995–2001 гг. по формуле (10.13) будет равен:

$$\overline{T}p_y = \sqrt[6]{\frac{1000,0}{835,1}} = \sqrt[6]{1,1975} = 1,023, \text{ или } 102,3\%.$$

При расчете средних темпов роста по периодам различной продолжительности (разноотстоящие ряды динамики) пользуются средними геометрическими взвешенными по продолжительности периодов. Формула средней геометрической взвешенной будет иметь вид:

$$\overline{T}_p = \sqrt[\sum t]{(K_{2/1})^{t_1} \cdot (K_{3/2})^{t_2} \cdot \ldots \cdot (K_{n/n-1})^{t_n}}\,, \qquad (10.14)$$

где t – интервал, в течение которого сохраняется данный темп роста;

$\sum t$ – сумма отрезков периода.

Средний темп прироста не может быть определен на основании последовательных темпов прироста или показателей среднего абсолютного прироста. Для его вычисления необходимо вначале найти средний темп роста, а затем уменьшить его на единицу, или 100%.

$$\overline{T}_{\text{пр}} = \overline{T}_p - 100. \qquad (10.15)$$

Пример. По данным табл. 10.6 средний темп прироста составил:

$$\overline{T}_{\text{пр}} = 102,3\% - 100\% = 2,3\%,$$

т.е. за период 1995–2001 гг. в среднем ежегодно темп прироста производства автомобилей достигал 2,3%.

Для проведения глубокого анализа динамики социально-экономических явлений следует параллельно использовать показатели скорости и интенсивности изменения уровней. Анализ, основанный на использовании какого-либо одного из этих показателей, неизбежно будет иметь односторонний характер.

Для комплексного статистического анализа необходимо использовать систему показателей, характеризующих абсолютную скорость и интенсивность изменения уровней ряда (рис. 10.2).

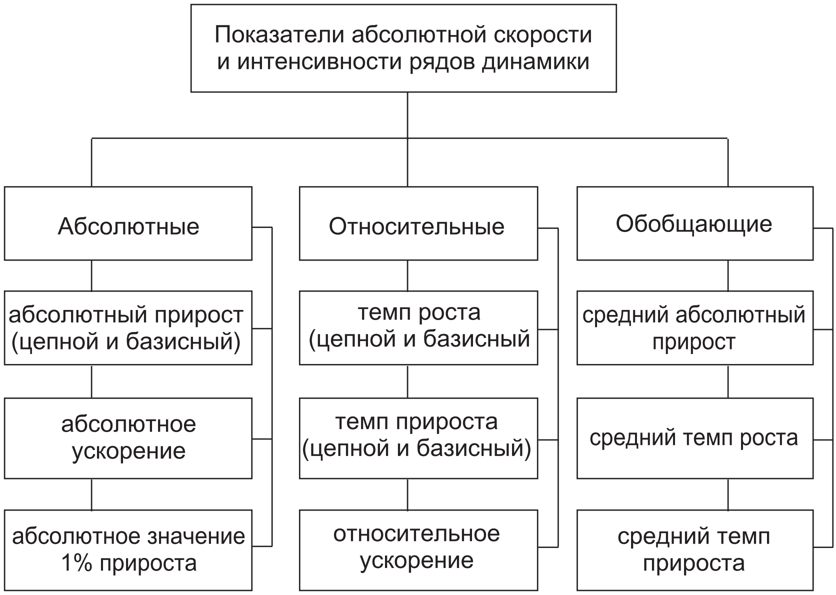

Рис. 10.2. Группировка показателей, характеризующих скорость и интенсивность изменения уровней ряда динамики

10.4 КОМПОНЕНТЫ РЯДА ДИНАМИКИ

Ряд динамики может быть подвержен влиянию факторов эволюционного и осциллятивного характера, а также находиться под влиянием факторов разного воздействия.

Влияния эволюционного характера – это изменения, определяющие некое общее направление развития, как бы многолетнюю эволюцию, которая пробивает себе дорогу через другие систематические и случайные колебания. Такие изменения динамического ряда называются *тенденцией развития*, или *трендом*.

Влияния осциллятивного характера – это *циклические (конъюнктурные)* и *сезонные колебания*. Циклические (или периодические) состоят в том, что значение изучаемого признака в течение какого-то времени возрастает, достигает определенного максимума, затем понижается, достигает определенного минимума, вновь возрастает до прежнего значения и т.д. Иначе циклические колебания можно схематически представить в виде синусоиды $y = \sin t$. Циклические колебания в экономических процессах примерно соответствуют так называемым циклам конъюнктуры. Сезонные колебания – это колебания, периодически повторяющиеся в некоторое определенное время каждого года, дня месяца или часа дня. Эти изменения отчетливо наблюдаются на графиках многих рядов динамики, содержащих данные за период не менее одного года.

Наконец, рассмотрим *нерегулярные колебания*, которые для социально-экономических явлений можно разделить на две группы:

- спорадически наступающие изменения, вызванные, например, войной или экологической катастрофой;
- случайные колебания, являющиеся результатом действия большого количества относительно слабых второстепенных факторов.

Следовательно, первоначальные значения ряда динамики подвергаются самым разнообразным воздействиям. Выделим его четыре основные компоненты:

- основную тенденцию (тренд) (T);
- циклическую, или конъюнктурную (K);
- сезонную (S);
- случайные колебания (E).

Если ряд динамики разбить на различные компоненты, то он представляется в следующем виде:

$$y = f(T, K, S, E).$$

В зависимости от взаимосвязи этих компонент между собой может быть построена аддитивная или мультипликативная модель ряда динамики.

Аддитивная модель ряда динамики $y = T + K + S + E$ характеризуется главным образом тем, что характер циклических и сезонных флюктуаций (колебаний) остается постоянным (рис. 10.3).

Мультипликативная модель ряда динамики $y = T \cdot K \cdot S + E$. В этой модели характер циклических и сезонных флюктуаций остается постоянным только по отношению к тренду (рис. 10.4).

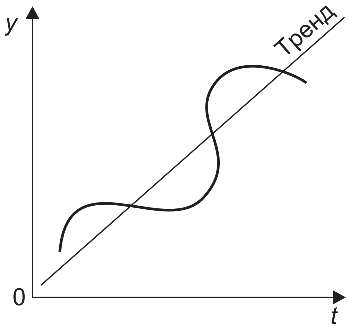

Рис. 10.3. Сочетание различных составляющих ряда динамики при аддитивной связи

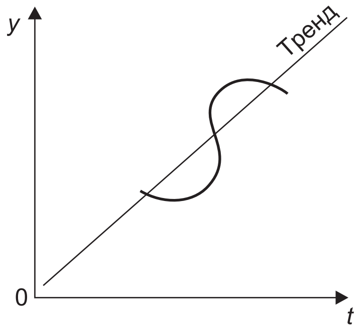

Рис. 10.4. Сочетание различных составляющих ряда динамики при мультипликативной связи

10.5 ВИДЫ ТРЕНДОВОЙ КОМПОНЕНТЫ И ПРОВЕРКА ГИПОТЕЗЫ О СУЩЕСТВОВАНИИ ТЕНДЕНЦИИ

Тренд – это долговременная компонента ряда динамики. Она характеризует основную тенденцию развития явления, при этом остальные компоненты рассматриваются только как мешающие процедуре его определения. При наличии ряда наблюдаемых значений для различных моментов времени следует найти подходящую трендовую кривую, которая сгладила бы остальные колебания.

Виды основной тенденции. В социально-экономических рядах динамики можно наблюдать *тенденцию трех видов:*

- среднего уровня;
- дисперсии;
- автокорреляции.

Тенденция среднего уровня аналитически выражается с помощью математической функции, вокруг которой варьируют фактические уровни исследуемого явления. В таком случае значения тренда в отдельные моменты времени будут являться математическим ожиданием ряда динамики. Часто тенденция среднего уровня называется *детерминированной составляющей исследуемого явления*, и соответствующий ряд динамики выражается следующим уравнением:

$$y_t = f_t + E_i. \tag{10.16}$$

Тенденция дисперсии представляет собой тенденцию изменения отклонений между эмпирическими уровнями и детерминированной компонентой ряда.

Тенденция автокорреляции характеризует изменения связи между отдельными уровнями ряда динамики. Графически это изменение не прослеживается. Однако прежде чем перейти к выделению тренда, следует проверить гипотезу о том, существует ли он вообще. Отсутствие основной тенденции (тренда) означает неизменность среднего уровня ряда во времени.

Методы выявления наличия тенденции. В настоящее время для проверки наличия тренда известно около десятка критериев, различающихся как по мощности, так и по сложности математического аппарата. Рассмотрим два из них: метод, основанный на проверке разности средних двух разных частей одного и того же ряда, и метод Фостера–Стюарта.

Метод проверки существенности разности средних основан на t-критерии Стьюдента. Ряд динамики разбивается на две равные или почти равные части. Проверяется гипотеза о существовании разности средних: $H_0 : \bar{y}_1 = \bar{y}_2$.

Воспользуемся методом проверки, разработанным для малых выборок, так как число членов анализируемого ряда, как правило, довольно незначительно. За основу проверки берется $t\alpha$-критерий Стьюдента. При $t \geq t\alpha$ гипотеза об отсутствии тренда отвергается, при $t < t\alpha$ гипотеза (H_0) принимается. Здесь t – расчетное значение, найденное для анализируемых данных, $t\alpha$ – табличное значение этого критерия при уровне вероятности ошибки, равном α. В случае равенства или при несущественном различии дисперсий двух исследуемых совокупностей ($\sigma^2{}_1 = \sigma^2{}_2$) исчисляется отношение средних с помощью выражения:

$$t = \frac{\bar{y}_1 - \bar{y}_2}{\text{у} \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}, \tag{10.17}$$

где $\bar{y}_1$ и $\bar{y}_2$ – средние для первой и второй половины ряда динамики;
n_1 и n_2 – число наблюдений в этих частях ряда;
σ – среднее квадратическое отклонение разности средних.

Значение $t\alpha$ берется с числом степеней свободы, равным $n_1 + n_2 - 2$. Необходимое значение σ можно определить на основе средней взвешенной величины дисперсий отдельных совокупностей:

$$\text{у} = \sqrt{\frac{(n_1 - 1)^2 \cdot \sigma_1^2 + (n_2 - 1)^2 \cdot \sigma_2^2}{n_1 + n_2 - 2}}. \qquad (10.18)$$

При оценивании дисперсий для первой и второй частей ряда динамики σ_1^2 и σ_2^2 возьмем число степеней свободы, равное $n_1 - 1$ и $n_2 - 1$ соответственно:

$$\text{у}_i^2 = \frac{\sum (y_i - \bar{y}_i)^2}{n - 1}, \ i = 1, 2, ..., n.$$

Проверка гипотезы о равенстве дисперсий реализуется с помощью F-критерия, который основан на сравнении расчетного отношения с табличным.

$$F = \frac{\text{у}_1^2}{\text{у}_2^2}, \qquad (10.19)$$

где $\sigma_1^2 > \sigma_2^2$.

Если расчетное значение F меньше, чем табличное, при заданном уровне вероятности, то можно принять гипотезу о равенстве дисперсий. Если же F больше, чем табличное значение, то гипотеза о равенстве дисперсий отклоняется и формула (10.17) для испытания разности средних не может быть применена.

Следует заметить, что данный метод дает вполне приемлемые результаты лишь в случае рядов с монотонной тенденцией. Когда же ряд динамики меняет общее направление развития, то точка поворота тенденции оказывается близкой к середине ряда, поэтому средние двух отрезков будут близки, а проверка может не показать наличие тенденции.

Метод Фостера–Стюарта кроме определения наличия тенденции явления позволяет выявить основную тенденцию дисперсии уровней ряда динамики, что важно знать при анализе и прогнозировании экономических явлений.

Расчет состоит из следующих этапов.

1. Сравнивается каждый уровень ряда со всеми предыдущими, при этом

если $y_i > y_{i-1}; y_{i-2}, ..., y_1$, то $U_i = 1$; $l_i = 0$;

при $y_i < y_{i-1}; y_{i-2}, ..., y_1$, то $U_i = 0$; $l_i = 1$.

2. Вычисляются значения величин S и d:

$$S = \sum S_i;\ \ d = \sum d_i, \tag{10.20}$$

где $S_i = U_i + l_i$;

$$d_i = U_i - l_i.$$

Анализируя формулу (10.20), нетрудно заметить, что величина S может принимать значения $0 \le S \le n - 1$, причем $S = 0$, когда все уровни ряда равны между собой, и $S = n - 1$, когда ряд динамики монотонно убывает или возрастает. Показатель S характеризует тенденцию изменения дисперсии ряда динамики.

Показатель d имеет нижний предел, равный $-(n - 1)$, и верхний составляет $(n - 1)$. В первом случае ряд является монотонно убывающим, во втором – монотонно возрастающим. Кроме того, показатель d может быть равен нулю:

- если все уровни ряда равны между собой, тогда $\Sigma U_i = \Sigma l_i$. (Данное условие выполняется для ряда, который в первой половине является монотонно убывающим, а во второй – монотонно возрастающим.);
- если уровни подъема и спада чередуются, причем каждое следующее значение уровня подъема (спада) больше (меньше) всех последующих.

Перечисленные случаи, при которых показатель $d = 0$, представляют лишь теоретический интерес, и вероятность их использования при проведении практических расчетов крайне незначительна. Показатель d характеризует изменение тенденций в среднем.

Оба показателя, S и d, асимптотически нормальны и имеют независимые распределения.

3. Проверяется с использованием t-критерия Стьюдента гипотеза о том, можно ли считать случайными разности $S - \mu$ и $d - 0$:

$$t_s = \frac{S - м}{у_1};\ t_d = \frac{d - 0}{у_2}, \qquad (10.21)$$

где μ – среднее значение величины S, определенное для ряда, в котором уровни расположены случайным образом;

σ_1 – стандартная ошибка величины S;

σ_2 – стандартная ошибка величины d.

Значения величин μ, σ_1 и σ_2 табулированы и приведены в приложении 13.

4. Сравниваются расчетные значения t_s и t_d с табличным при заданном уровне значимости. Если $t_s < t_{табл}$ и $t_d < t_{табл}$, то гипотеза об отсутствии тренда в средней и дисперсии подтверждается.

Пример. Рассмотрим определение наличия тенденции в ряду динамики производства реализованной продукции на производственном объединении за 1982 – 2001 гг. (табл. 10.7).

Таблица 10.7

Реализованная продукция производственного объединения
(определение U_t и l_i)

Год	Млн руб. Y_i	U_t	l_i	Год	Млн руб. Y_i	U_t	l_i
1982	63,5	0	0	1992	63,0	0	0
1983	62,1	0	1	1993	59,9	0	1
1984	61,6	0	1	1994	62,0	0	0
1985	61,3	0	1	1995	63,4	0	0
1986	61,5	0	0	1996	64,5	0	0
1987	61,3	0	0	1997	58,0	0	1
1988	62,4	0	0	1998	54,5	0	1
1989	65,5	1	0	1999	56,0	0	0
1990	64,8	0	0	2000	55,2	0	0
1991	64,3	0	0	2001	56,1	0	0

Из формулы (10.20) определяем $S = 7$, $d = 5$. По данным приложения 13 при $n = 20$, $\mu = 5{,}195$, $\sigma_1 = 1{,}677$, $\sigma_2 = 2{,}279$. Подставляя полученные значения в формулу (10.21), рассчитаем:

$$t_s = \frac{7-5{,}196}{1{,}677} = 1{,}076; \quad t_d = \frac{5-0}{2{,}279} = 2{,}194.$$

Ближайшее табличное значение $t_{\text{табл}}$ для двустороннего критерия при уровне значимости 0,10 равно $t_{\text{табл}} = 1{,}725$, т.е. $t_{\text{табл}} > t_s$, $t_{\text{табл}} < t_d$. Следовательно, гипотеза об отсутствии тенденции в дисперсии показателя реализованной продукции подтвердилась, а в средней – отвергнута.

10.6 МЕТОДЫ АНАЛИЗА ОСНОВНОЙ ТЕНДЕНЦИИ (ТРЕНДА) В РЯДАХ ДИНАМИКИ

После того как установлено наличие тенденции в ряду динамики, производится ее описание с помощью методов сглаживания. Методы сглаживания разделяются на две основные группы:

- сглаживание или механическое выравнивание отдельных членов ряда динамики с использованием фактических значений соседних уровней;
- выравнивание с применением кривой, проведенной между конкретными уровнями таким образом, чтобы она отображала тенденцию, присущую ряду, и одновременно освободила его от незначительных колебаний.

Рассмотрим каждый из них.

Метод усреднения по левой и правой половине. Разделяют ряд динамики на две части, находят для каждой из них среднее арифметическое значение и проводят через полученные точки линию тренда на графике.

Метод укрупнения интервалов. Если рассматривать уровни экономических показателей за короткие промежутки времени, то в силу влияния различных факторов, действующих в разных направлениях, в рядах динамики наблюдается снижение и повышение этих уровней.

Это мешает видеть основную тенденцию развития изучаемого явления. Поэтому для наглядного представления тренда применяется метод укрупнения интервалов, основанный на укрупнении периодов времени, к которым относятся уровни ряда. Например, ряд ежесуточного выпуска продукции заменяется рядом месячного выпуска продукции и т.д.

Метод простой скользящей средней. Сглаживание ряда динамики с помощью скользящей средней заключается в том, что вычисляется средний уровень из определенного числа первых по порядку уровней ряда, затем – средний уровень из такого же числа уровней начиная со второго, далее – начиная с третьего и т.д. Таким образом, при расчетах среднего уровня как бы «скользят» по ряду динамики от его начала к концу, каждый раз отбрасывая один уровень вначале и добавляя один следующий. Отсюда название – *скользящая средняя.*

Каждое звено скользящей средней – это средний уровень за соответствующий период, который относится к *середине выбранного периода*.

Для каждого конкретного ряда динамики (y_1, y_2, ..., y_n) алгоритм расчета скользящей средней следующий.

1. Определить интервал сглаживания, т.е. число входящих в него уровней m ($m < n$), используя правило: если необходимо сгладить мелкие, беспорядочные колебания, то интервал сглаживания берут по возможности большим, и, наоборот, интервал сглаживания уменьшают, когда нужно сохранить более мелкие волны и освободиться от периодически повторяющихся колебаний, возникающих, например, из-за автокорреляции уровней.

2. Вычислить среднее значение уровней, образующих интервал сглаживания, которое одновременно является сглаживающим значением уровня, находящегося в центре интервала сглаживания, при условии, что m – нечетное число, по одной из формул:

$$y_t = \frac{\sum\limits_{i=t-p}^{t+p} y}{m} \quad \text{или} \quad \bar{y}_t = \bar{y}_{t-1} + \frac{y_{t+p} - y_{t-p-1}}{m}, \tag{10.22}$$

где y_i – фактическое значение i-го уровня;
m – число уровней, входящих в интервал сглаживания ($m = 2p + 1$);
y_t – текущий уровень ряда динамики;
i – порядковый номер уровня в интервале сглаживания;
p – при нечетном m равно: $p = (m – 1) / 2$.

Определение скользящей средней по четному числу членов ряда динамики несколько сложнее, так как средняя может быть отнесена только к середине между двумя датами, находящимися в середине интервала сглаживания.

Если число членов скользящей средней обозначить через $2m$, то серединным будет уровень, относящийся к $m + 1/2$ члену ряда, т.е. имеет место сдвиг периода, к которому относится уровень. Например, средняя, найденная для четырех членов, относится к середине между вторым и третьим периодами, следующая средняя – к середине между третьим и четвертым и т.д. Чтобы ликвидировать такой сдвиг, применяют так называемый *способ центрирования*. Центрирование заключается в нахождении средней из двух смежных скользящих средних для отнесения полученного уровня к определенной дате. При центрировании необходимо находить скользящие суммы, скользящие средние нецентрированные по этим суммам и средние из двух смежных нецентрированных скользящих средних.

3. Сдвинуть интервал сглаживания на одну точку вправо, потом вычислить по формуле (10.22) сглаженное значение для $t + 1$ члена, снова произвести сдвиг и т.д. В результате последовательного применения приведенной итеративной процедуры получится $n - (m - 1)$ новых сглаженных уровней.

Первые и последние p членов ряда с помощью данного алгоритма сгладить нельзя, так как их значения теряются.

Пример. Покажем расчет 5-летней и 4-летней скользящих средних на данных табл. 10.8. Как видим, скользящая средняя дает более или менее плавное изменение уровней (рис. 10.5).

Метод простой скользящей средней вполне приемлем, если графическое изображение ряда динамики напоминает прямую линию. В этом случае не искажается динамика исследуемого явления. Однако когда тренд выравниваемого ряда имеет изгибы и к тому же желательно сохранить мелкие волны, использовать для сглаживания ряда метод простой скользящей средней нецелесообразно, так как простая скользящая средняя может привести к значительным искажениям исследуемого процесса. В таких случаях более надежным является использование взвешенной скользящей средней.

Метод взвешенной скользящей средней. Взвешенная скользящая средняя отличается от простой скользящей средней тем, что уровни, входящие в интервал усреднения, суммируются с различными весами. Это связано с тем, что аппроксимация сглаживаемого ряда

Таблица 10.8

Сглаживание урожайности зерновых культур в хозяйстве за 1986–2001 гг. методом скользящей средней

Год	Урожайность, ц/га	Пятилетние скользящие суммы	Пятилетние скользящие средние	Четырехлетние скользящие суммы	Четырехлетние скользящие средние (нецентрированные)	Четырехлетние скользящие средние (центрированные)	Пятилетние взвешенные скользящие средние
А	1	2	3	4	5	6	7
1986	9,5	–	–	–	–	–	–
1987	13,7	–	–	–	12,3	–	–
1988	12,1	–	12,5	–	13,2	12,8	12,8
1989	14,0	–	13,7	49,3	13,7	13,5	13,4
1990	13,2	63,5	14,1	53,0	14,6	14,1	14,1
1991	15,6	68,6	14,4	54,9	14,6	14,6	14,7
1992	15,4	70,3	15,2	58,2	15,7	15,1	15,1
1993	14,0	72,2	15,6	58,2	15,6	15,6	15,4
1994	17,6	75,8	14,7	62,6	14,5	15,0	15,6
1995	15,4	78,0	15,1	62,4	15,3	14,9	14,9
1996	10,9	73,5	15,3	57,9	14,7	15,0	14,3
1997	17,5	75,4	15,5	61,4	15,5	15,1	15,2
1998	15,0	76,4	15,2	58,8	16,3	15,8	16,2
1999	18,5	77,3	16,0	61,9	15,65	15,97	16,3
2000	14,2	76,1	–	65,2	–	–	–
2001	14,9	80,1	–	62,6	–	–	–

Рис. 10.5. Динамика урожайности зерновых культур в хозяйстве за 1986–2001 гг.

динамики в пределах интервала сглаживания осуществляется с использованием уровней, рассчитанных по полиному $\overline{y}_i = a_0 + a_i \cdot i + a_2 \cdot i^2 + ...$ (здесь i – порядковый номер уровня в интервале сглаживания). Полином 1-го порядка $\overline{y}_i = a_0 + a_i \cdot i$ есть уравнение прямой, следовательно, метод простой скользящей средней является частным случаем метода взвешенной скользящей средней. Коэффициенты полиномов находятся по способу наименьших квадратов (глава 9).

На первом этапе сглаживания по методу взвешенной средней определяются интервал сглаживания и порядок аппроксимирующего полинома – параболы. Считается, что при использовании полиномов высоких степеней и при меньших размерах интервалов сглаживание ряда динамики будет более «гибким».

Центральная ордината параболы принимается за сглаженное значение соответствующего фактическим данным уровня. Поскольку отсчет времени в пределах интервала сглаживания производится от его середины, т.е. ($t = i$) i =..., –2, –1, 0, 1, 2, ..., то сглаженное значение уровня равно параметру a_0 подобранной параболы и является соответствующей скользящей средней. Поэтому для сглаживания нет необходимости прибегать к процедуре подбора системы парабол, так как величину a_0 можно получить как взвешенную среднюю из m уровней.

Пример. Если в интервал сглаживания входят пять последовательных уровней ряда со сдвигом во времени на один шаг, а выравнивание проводится по полиному 2-го порядка, то коэффициенты полинома находятся из условия

$$\sum_{t=1}^{y}\left(y-a_0-a_1\cdot i-a_2\cdot i^2\right)^2\to\min.$$

Учитывая, что для нечетных $k \Sigma i^k = 0$, приходим к системе:

$$\begin{cases}\sum y-5a_0-a_2\sum i^2=0;\\ \sum yi-a_1\sum i^2=0;\\ \sum yi^2-a_0\sum i^2-a_2\sum i^4=0.\end{cases}$$

Для определения a_0 необходимо найти значения $\sum\limits_{-p}^{p} i^2$ и $\sum\limits_{-p}^{p} i^4$.

Так как интервал сглаживания равен $m=5$, то $\sum\limits_{-p}^{p} i^2=10$ и $\sum\limits_{-p}^{p} i^4=34$.

Система нормальных уравнений для определения a_0 и a_2 в этом случае записывается так:

$$\begin{cases}5a_0+10a_2=\sum\limits_{-p}^{p} yi;\\ 10a_0+34a_2=\sum\limits_{-p}^{p} y\cdot i^2.\end{cases}$$

Решение этой системы относительно a_0 может быть представлено следующим образом:

$$a_0 = \frac{34\sum y - 10\sum y^2}{5 \cdot 34 - 10^2} = \frac{17\sum y - 5\sum yi^2}{35} = \frac{1}{35}\left(-3y_{t-2} + 12y_{t-1} + 17y_t + 12y_{t+1} - 3y_{t+2}\right). \quad (10.23)$$

Аналогичным путем получим выражения и для других интервалов сглаживания по параболе второго и третьего порядка. Так, например, для

$$\begin{cases} m = 7 \;\; \bar{y}_t = \frac{1}{21}\left(-2y_{t-3} + 3y_{t-2} + 6y_{t-1} + 7y_1 + 6y_{t+1} + 3y_{t+2} - 2y_{t+3}\right); \\ m = 9 \;\; \bar{y}_t = \frac{1}{231}(-21y_{t-4} + 14y_{t-3} + 39y_{t-2} + 54y_{t-1} + 59y_1 + 54y_{t+1} + \\ \qquad + 39y_{t+2} + 14y_{t+3} - 21y_{t+4}). \end{cases}$$

Согласно приведенным формулам веса симметричны относительно центрального уровня (y_t) и их сумма с учетом общего множителя, вынесенного за скобки, равна 1.

По данным рассмотренного выше примера с урожайностью зерновых получим следующие значения взвешенных скользящих средних для $m = 5$ (табл. 10.8 графа 7). Пятичленная скользящая средняя показывает, что на протяжении периода с 1986 по 2001 г. наблюдался рост урожайности зерновых культур.

Выбор уравнения тренда, отображающего развитие социально-экономических явлений во времени. Для отображения основной тенденции развития явлений во времени применяются полиномы разной степени, экспоненты, логистические кривые и другие функции.

Полиномы имеют следующий вид:

- 1-й степени – $\bar{y}_t = a_0 + a_1 t;$
- 2-й степени – $\bar{y}_t = a_0 + a_1 t + a_2 t^2;$
- 3-й степени – $\bar{y}_t = a_0 + a_1 t + a_2 t^2 + a_3 t^3;$
- *n*-й степени – $\bar{y}_t = a_0 + a_1 t + a_2 t^2 + a_3 t^3 + ... + a_n t^n,$

где $a_0, a_1, a_2, ..., a_n$ – параметры полиномов;
t – условное обозначение времени.

В статистической практике параметры полиномов невысокой степени иногда имеют конкретную интерпретацию характеристик динамического ряда. Так, параметр a_0 трактуется как характеристика средних условий ряда динамики, параметры a_1, a_2, a_3 – изменение ускорения.

В статистике выработано правило выбора степени полинома модели развития, основанное на определении величин конечных разностей уровней динамических рядов. Согласно этому правилу *полином 1-й степени* (прямая) применяется как модель такого ряда динамики, у которого первые разности (абсолютные приросты) постоянны; *полиномы 2-й степени* – для отражения ряда динамики с постоянными вторыми разностями (ускорениями); *полиномы 3-й степени* – с постоянными третьими разностями и т.д.

Для полиноминальных моделей характерно отсутствие прямой связи между абсолютными приростами и приростами уровней рядов динамики.

Предполагаемой функцией, отражающей процесс роста явления, может быть и экспонента $\overline{y}_t = a_0 a_1{}^t$ или $\overline{y}_t = a_0 \cdot (a_1)^{b_1 t + b_2 t}$. Экспоненты характеризуют прирост, зависящий от величины основания функции.

Отдельные уравнения выражают различные типы динамики.

Монотонное возрастание или убывание процесса характеризуют функции:

- линейная;
- параболическая;
- степенная;
- экспоненциальная простая (показательная) и производная от нее логарифмическая линейная;
- сложная экспоненциальная и производная от нее логарифмическая парабола;
- гиперболическая (главным образом убывающих процессов);
- комбинация их видов.

Для моделирования динамических рядов, проявляющих быстрое развитие в начале ряда и затухающее его развитие к концу, т.е. тех, которые характеризуются стремлением к некоторой предельной величине, применяются *логистические функции*.

Логистическую функцию часто записывают в следующем виде:

$$\overline{y}_t = \frac{K}{1 + C^{-a_0 t}} \quad \text{или} \quad \overline{y}_t = \frac{K}{1 + 10^{a_0 + a_1 t}},$$

где C – основание натурального логарифма.

Логистическая кривая симметрична относительно точки перегиба и при $t = -\infty$ стремится к нулю, а при $t = +\infty$ стремится к некоторой постоянной величине, к которой кривая асимптотически приближается. Если найти вторую производную от $\bar{y}_t$ по t и приравнять ее к нулю, то для логистической кривой, выражаемой через местоположение точки перегиба кривой, $t = \lg a_1 : a_0$; $\bar{y}_t = n : 2$.

Тип процессов, характеризующийся наличием экстремальных значений, описывается *кривой Гомперца*, имеющей следующее уравнение:

$$\bar{y}_t = K \cdot (a_0)^{a_1 t}.$$

Возможны четыре варианта этой кривой. Для экономистов наибольшее значение имеет кривая, у которой $\lg a_0 < 0$ и $a_1 < 1$. Развитие уровня такой кривой имеет следующие этапы. Если коэффициент a_1 меньше 1 при отрицательном значении $\lg a_0$, то на первом этапе прирост кривой незначителен. Он медленно увеличивается по мере роста t, но на следующем этапе прирост увеличивается быстрее, а затем, после точки перегиба, начинает уменьшаться, и на подходе к линии асимптоты прирост кривой опять незначителен.

Прологарифмировав функцию Гомперца $\lg \bar{y}_t = \lg k + \lg a_0 \cdot a_1^t$, получим модифицированную экспоненту. Вводя в модифицированную экспоненту величину, обратную $\bar{y}_t$, получим логистическую кривую. Следовательно, логистическая кривая имеет сходство с кривой Гомперца. Различие между ними состоит в том, что изменение во времени первых разностей кривой Гомперца асимметрично, а у логистической кривой их изменение во времени имеет симметричный вид, напоминающий нормальное распределение.

Для выбора уравнения можно воспользоваться *формулой стандартной ошибки*

$$S = \sqrt{\frac{\sum (y_t - \bar{y}_t)^2}{n - p - 1}}, \tag{10.24}$$

где p – число параметров уравнения.

Можно также применить *критерий наименьшей суммы квадратов отклонения* эмпирических уровней от теоретических

$$S = \sum_{t=1}^{n} \left(y_t - \bar{y}_t\right)^2 \to \min.$$

Из множества возможных уравнений тренда можно выбрать то уравнение, которому соответствует минимальное значение, т.е. критерий наименьших квадратов отклонений, либо использовать формулу средней ошибки аппроксимации:

$$\varepsilon_t = \frac{1}{n}\sum_{t=1}^{n}\frac{\left|y_t - \bar{y}_t\right|}{y_t}\cdot 100. \qquad (10.25)$$

Все эти характеристики имеют один и тот же смысл: показывают, как близко аналитическая функция выравнивания огибает все значения исходного ряда. Поэтому, проводя сравнительную оценку моделей тренда, можно использовать лишь одну из перечисленных характеристик. Результаты такой оценки, полученные на основе прочих характеристик, как правило, совпадают. Наиболее часто в качестве меры точности аппроксимации выбирают остаточную дисперсию или остаточное среднее квадратическое отклонение.

Расчет параметров полиномов различными методами. После того как выяснен характер кривой развития, необходимо определить ее параметры. Элементарный метод определения параметров уравнения тренда, описанного полиномом или экспонентой, состоит в решении системы уравнений *по известным уровням ряда динамики*.

Пример. Дан ряд динамики, представленный табл. 10.9. Приняв условные обозначения времени через *t* и взяв две точки – конечный и начальный уровни, можно построить *уравнение прямой по этим двум точкам*.

Таблица 10.9

Динамика производства готовой продукции на фирме

	1996	1997	1998	1999	2000	2001
Готовая продукция фирмы, тыс. руб.	18	21	26	22	25	28
t	1	2	3	4	5	6

Уравнение прямой имеет вид: $\bar{y}_t = a_0 + a_1 t$. Для 1996 г. его уровень составил: $a_0 + a_1 \cdot 1 = 18$; для 2001 г.: $a_0 + 6 \cdot a_1 = 28$.

Решая эти уравнения как систему уравнений, получим: $10 = 5a_1$; $a_1 = 2$; $a_0 = 28 - 6a_1 = 28 - 12 = 16$. Следовательно, приближенная модель динамики готовой продукции выражается уравнением $\overline{y}_t = 16 + 2t$. Здесь параметр a_1 соответствует абсолютному приросту.

Можно предположить и развитие по параболе второго порядка: $\overline{y}_t = a_0 + a_i t + a_2 t^2$, но тогда следует взять три точки, например 1996, 1999, 2001 гг., т.е. уровни при $t = 1$, $t = 4$, $t = 6$.

Составим систему трех уравнений с тремя неизвестными:

$$\begin{cases} a_0 + 1a_1 + 1a_2 = 18; \\ a_0 + 4a_1 + 16a_2 = 22; \\ a_0 + 6a_1 + 36a_2 = 28. \end{cases}$$

Решая эту систему, получим: $a_0 = 18$, $a_1 = 0{,}3$ и $a_2 = 0{,}3$, а само уравнение применительно к нашему примеру выразится $\overline{y}_t = 18 + 0{,}3t + 0{,}3t^2$, что в приближенной форме определит модель динамики данного явления.

Отрицательным моментом в таком моделировании тренда служат разные числовые выражения параметров в различных точках их определения.

Другим способом определения параметров уравнения является *метод средних значений (линейных отклонений)*, заключающийся в следующем: ряд расчленяется на две примерно равные части и вводится требование, чтобы сумма выравненных значений в каждой части совпала с суммой фактических значений, т.е. чтобы сумма отклонений фактических данных от выравненных равнялась нулю.

В случае выравнивания по прямой линии

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

где a_0 и a_1 – параметры,

получим:

$$\Sigma(y - a_0 - \mathrm{a}_1 t) = 0, \tag{10.26}$$

откуда

$$na_0 + a_1 \Sigma t = \Sigma y.$$

Если применить это требование к каждой из двух частей ряда, то, вычислив для каждой части динамического ряда Σt и Σy, получим два уравнения с двумя неизвестными. В результате решения этой системы уравнений находим параметры a_0 и a_1, т.е. начальный уровень и скорость ряда. При этом значение $t = 1, 2, 3, 4,..., n$.

Разобьем приведенный в табл. 10.8 ряд динамики урожайности зерновых культур на два периода:

1-й – 1986–1993 гг.;

2-й – 1994–2001 гг.,

тогда: $\Sigma_1 y = 107{,}5$; $\Sigma_2 y = 124{,}0$; $\Sigma_1 t = 45$; $\Sigma_2 t = 91$.

Для определения параметров a_0 и a_1 решим систему:

$$\begin{cases} 8a_0 + 45a_1 = 107{,}5; \\ 8a_0 + 91a_1 = 124{,}0. \end{cases}$$

Вычтем из второго уравнения первое. В результате получим:

$$a_1 = 0{,}359; \ a_0 = 11{,}42.$$

Искомое уравнение будет иметь следующий вид:

$$\bar{y}_t = 11{,}42 + 0{,}359t.$$

Метод средних значений прост и требует минимального количества вычислений. Его недостаток заключается в том, что при произвольном расчленении ряда на две части мы будем получать разные результаты. Метод средних значений, как и выравнивание ряда динамики с помощью среднего прироста и темпа роста, может применяться для ориентировочных расчетов.

Выравнивание ряда динамики с помощью метода конечных разностей. Этот метод заключается в следующем.

Пусть ряд динамики y_t описывается полиномом p-й степени. Для полинома p-й степени вычислим первые разности:

$$\Delta_t^{(1)} = y_{t+1} - y_t;$$

вторые разности:

$$\Delta_t^{(2)} = \Delta^{(1)}_{t+1} - \Delta_t^{(1)}$$

и т.д.

Общая формула *p*-й разности:

$$\Delta_t^{(p)} = y_p + py_{p-1} + \frac{p(p-1)}{2!} \cdot y_{p-2} - \\ - \frac{p(p-1)(p-2)}{3!} \cdot y_{p-3} + \ldots + (-1)^p y_0. \quad (10.27)$$

Любой член y_i ($i = 0, 1, 2, 3, ..., n$) ряда динамики можно выразить через начальный уровень ряда y_0 и конечные разности:

$$y_i = y_0 + \Delta_0^{(1)};\ y_2 = y_0 + \Delta_0^{(1)} + \Delta_1^{(1)},$$

но $\Delta_1^{(1)} = \Delta_0^{(1)} + \Delta_0^{(2)}$, поэтому $y_2 = y_0 + 2\Delta_0^{(1)} + \Delta_0^{(2)}$

и т.д.

Отсюда получаем:

$$\overline{y}_i = y_0 + t\Delta_0^{(1)} + \frac{t(t-1)}{2!}\Delta_0^{(2)} + \frac{t(t-1)(t-2)}{3!}\Delta_0^{(3)} + \ldots + \Delta_0^{(t)}. \quad (10.28)$$

Если первые разности не равны, но варьируют с незначительными отклонениями друг от друга, а средняя арифметическая вторых разностей настолько мала, что ею можно пренебречь, то первые разности можно считать практически равными.

Окончательная формула для расчета уровней ряда динамики при равных или почти равных первых разностях будет:

$$\overline{y}_t = \overline{y} + \overline{\Delta}^{(1)} \cdot t^1. \quad (10.29)$$

Если, анализируя вторые разности, мы придем к выводу, что они практически равны, то, вычислив коэффициенты параболы 2-го порядка, получим тренд ряда динамики:

$$\bar{y}_t = \bar{y} - \frac{n^2 - 1}{24}\bar{\Delta}^{(2)} + \bar{\Delta}^{(1)} \cdot t^1 + \frac{\bar{\Delta}^{(2)}}{2} \cdot t^2, \qquad (10.30)$$

где $\bar{y}_t$ – выравненное значение ряда динамики;
$\bar{y}$ – средний уровень ряда динамики;
$\bar{\Delta}^{(1)}$ – средняя арифметическая первых разностей;
$\bar{\Delta}^{(2)}$ – средняя арифметическая вторых разностей;
n – число уровней;
t – независимая переменная (время).

Пример. Рассмотрим сглаживание методом конечных разностей на следующих данных (табл. 10.10).

Таблица 10.10

Удельный вес воздушных судов, прибывших без опоздания по сравнению с расписанием за 1991–2001 гг.
(сглаживание методом конечных разностей)

Год	Удельный вес прибытий, % y_i	Условное обозначение времени t	Разности		Выравненные значения $\bar{y}_t$
			$\Delta^{(1)}$	$\Delta^{(2)}$	
1991	91,6	–5	–	–	91,6
1992	91,5	–4	–0,1	–	91,4
1993	91,3	–3	–0,2	–0,1	91,3
1994	91,1	–2	–0,2	0	91,1
1995	91,0	–1	–0,1	0,1	91,0
1996	90,8	0	–0,2	–0,1	90,8
1997	90,6	1	–0,2	0	90,6
1998	90,4	2	–0,2	0	90,4
1999	90,2	3	–0,2	0	90,2
2000	90,0	4	–0,2	0	90,1
2001	89,9	5	–0,1	0,1	89,9
Итого	998,4	0	–1,7	0,0	998,4
Среднее значение $\bar{y}$	90,76	–	–0,17	–	–

Способом конечных разностей получили уравнение, выражающее тенденцию динамики данного в ряде процесса:

$$y_t = 90{,}76 - 0{,}17t.$$

Метод наименьших квадратов при расчете параметров полиномов. Этот метод при моделировании рядов динамики можно рассматривать как некоторый прием получения оценки детерминированной компоненты $f(t)$, которая характеризует тренд изучаемого явления.

В экономике часто применяется функция вида:

$$y_t = a_0 + \sum_{i=1}^{p} a_i \cdot t^i. \tag{10.31}$$

Отсюда a_0, a_1, a_2, ..., a_p находятся методом наименьших квадратов, сущность которого нам известна из регрессионного анализа.

Согласно этому методу для нахождения параметров полинома p-й степени необходимо решить систему так называемых нормальных уравнений:

$$\begin{cases} na_0 + a_1\sum t + a_2\sum t^2 + ... + a_p\sum t^p = \sum y; \\ a_0\sum t + a_1\sum t^2 + a_2\sum t^3 + ... + a_p\sum t^{p-1} = \sum yt; \\ \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \\ a_0\sum t^p + a_1\sum t^{p+1} + a_2\sum t^{p+2} + ... + a_p\sum t^{2p} = \sum yt^p, \end{cases} \tag{10.32}$$

где n – число членов в ряду динамики; $t = 1, 2, ..., n$.

Система (10.32), состоящая из p уравнений, содержит в качестве известных величины Σy, Σyt, ..., Σyt^p, т.е. суммы наблюдаемых значений уровней динамического ряда, умноженные на показатели времени в степени 0, 1,2, ..., p, и p-неизвестных величин a_i. Решение этой системы относительно a_0, a_1, ..., a_p и дает искомые значения параметров.

Системы для расчета параметров полиномов невысоких степеней намного проще. Обозначим последовательные параметры полиномов как a_0, a_1, a_2. Тогда системы нормальных уравнений для оценивания параметров прямой ($\overline{y}_i = a_0 + a_i t$) примут вид:

$$\begin{cases} na_0 + a_1 \sum t = \sum y; \\ a_0 \sum t + a_1 \sum t^2 = \sum yt; \end{cases} \tag{10.33}$$

для параболы 2-го порядка ($\overline{y}_i = a_0 + a_1 t + a_2 t^2$):

$$\begin{cases} na_0 + a_1 \sum t + a_2 \sum t^2 = \sum y; \\ a_0 \sum t + a_1 \sum t^2 + a_2 \sum t^3 = \sum yt; \\ a_0 \sum t^2 + a_1 \sum t^3 + a_2 \sum t^4 = \sum yt^2. \end{cases} \tag{10.34}$$

Решение системы (10.33) относительно искомых параметров a_0 и a_1:

$$\begin{cases} a_0 = \dfrac{\sum y \sum t^2 - \sum t \sum yt}{n \sum t^2 - (\sum t)^2}; \\ a_0 = \dfrac{n \sum yt - \sum y \sum t}{n \sum t^2 - (\sum t)^2}. \end{cases}$$

Аналогичным путем можно было бы подойти и к системе (10.34). Однако такой путь расчета параметров достаточно трудоемок, если он не выполняется с помощью пакета прикладных программ. Поэтому перейдем к упрощенным приемам расчета параметров, применение которых дает ощутимую экономию труда без какой-либо потери точности результатов.

Упрощенный расчет параметров уравнений заключается в переносе начала координат в середину ряда динамики. В этом случае упрощаются сами нормальные уравнения, кроме того, уменьшаются абсолютные значения величин, участвующих в расчете. В самом деле, если до переноса начала координат t было равно 1, 2, 3, ..., n, то после переноса t = ... –4, –3, –2, –1, 0, 1, 2, 3, 4, ..., если число членов ряда нечетное (табл. 10.10). Когда же число ряда четное (табл. 10.17), то t = ..., –5, –3, –1, 1, 3, 5, ... Следовательно, Σt и все Σt^p, у которых p – нечетное число, равны 0. Таким образом, все члены уравнений, содержащие Σt с такими степенями, могут быть исключены. Системы нормальных уравнений теперь упрощаются для прямой:

$$\begin{cases} a_0 n = \sum y; \\ a_1 \sum t^2 = \sum ty; \end{cases} \quad (10.35)$$

для параболы второго порядка:

$$\begin{cases} a_0 n + a_2 \sum t^2 = \sum y; \\ a_1 \sum t^2 = \sum ty; \\ a_0 \sum t^2 + a_2 \sum t^4 = \sum t^2 y. \end{cases} \quad (10.36)$$

Решая системы (10.35), (10.36) относительно неизвестных параметров, получим величины параметров соответствующих полиномов.

Параметр a_1 выражает начальную скорость роста, а коэффициент a_2 – постоянную скорость изменения прироста. Если уровень явления растет с ускорением, то величина этого ускорения в среднем за изучаемый период равна $2a_2$ единицам.

При сглаживании ряда динамики по экспоненте ($\bar{y}_t = a_0 e^{a_1 t}$) для определения параметров применяется метод наименьших квадратов к логарифмам исходных данных. Так, для нахождения параметров экспоненты необходимо решить следующую систему уравнений:

$$\begin{cases} \sum \lg y = n \lg a_0 + \lg a_1 \sum t; \\ \sum t \lg y = \sum t \lg a_0 + \lg a_1 \sum t^2. \end{cases} \quad (10.37)$$

Если $\Sigma t = 0$, то параметры уравнения $\lg a_0$ и $\lg a_1$ находим по формулам:

$$\lg a_0 = \frac{\sum \lg y}{n}; \ \lg a_1 = \frac{\sum t \lg y}{\sum t^2}.$$

Экспонента отражает постоянный относительный рост, равный e^{a_1} единицам.

Пример. Рассмотрим аналитическое выравнивание по прямой ряда динамики урожайности зерновых культур в хозяйстве за 1987–2001 гг.

Исходные данные и расчет параметров уравнения прямой представлены в табл. 10.11.

Для выравнивания данного ряда по прямой используем уравнение $\bar{y}_i = a_0 + a_1 t$. Расчет параметров уравнения проведем по упрощенному методу, т.е. $\Sigma t = 0$. В нашем примере число исходных данных

Таблица 10.11

Динамика урожайности зерновых культур в хозяйстве
(определение параметров уравнения методом наименьших квадратов)

Год	Урожайность, ц/га, y_i	t	t^2	$y_i \cdot t$	$\bar{y}_t$
А	1	2	3	4	5
1987	13,7	–7	49	–95,9	13,6
1988	12,1	–6	36	–72,6	13,8
1989	14,0	–5	25	–70,0	13,9
1990	13,2	–4	16	–52,8	14,1
1991	15,6	–3	9	–46,8	14,3
1992	15,4	–2	4	–30,8	14,5
1993	14,0	–1	1	–14,0	14,6
1994	17,6	0	0	0	14,8
1995	15,4	1	1	15,4	15,0
1996	10,9	2	4	21,8	15,1
1997	17,5	3	9	52,5	15,3
1998	15,0	4	16	60,0	15,5
1999	18,5	5	25	92,5	15,7
2000	14,2	6	36	85,2	15,8
2001	14,9	7	49	104,3	16,0
Итого	222,0	0	280	48,8	222,0

ряда нечетное ($n = 15$) (табл. 10.11). При этом система уравнений примет вид $\begin{cases} na_0 = \sum y \\ a_1 \sum t^2 = \sum ty \end{cases}$. Используя итоги граф 1, 3, 4, определим параметры уравнения прямой $a_0 = \frac{\sum y}{n}$; $a_1 = \frac{\sum ty}{\sum t^2}$:

$$a_0 = \frac{222,0}{15} = 14,8; \; a_1 = \frac{48,8}{280} = 0,17.$$

По рассчитанным параметрам запишем уравнение прямой ряда динамики урожайности зерновых культур:

$$\overline{y}_i = 14,8 + 0,17t.$$

Данное уравнение показывает, что в течение исследуемого периода урожайность в хозяйстве увеличивалась в среднем на 0,17 ц/га в год.

Используя приведенное уравнение, рассчитаем для каждого года теоретические значения:

для 1987 г. $\overline{y}_{t=-7} = 14,8 + 0,17(-7) = 13,6$;

для 1988 г. $\overline{y}_{t=-6} = 14,8 + 0,17(-6) = 13,8$ (см. графу 5 табл. 10.11).

Правильность расчета уровней выравниваемого ряда динамики может быть проверена следующим образом: сумма значений эмпирического ряда должна совпасть с суммой выравненных значений ряда, т.е. $\sum_{i=1}^{n} y_i = \sum_{i=1}^{n} \overline{y}_t$ (см. итоги граф 1 и 5).

Во многих случаях моделирование рядов динамики с помощью полиномов или экспоненциальной функции не дает удовлетворительных результатов, так как в рядах динамики содержатся заметные периодические колебания вокруг общей тенденции или наблюдается автокорреляция не в самих уровнях, а в их отклонениях от теоретических значений, полученных по определенным аналитическим формулам. В таких случаях следует использовать гармонический анализ.

Целью данного анализа являются выявление и измерение периодических колебаний в рядах динамики и автокорреляции в остатках ряда.

Функцию, заданную в каждой точке изучаемого интервала времени, можно представить бесконечным рядом синусоидальных и косинусоидальных функций. Нахождение конечной суммы уровней с использованием функций косинусов и синусов времени называется *гармоническим анализом*.

Другими словами, гармонический анализ представляет собой операцию по выражению заданной периодической функции в виде *ряда Фурье по гармоникам разных порядков*. Каждый член ряда представляет собой слагаемое постоянной величины с функциями косинусов и синусов определенного периода.

В простейшем случае динамика явлений, обладающих периодичностью, может быть аппроксимирована синусоидой:

$$y = A\sin(\alpha t + \beta),$$

где t – время;

A – полуамплитуда колебания, т.е. наибольшее и наименьшее отклонения от оси t;

$t = \pi/\alpha$ – период (длина волны) колебательного движения;

β – начальная фаза колебания.

При $t = 0$ получаем $y_{t=0} = A\sin\beta$.

Аппроксимация динамики экономических явлений рядом Фурье состоит в выборе таких гармонических колебаний, наложение которых друг на друга (сумма) отразит периодические колебания фактических уровней динамического ряда. С помощью ряда Фурье можно представить динамику явлений в виде некоторой функции времени, в которой слагаемые расположены по убыванию периодов:

$$\bar{y}_t = a_0 + \sum_{t=0}^{m} (a_k \cos kt + b_k \sin kt). \qquad (10.38)$$

В этом уравнении величина k определяет гармонику ряда Фурье и может быть взята целым числом (чаще всего от 1 до 4). Параметры уравнения рассчитываются методом наименьших квадратов.

Найдя частные производные этой функции и приравняв их к нулю, получим систему нормальных уравнений, из которой вычислим параметры:

$$a_0 = \frac{1}{n}\sum y; \quad a_k = \frac{2}{n}\sum y\cos\ kt; \quad b_k = \frac{2}{n}\sum y\sin\ kt.$$

Последовательные значения t обычно определяются от 0 с увеличением (приростом), равным $2\pi/n$, где n – число уровней ряда динамики.

Для изучения специфического периодического явления – сезонности берется $n = 12$, по числу месяцев в году.

Тогда ряд динамики годового производства можно записать так:

$$0;\ \frac{p}{6};\ \frac{p}{3};\ \frac{p}{2};\ \frac{2p}{3};\frac{5p}{6};\ \pi;\ \frac{7p}{6};\ \frac{4p}{3};\ \frac{3p}{2};\ \frac{5p}{3};\ \frac{11p}{6}.$$

$$y_0;\ y_1;\ y_2;\ y_3;\ y_4;\ y_5;\ y_6;\ y_7;\ y_8;\ y_9;\ y_{10};\ y_{11}.$$

Для определенных в каждом конкретном случае t находят значения синусов и косинусов разных гармоник, которые для удобства располагают в табл. 10.12.

Таблица 10.12

Коэффициент гармонического анализа месячных наблюдений для расчета параметров a_k и b_k

t	cos$t$	cos2t	cos3t	cos4t	sin$t$	sin2t	sin3t	sin4t
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

Полагая гармоники k соответственно равными 1, 2, 3 и т.д., находим все значения coskt и sinkt. Тогда, например, первая гармоника ряда Фурье примет вид:

$$\bar{y}_t = a_0 + a_1 \cos t + b_1 \sin t,$$

здесь:

$$a_0 = \frac{\sum y}{12};\; a_1 = \frac{\sum y \cos t}{6};\; b_1 = \frac{\sum y \sin t}{6}. \quad (10.39)$$

Ряд Фурье с двумя гармониками:

$$\bar{y}_t = a_0 + a_1 \cos t + b_1 \sin t + a_2 \cos 2t + b_2 \sin 2t, \quad (10.40)$$

где $a_2 = \frac{\sum y \cos 2t}{6};\; b_2 = \frac{\sum y \sin 2t}{6}.$

Исчисление параметров ряда Фурье может производиться и другими способами, а также путем использования различных шаблонов.

Пример. Полагая наличие периодичности, проведем гармонический анализ динамики отклонений от линейной тенденции данных об урожайности ярового ячменя в одном из хозяйств на 1990–2001 гг. (U_i) (табл. 10.13). Проведем расчеты первой гармоники (для значений синусов и косинусов используем данные табл. 10.12).

Отсюда можно определить параметры:

$$a_0 = 0;\; a_1 = 0{,}81;\; b_1 = -0{,}423.$$

Следовательно, 1-я гармоника описывается уравнением

$${}_1\bar{U}_t = 0{,}81 \cos t - 0{,}423 \sin t.$$

Аналогично рассчитываются гармоники 2-го и высших порядков, и значения их последовательно присоединяются к значениям 1-й гармоники. Запишем уравнения искомых отклонений с 2-й и 3-й гармоник.

Таблица 10.13

Отклонения от линейной тенденции данных об урожайности ярового ячменя и расчет параметров a_1 и b_1 в модели сезонной волны

Год	$U_i = y_t - y_t$	t	$\cos\ t$	$\sin\ t$	$a_1\cos\ t$	$b_1\sin\ t$	$\overline{U}_t$
1990	2,1	0	2,1	0	0,81	0	0,81
1991	–2,0	$\pi/6$	–1,732	–1,0	0,701	–0,212	0,49
1992	0,5	$\pi/3$	0,25	0,433	0,405	–0,366	0,04
1993	–0,1	$\pi/2$	0	0,1	0	–0,423	–0,423
1994	2,8	$2\pi/3$	–1,4	2,425	–0,405	–0,366	–0,797
1995	–2,5	$5\pi/6$	2,165	–1,25	–0,701	–0,212	–0,913
1996	–3,1	π	3,1	0	–0,81	0	–0,81
1997	–2,0	$7\pi/6$	1,732	1,0	–0,701	0,121	–0,489
1998	3,4	$4\pi/3$	–1,7	–2,944	–0,405	0,366	–0,04
1999	–0,6	$3\pi/2$	0	0,6	0	0,423	0,423
2000	2,6	$5\pi/3$	1,3	–2,252	0,405	0,366	0,771
2001	–1,1	$11\pi/6$	–0,953	0,55	0,701	0,212	0,912
Итого	0,0	–	4,862	–2,538	–	–	0,0

Для 2-й гармоники:

$$_2\overline{U}_t = 0,81\ \cos t - 0,423\ \sin t - 1,46\ \cos 2t - 0,27\ \sin 2t.$$

Для 3-й гармоники:

$$_3\overline{U}_t = 0,81\cos t - 0,42\sin t - 1,46\cos 2t - 0,27\sin 2t +$$
$$+1,38\cos 3t - 0,32\sin 3t.$$

Подставив в уравнение конкретные значения cost, sin, cos2, sin2, cos3t, sin3t, получим выравненные уровни отклонений урожайности ярового ячменя за 1990–2001 гг. Затем, рассчитав остаточные дисперсии ($\sigma^2_{\text{ост}} = \Sigma(y_i - \overline{y}_t)^2 : n$) для трех гармоник, можно сделать вывод, какая гармоника ряда Фурье наиболее близка к фактическим уровням ряда.

10.7 МЕТОДЫ ВЫЯВЛЕНИЯ ПЕРИОДИЧЕСКОЙ КОМПОНЕНТЫ. МОДЕЛИ СЕЗОННЫХ КОЛЕБАНИЙ

Для проверки предположения о существенности периодической компоненты ряда динамики целесообразно использовать такие критерии случайности, которые имеют наибольшую мощность относительно альтернативной гипотезы о цикличности ряда. Наиболее простым для применения и зрительно понятным является критерий «пиков» и «ям». В основе этого критерия лежит подсчет числа экстремальных точек ряда p, который осуществляется следующим образом:

$$\hat{p} = \sum_{t=2}^{n-1} p_t,$$

где $p_t = \begin{cases} 1, \text{ если } y_{t-1} < y_t > y_{t+1} \\ \text{или } y_{t-1} > y_t < y_{t+1} \\ 0 \text{ в противном случае} \end{cases}$

$t = 1{+}n$;

n – число наблюдений в ряду динамики.

Для случайного ряда математическое ожидание числа экстремальных точек

$$\bar{p} = \frac{2(n-2)}{3}.$$

Проверка гипотезы сводится к сравнению $\bar{p}$ с расчетным значением $\hat{p}$. Если эти значения близки, то можно отказаться от дальнейшей проверки и признать ряд случайным. Если же $\hat{p}$ и $\bar{p}$ значительно отличаются друг от друга, то проводится дальнейшая проверка гипотезы, основанная на подсчете фаз различной длины.

Фазой называется интервал между двумя соседними уровнями, для которых $p_t = 1$. Для определения *длины фазы* l достаточно просто найти разности индексов двух соседних экстремальных точек, затем подсчитать число фаз N_1, N_2, N_3 длин $l_1 = 1$, $l_2 = 2$, $l_3 = 3$. Теоретическое значение числа фаз длины l для случайного ряда следующее:

$$\hat{N}_l = \frac{2(n-l-2)(l^2-3l+1)}{(l+3)!}. \tag{10.41}$$

Естественная процедура проверки случайности сводится к сравнению наблюдаемых значений N_1, N_2, N_3 с теоретическим значением N_l. Однако при небольшом числе наблюдений n критерий χ^2 здесь непосредственно использовать нельзя, так как в этом случае длины фаз l_i не являются независимыми. Доказано, что при разбиении длины фазы на три группы: $l = 1$, $l = 2$, $l = 3$ (две степени свободы) – статистика χ^2 может быть использована в обычной форме ($\nu = 2{,}5$) при $\chi^2 = 6{,}3$. Расчетные значения χ^2 в случае трех групп длин фазы определяются по формуле

$$\chi^2 = \sum_{l=1}^{3} \frac{(N_l - \hat{N}_l)}{N_l}. \tag{10.42}$$

Если $\chi^2 \geq 6{,}3$, то колебания исходного ряда нельзя считать чисто случайными и ряд содержит периодическую составляющую. Этот критерий весьма чувствителен к периодическим колебаниям и имеет практически нулевую эффективность относительно альтернативы наличия тренда, поэтому он может применяться непосредственно к исходному ряду динамики в отличие от других критериев, которые требуют, чтобы из ряда динамики предварительно была выделена систематическая составляющая. После того как установлена периодическая составляющая, проводится ее анализ.

При рассмотрении квартальных или месячных данных многих социально-экономических явлений часто обнаруживаются определенные, постоянно повторяющиеся колебания, которые существенно не изменяются за длительный период времени. Они являются результатом влияния природно-климатических условий, общих экономических факторов, а также ряда многочисленных разнообразных факторов, которые частично являются регулируемыми. В статистике периодические колебания, которые имеют определенный и постоянный период, равный годовому промежутку, носят название *сезонных колебаний,* или *сезонных волн*, а динамический ряд в этом случае называют *тренд-сезонным*, или просто *сезонным рядом динамики*.

Сезонные колебания характеризуются специальными показателями, которые называются *индексами сезонности* (I_s). Совокупность этих показателей отражает сезонную волну. Индексами сезонности являются процентные отношения фактических внутригодовых уровней к постоянной или переменной средней.

Для выявления сезонных колебаний обычно берут данные за несколько лет, распределенные по месяцам. Данные за несколько лет (не менее трех) используют для того, чтобы выявить устойчивую сезонную волну, которая бы не отражала случайные условия одного года.

Для вычисления индексов сезонности применяют различные методы. Если ряд динамики не содержит ярко выраженной тенденции в развитии, то индексы сезонности вычисляют непосредственно по эмпирическим данным без их предварительного выравнивания.

Для каждого месяца рассчитывается средняя величина уровня, например за три года ($\bar{y}_i$), затем из них вычисляется среднемесячный уровень для всего ряда ($\bar{y}$) и в заключение определяется процентное отношение средних для каждого месяца к общему среднемесячному уровню ряда, т.е.

$$I_S = \frac{\bar{y}_i}{\bar{y}} 100\%. \qquad (10.43)$$

Пример. Расчет индексов сезонности покажем на месячных данных о внутригодовой динамике числа браков, расторгнутых населением города за 1999–2001 гг. (табл. 10.14).

Таблица 10.14

Динамика браков, расторгнутых населением города
(расчет индексов сезонности)

Месяц	Число расторгнутых браков				Индекс сезонности $(\bar{y}_i : \bar{y})100\%$
	1999 y_i	2000 y_i	2001 y_i	в среднем за три года $\bar{y}_i$	
А	1	2	3	4	5
Январь	195	158	144	165,7	122,4
Февраль	164	141	136	147,0	108,6
Март	153	153	146	150,7	111,3
Апрель	136	140	132	136,0	100,4
Май	136	136	136	136,0	100,4
Июнь	123	129	125	125,7	92,8
Июль	126	128	124	126,0	93,1
Август	121	122	119	120,7	89,1
Сентябрь	118	118	118	118,0	87,2
Октябрь	126	130	128	128,0	94,5
Ноябрь	129	131	135	131,7	97,3
Декабрь	138	141	139	139,3	102,9
Средний уровень ряда $\bar{y}$	138,77	135,6	131,8	$\bar{y}$ = 135,4	100,0

По данным табл. 10.14 вычислим усредненные значения уровней по одноименным периодам способом средней арифметической простой:

$$\text{январь}: \ \bar{y}_1 = \frac{195+158+144}{3} = \frac{497}{3} = 165{,}7;$$

$$\text{февраль}: \ \bar{y}_2 = \frac{164+141+136}{3} = \frac{441}{3} = 147,0 \text{ и т.д. (графа 4 табл.10.14).}$$

Затем по вычисленным помесячным средним уровням ($\bar{y}_i$) определим общий средний уровень ($\bar{y}$):

$$\bar{y} = \frac{\sum y_i}{n} = \frac{1624,8}{12} = 135,4$$

или

$$\bar{y} = \frac{\sum(\bar{y}_i)}{m} = \frac{406,1}{3} = 135,4,$$

где m – число лет;

$\Sigma\bar{y}_i$ – сумма среднегодовых уровней ряда динамики.

Далее рассчитаем по месяцам года индексы сезонности:

$$\text{январь}: I_{S_1} = \frac{165,7}{135,4} \cdot 100 = 122,4\%;$$

$$\text{февраль}: \ I_{S_2} = \frac{147,0}{135,4} \cdot 100 = 108,6\% \text{ и т.д. (графа 5 табл. 10.14).}$$

Совокупность исчисленных индексов сезонности характеризует сезонную волну браков, расторгнутых населением города, во внутригодовой динамике. Для наглядного получения представления о сезонной волне желательно изобразить полученные данные в виде линейной диаграммы.

Если же ряд динамики содержит определенную тенденцию в развитии, то прежде чем вычислить сезонную волну фактические данные нужно обработать так, чтобы выявить общую тенденцию. Обычно для этого прибегают к аналитическому выравниванию ряда динамики.

Пример. При использовании способа аналитического выравнивания алгоритм вычислений индексов сезонности следующий (табл. 10.15):

- по соответствующему полиному вычисляют для каждого месяца (квартала) выравненные уровни на момент времени (t) (графа 2);

- определяют отношения фактических месячных (квартальных) данных (y_i) к соответствующим выравненным данным ($\bar{y}_t$) в процентах (графа 3);

$$I_i = (y_i : \bar{y}_t) \cdot 100;$$

- находят средние арифметические из процентных соотношений, рассчитанных по одноименным периодам в процентах (графа 4):

$$I_i = (I_1 + I_2 + I_3 + ... + I_n) : n,$$

где n – число одноименных периодов.

Таблица 10.15

Динамика поквартальной продажи безалкогольных напитков в одной из республик за 1999–2001 гг.
(расчет сезонной волны)

Год и квартал	Млн дкл y_i	Теоретические уровни $\bar{y}_t = 88,3 + 0,13t$	Индекс сезонности по каждому кварталу года $(y_i : \bar{y}_t) \cdot 100\%$	Индекс сезонности по одноименным кварталам $\frac{\sum (y_i : \bar{y}_t) \cdot 100\%}{n}$
А	1	2	3	4
1999				
I	58,4	86,2	67,8	67,6
II	125,6	86,6	145,0	140,4
III	108,1	87,0	124,3	121,7
IV	60,8	87,3	69,6	70,3
2000				
I	57,7	87,7	65,8	67,6
II	115,4	88,1	131,0	140,4
III	103,9	88,5	117,4	121,7
IV	60,6	88,9	68,2	70,3

Продолжение

Год и квартал	Млн дкл y_i	Теоретические уровни $\overline{y}_t = 88,3 + 0,13t$	Индекс сезонности по каждому кварталу года $(y_i : \overline{y}_t) \cdot 100\%$	Индекс сезонности по одноименным кварталам $\frac{\sum (y_i : \overline{y}_t) \cdot 100\%}{n}$
А	1	2	3	4
2001				
I	61,8	89,2	69,3	67,6
II	130,2	89,7	145,2	140,4
III	111,0	90,0	123,3	121,7
IV	66,1	90,4	73,1	70,3
Итого	1059,6	1059,6	1200	1200

В общем виде формулу расчета индекса сезонности данным способом можно записать так:

$$I_S = \left[\sum \frac{y_i}{\overline{y}_t} \right] : n. \qquad (10.44)$$

Расчет заканчивается проверкой правильности вычислений индексов. Так как средний индекс сезонности для всех месяцев (кварталов) должен быть 100%, то сумма полученных индексов по месячным данным равна 1200, а сумма по четырем кварталам – 400.

В результате проведенных расчетов в табл. 10.15 получили ряд индексов (графа 4), характеризующих сезонную волну продажи безалкогольных напитков.

В табл. 10.16 дана классификация наиболее распространенных методов измерения сезонной волны.

Таблица 10.16

Классификация методов измерения сезонной волны

Методы измерения сезонной волны, основанные на применении
1. Средней арифметической: метод абсолютных разностей метод отношений средних помесячных к средней за весь период метод отношений помесячных уровней к средней данного года
2. Относительных величин: метод относительных величин метод относительных величин на основе медианы метод У. Персонса (цепной метод)
3. Механического выравнивания: метод скользящих средних метод скользящих сумм и скользящих средних
4. Аналитического выравнивания: выравнивание по прямой выравнивание по параболе и экспоненте выравнивание по ряду Фурье

Подобно сезонной компоненте ряда динамики *циклическая компонента* также представляет собой волнообразные движения (на графике), но она более продолжительна и менее предсказуема, чем сезонные колебания. Сущность классического метода устранения циклической компоненты ряда динамики заключается в исключении (или в усреднении) основной тенденции и сезонной компоненты из ряда динамики, так как при этом остается циклическая и, как правило, *нерегулярная компонента*. Поскольку эти компоненты составляют то, что остается после подобных расчетов, этот метод называется *остаточным.*

Экономисты уделяют большое внимание анализу деловых циклов и их причинам, но мы не ставим своей целью рассмотрение многочисленных теорий этого анализа.

10.8 РЕГРЕССИОННЫЙ АНАЛИЗ СВЯЗНЫХ ДИНАМИЧЕСКИХ РЯДОВ

Многомерные временные ряды, показывающие зависимость результативного признака от одного или нескольких факторных, называют *связными рядами динамики.* Применение методов наименьших квадратов для обработки рядов динамики не требует выдвижения никаких предположений о законах распределения исходных данных. Однако при использовании метода наименьших квадратов для обработки связных рядов следует учитывать наличие автокорреляции (авторегрессии), которая не учитывалась при обработке одномерных рядов динамики, поскольку ее наличие способствовало более плотному и четкому выявлению тенденции развития рассматриваемого социально-экономического явления во времени.

Выявление автокорреляции в уровнях ряда динамики. В рядах динамики экономических процессов между уровнями, особенно близко расположенными, существует взаимосвязь. Ее удобно представить в виде корреляционной зависимости между рядами $y_1, y_2, y_3,..., y_n$ и этим же рядом, сдвинутым относительно первоначального положения на h моментов времени $y_{1+h}, y_{2+h}, y_{3+h}, ..., y_{n+h}$. Временное смещение L называется *сдвигом*, а само явление взаимосвязи – *автокорреляцией*.

Автокорреляционная зависимость особенно существенна между последующими и предшествующими уровнями ряда динамики. Поскольку классические методы математической статистики применимы лишь в случае независимости отдельных членов ряда между собой, то при анализе нескольких взаимосвязанных рядов динамики важно установить наличие и степень их автокорреляции.

Различаются *два вида автокорреляции:*

- автокорреляция в наблюдениях за одной или более переменными;
- автокорреляция ошибок или автокорреляция в отклонениях от тренда.

Наличие последней приводит к искажению величин средних квадратических ошибок коэффициентов регрессии, что затрудняет построение доверительных интервалов для коэффициентов регрессии, а также проверку их значимости.

Автокорреляцию измеряют при помощи *нециклического коэффициента автокорреляции*, который может рассчитываться не только между соседними уровнями, т.е. сдвинутыми на один период, но и между сдвинутыми на любое число единиц времени (L). Этот сдвиг, именуемый *временным лагом*, определяет и порядок коэффициентов автокорреляции: первого порядка (при $L = 1$), второго порядка (при $L = 2$) и т.д. Однако наибольший интерес для исследования представляет вычисление нециклического коэффициента (первого порядка), так как наиболее сильные искажения результатов анализа возникают при корреляции между исходными уровнями ряда (y_{t1}) и теми же уровнями, сдвинутыми на одну единицу времени, т.е. y_{t-1} (или y_{t+1}).

Тогда формулу коэффициента автокорреляции можно записать следующим образом:

$$r_a = \frac{\overline{y_t \cdot y_{t+1}} - \bar{y}_t \cdot \bar{y}_{t+1}}{\text{у}_{y_t} \cdot \text{у}_{y_{t+1}}}, \tag{10.45}$$

где σ_{yt}, σ_{yt+1} – среднее квадратическое отклонение рядов y_t и y_{t+1} соответственно.

Если значение последнего уровня (y_n) ряда мало отличается от первого (y_1), то сдвинутый ряд не укорачивается, его можно условно дополнить, приняв $y_n = y_1$. Тогда $y_t = y_{t+1}$ и $\sigma_{yt} = \sigma_{yt+1}$, поскольку рассчитываются они для одного и того же ряда. При такой замене, т.е. если $\bar{y}_t = \bar{y}_{t+1}$ и $\sigma_{y_t} = \sigma_{y_{t+1}}$, формула коэффициента автокорреляции примет вид:

$$r_a = \frac{\overline{y_t \cdot y_{t+1}} - (\bar{y}_t)^2}{\sigma_{y_t}^2} \quad \text{или} \quad r_a = \frac{\sum y_t \cdot y_{t+1} - n(\bar{y}_t)^2}{\sum y_t^2 - n(\bar{y}_t)^2}. \tag{10.46}$$

Если ряд динамики состоит из уровней, среднее значение которых равно нулю ($\bar{y} = 0$), то выражение (10.46) значительно упрощается:

$$r_a = \frac{\sum_{t=1}^{n-1} y_t \cdot y_{t+1}}{\sum_{t=1}^{n} y_t^2}. \tag{10.47}$$

Для суждения о наличии или отсутствии автокорреляции в исследуемом ряду фактическое значение коэффициентов автокорреляции сопоставляется с табличным (критическим) для 5%-го или 1%-го уровня значимости (вероятность допустить ошибку при принятии нулевой гипотезы о независимости уровней ряда).

Одна из специальных таблиц, в которой определена критическая область проверяемой гипотезы (об отсутствии автокорреляции), составленная Р. Андерсеном в 1942 г., приведена в приложении 12.

Если фактическое значение коэффициента автокорреляции меньше табличного, то гипотеза об отсутствии автокорреляции в ряду может быть принята. Когда же фактическое значение больше табличного, можно сделать вывод о наличии автокорреляции в ряду динамики.

Пример. Покажем расчет коэффициента автокорреляции первого порядка на данных объема собственной продукции общественного питания в одной из областей за 1992–2001 гг. Исходные данные и расчет необходимых величин для формулы (10.46) представлены в табл. 10.17.

По итоговым данным табл. 10.17 находим:

$$1)\ \overline{y_t y_{t+1}} = \frac{49,20}{10} = 4,92;$$

$$2)\ \bar{y} = \frac{21,7}{10} = 2,17;$$

$$3)\ \bar{y}^2 = 2,17^2 = 4,71;$$

$$4)\ \overline{y_t^2} = \frac{51,47}{10} = 5,15;$$

$$5)\ \text{y}_{y_1}^2 = \overline{y_t^2} - \bar{y}^2 - = 5,15 - 4,71 = 0,44$$

и их значения подставляем в формулу коэффициента автокорреляции (10.46):

$$r_a = \frac{4,92 - 4,71}{0,44} = 0,48.$$

Таблица 10.17

Динамика объема собственной продукции общественного питания в одной из областей за 1992–2001 гг.
(расчет коэффициента автокорреляции)

Год	Собственная продукция		Расчетные величины	
	млн руб. y_t	уровни со сдвигом в 1 год y_{t+1}	$y_t\, y_{t+1}$	y^2_t
1992	1,3	1,4	1,82	1,69
1993	1,4	1,5	2,10	1,96
1994	1,5	1,7	2,55	2,25
1995	1,7	2,1	3,57	2,89
1996	2,1	2,2	4,62	4,41
1997	2,2	2,5	5,50	4,84
1998	2,5	2,7	6,75	6,25
1999	2,7	3,0	8,10	7,29
2000	3,0	3,3	9,90	9,00
2001	3,3	1,3	4,29	10,89
Итого	21,7	21,7	49,20	51,47

Приведем сопоставление полученного коэффициента автокорреляции с табличной величиной при численности n = 10. При уровне значимости P = 0,05 (5%-й уровень) величина r_a может только в пяти случаях из ста превышать 0,36 (см. приложение 12). Следовательно, фактический коэффициент автокорреляции, равный 0,48, превысил табличное значение, что говорит о наличии автокорреляции в ряду динамики собственной продукции.

Способ выявления автокорреляции в отклонениях от тренда или от регрессионной модели. Таким способом является использование *критерия Дарбина–Уотсона,* который рассчитывается по формуле:

$$d = \frac{\sum_{t=1}^{n}\left(l_{t+1} - l_t\right)^2}{\sum_{1}^{n} l_t^2}, \tag{10.48}$$

где $l_t = y_t - \overline{y}_t$.

Теоретическое основание применения этого критерия обусловлено тем, что в динамических рядах как сами наблюдения, так и отклонения от них распределяются в хронологическом порядке.

При условии, что отклонения уровней от тенденции (так называемые остатки) случайны, значения D, лежащие в интервале 0 – 4, всегда будут находиться ближе к 2. Если автокорреляция положительная, то $D < 2$; отрицательная $-2 \le D \le 4$. Следовательно, оценки, получаемые по критерию, являются не точечными, а интервальными. Их значения для трех уровней значимости (α = 0,01, α = 0,025 и α = 0,05) с учетом числа наблюдений даны в специальных таблицах (см. приложение 14).

Пример. Рассчитаем критерий Дарбина–Уотсона по данным табл. 10.18.

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

$$\begin{cases} na_0 = \sum y; \\ a_1 \sum t^2 = \sum ty; \end{cases}$$

$$\begin{cases} 16a_0 = 1212; \\ 1360a_1 = 2554; \end{cases}$$

$$\begin{cases} a_0 = 75{,}8; \\ a_1 = 1{,}88. \end{cases}$$

Следовательно, $\overline{y}_i$ = 75,8 + 1,88t.

Таблица 10.18

Динамика среднегодовой стоимости основных промышленно-производственных фондов объединения за 1986–2001 гг.

Год	Млн руб. y_i	t	t^2	yt	$\bar{y}_t$	l_t	l_{t+1}	l_t^2	$l_{t+1}-l_t$	$(l_{t+1}-l_t)^2$
1986	47	–15	225	–705	47,6	–0,6	–0,4	0,36	–0,2	0,04
1987	51	–13	169	–663	51,4	–0,4	–0,1	0,16	0,3	0,09
1988	55	–11	121	–605	55,1	–0,1	0,1	0,01	0,2	0,04
1989	59	–9	81	–531	58,9	0,1	–0,6	0,01	–0,7	0,49
1990	62	–7	49	–434	62,6	–0,6	–0,4	0,36	0,2	0,04
1991	66	–5	25	–330	66,4	–0,4	–0,2	0,16	0,2	0,04
1992	70	–3	9	–210	70,2	–0,2	1,1	0,04	1,3	1,69
1993	75	–1	1	–75	73,9	1,1	1,3	1,21	0,2	0,04
1994	79	1	1	79	77,7	1,3	0,6	1,69	–0,7	0,49
1995	82	3	9	246	81,4	0,6	0,8	0,36	0,2	0,04
1996	86	5	25	430	85,2	0,8	0	0,64	–0,8	0,64
1997	89	7	49	623	89,0	0	0,7	0	0,7	0,49
1998	92	9	81	828	82,7	–0,7	–0,5	0,49	0,2	0,04
1999	96	11	121	1056	96,5	–0,5	–0,2	0,25	0,3	0,09
2000	100	13	169	1300	100,2	–0,7	–0,1	0,04	0,3	0,09
2001	103	15	225	1545	104,0	–1,0		1,0	–	–
Итого	1212	–	1360	2554	–	–	–	6,78	–	4,35

Величина критерия Дарбина–Уотсона $D = 0{,}64$ (4,35 : 6,78), т.е. $D < 2$, и это подтверждает наличие в исходном динамическом ряду положительной автокорреляции.

Способы исключения или уменьшения автокорреляции (авторегрессии) в рядах динамики.

К ним относятся:

- метод включения времени в качестве дополнительного фактора;
- метод последовательных разностей;
- метод авторегрессионных преобразований.

Рассмотрим эти способы. В соответствии с теоремой, доказанной Фришем и Воу, время вводится в систему связных динамических рядов в явной форме в качестве дополнительного фактора, и эта процедура называется *введением фактора времени в уравнение регрессии*. Уровни исходных динамических рядов могут быть представлены показателями в любой форме, в том числе логарифмической, а время всегда вводится в линейной форме. Считается, что введение фактора времени исключает основную тенденцию развития всех явлений, представленных исследуемыми рядами динамики. Доказано, что введение времени аналогично использованию отклонения фактических данных от трендов.

Применение метода наименьших квадратов к обработке многомерных временных рядов не отличается от методологии применения его к обычным статистическим рядам. В рассматриваемом случае минимизируется следующее выражение:

$$S = \Sigma[y_i - f(x_1 x_2,..., x_n, t)]^2 \to \min.$$

Пример. По условным данным о ликвидных активах и собственному капиталу найдем линейное уравнение связи и рассчитаем неизвестные параметры (табл. 10.19).

$$\begin{cases} na_0 + a_1\sum x + a_2\sum t = \sum y; \\ a_0\sum x + a_1\sum x^2 + a_2\sum xt = \sum yx; \\ a_0\sum t + a_1\sum xt + a_2\sum t^2 = \sum yt; \end{cases}$$

$$\begin{cases} 9a_0 + 281a_1 + 43a_2 = 487; \\ 281a_0 + 11287a_1 + 1791a_2 = 18439; \\ 45a_0 + 1791a_1 + 285a_2 = 2926; \end{cases}$$

Таблица 10.19

Расчет параметров уравнения регрессии

Год	Ликвидные активы, млрд руб. x	Собственный капитал, млрд руб. y	t	xy	x^2	ty	t^2	tx	$\bar{y}_{xt}$
1993	9	27	1	243	81	27	1	9	30,5
1994	13	36	2	468	169	72	4	26	32,4
1995	17	29	3	493	289	87	9	51	34,2
1996	22	41	4	902	484	164	16	88	38,7
1997	29	54	5	1566	841	270	25	145	48,3
1998	36	71	6	2566	1296	426	36	216	58,0
1999	44	50	7	2200	1936	350	49	308	70,3
2000	51	81	8	4131	2601	648	64	408	79,9
2001	60	98	9	5880	3600	882	81	540	94,8
Итого	281	487	45	18439	11297	2926	285	791	487,1

$$\begin{cases} a_0 = 15,63; \\ a_1 = 2,61; \\ a_2 = -8,60; \end{cases}$$

$$\overline{y}_{xt} = 15,63 + 2,61x - 8,6t.$$

При исключении автокорреляции *методом последовательных разностей* подвергаются обработке методом наименьших квадратов не сами уровни исходных рядов $y_t, y_{t+1}, ..., y_{t+n}$, и $x_t, x_{t+1}, ..., x_{t+n}$, а последовательные разности между ними:

$$\begin{array}{ll} \Delta y_1 = y_t - y_{t-1}; & \Delta x_t = x_t - x_{t-1}; \\ \Delta y_2 = y_{t-1} - y_{t-2}; & \Delta x_2 = x_{t-1} - x_{t-2}; \\ \dots\dots\dots & \dots\dots\dots \\ \dots\dots\dots & \dots\dots\dots \\ \Delta y_k = y_{t-k} - y_{t-k-1}; & \Delta x_k = x_{t-k} - x_{t-k-1}. \end{array}$$

При использовании этого метода исходят из предположения, что все разности между уровнями динамических рядов, начиная с первой, будут содержать только случайную компоненту. Причем первые разности содержат случайную компоненту в линейной форме, вторые – описываемую параболой 2-го порядка, третьи – показательной функцией.

Метод авторегрессионных преобразований заключается в том, что определяют уравнение связи между отклонениями от тенденций двух связных рядов динамики:

$$\begin{array}{ll} y_1 - \overline{y}_{t1}; & x_1 - \overline{x}_{t1}; \\ y_2 - \overline{y}_{t2}; & x_2 - \overline{x}_{t2}; \\ \dots\dots\dots & \dots\dots\dots \\ \dots\dots\dots & \dots\dots\dots \\ y_n - \overline{y}_{tn}; & x_n - \overline{x}_{tn}. \end{array}$$

В этом случае также получают уравнения регрессии, не искаженные влиянием автокорреляции.

Введение времени в качестве дополнительной переменной является наиболее действенным способом обработки связных рядов динамики. Во всяком случае, при линейной связи между исследуемыми рядами этот способ более точен, чем использование последовательных разностей или отклонений от трендов.

При обработке методом наименьших квадратов последовательных разностей или отклонений от трендов исследователь имеет дело с чисто случайными величинами, взаимосвязь между которыми является часто весьма сомнительной, так как исключение в обоих случаях тенденций нарушает существование причинно-следственной связи между явлениями.

10.9 КОРРЕЛЯЦИЯ РЯДОВ ДИНАМИКИ

При изучении развития явления во времени часто возникает необходимость оценить степень взаимосвязи в изменениях уровней двух или более рядов динамики различного содержания, но связанных между собой.

Эта задача решается методами коррелирования:

- уровней ряда динамики;
- отклонений фактических уровней от тренда;
- последовательных разностей, т.е. путем исчисления парного коэффициента корреляции.

Расчет парного коэффициента корреляции по уровням ряда динамики. Этот расчет правильно показывает тесноту связи между рядами динамики лишь в том случае, если в каждом из них отсутствует автокорреляция.

В этом случае величину коэффициента корреляции находят по формуле:

$$r = \frac{\overline{xy} - \bar{x}\,\bar{y}}{\text{у}_x \cdot \text{у}_y}, \qquad (10.49)$$

где x_i – уровни факторного ряда динамики;

y_i – уровни результативного ряда динамики.

Следовательно, прежде чем коррелировать ряды динамики (по уровням), необходимо проверить каждый из рядов на наличие или отсутствие в них автокорреляции (при помощи коэффициента автокорреляции, описанного в разделе 10.8). В случае наличия автокорреляции между уровнями ряда последняя должна быть устранена.

Рассмотрим способы ее исключения в рядах динамики.

Расчет парного коэффициента корреляции по отклонениям фактических уровней от выравненных по уравнению (тренду). Этот способ состоит в том, что коррелируют не сами уровни, а отклонения фактических уровней от выравненных, отражающих тренд, т.е. коррелируют остаточные величины. Для этого каждый ряд динамики выравнивают по определенной, характерной для него аналитической формуле, затем из эмпирических уровней вычитают выравненные (т.е. находят $d_x = x_t - \overline{x}_t$; $d_y = y_t - \overline{y}_t$) и определяют тесноту связи между рассчитанными отклонениями (d_x и $d_у$) по формуле

$$r_{d_x d_y} = \frac{\sum d_x \cdot d_y}{\sqrt{\sum d_x^2 \cdot \sum d_y^2}}. \qquad (10.50)$$

Расчет парного коэффициента корреляции по абсолютным отклонениям уровней ряда динамики. Исключить влияние автокорреляции можно путем вычитания из каждого уровня предшествующего ему, т.е. находя разности уровней ($y_i - y_{i-1}$). Алгебраически легко показать, что при переходе от уровней к их разностям исключается влияние общей тенденции на колеблемость. При этом при изменении уровней по прямой можно коррелировать первые разности, при изменении по *параболе* *n*-го порядка – *n*-е разности. Формула коэффициента разностей, используемая для измерения тесноты связи между исследуемыми рядами, имеет вид:

$$r_{\Delta_x \Delta_y} = \frac{\sum \Delta_x \cdot \Delta_y}{\sqrt{\sum \Delta_x^2 \cdot \sum \Delta_y^2}}. \qquad (10.51)$$

Коэффициент корреляции, рассчитанный для измерения тесноты зависимости изменения уровней двух рядов, является своего рода средним, обобщающим показателем. Однако для длительного периода эта зависимость не является постоянной, она может меняться во времени. Поэтому чтобы судить о том, в какие периоды зависимость между изменениями уровней двух рядов слабая или сильная, рекомендуется рассчитывать серию скользящих коэффициентов корреляции для определенного интервала времени.

10.10 ЭЛЕМЕНТЫ ПРОГНОЗИРОВАНИЯ И ИНТЕРПОЛЯЦИИ

Исследование динамики социально-экономических явлений, выявление и характеристика основной тенденции развития и моделей взаимосвязи дают основание для прогнозирования – определения будущих размеров уровня экономического явления.

Важное место в системе методов прогнозирования занимают статистические методы. Применение прогнозирования предполагает, что закономерность развития, действующая в прошлом (внутри ряда динамики), сохранится и в прогнозируемом будущем, т.е. прогноз основан на *экстраполяции*. Экстраполяция, проводимая в будущее, называется *перспективной* и в прошлое – *ретроспективной*. Обычно, говоря об экстраполяции рядов динамики, подразумевают чаще всего перспективную экстраполяцию.

Теоретической основой распространения тенденции на будущее является известное свойство социально-экономических явлений, называемое *инерционностью.* Именно инерционность позволяет выявить сложившиеся взаимосвязи как между уровнями динамического ряда, так и между группой связных рядов динамики. На основе рядов динамики получаются весьма надежные прогнозы, если уровни ряда динамики сопоставимы и получены на основе единой методологии.

Применение экстраполяции в прогнозировании базируется на следующих предпосылках:

- развитие исследуемого явления в целом описывается плавной кривой;
- общая тенденция развития явления в прошлом и настоящем не претерпет серьезных изменений в будущем.

Поэтому надежность и точность прогноза зависят от того, насколько близкими к действительности окажутся эти предположения, а также как точно удастся охарактеризовать выявленную в прошлом закономерность. Экстраполяцию следует рассматривать как начальную стадию построения окончательных прогнозов. Механическое использование экстраполяции может стать причиной погрешности и неправильных выводов. Всегда следует учитывать все необходимые условия, предпосылки и гипотезы, связывая их с содержательным экономико-теоретическим анализом.

Разумеется, чем шире раздвигаются временные рамки прогнозирования, тем очевиднее становится недостаточность простого экстраполяционного метода (изменение тенденций, неизвестны точки поворота кривых, влияние новых факторов и т.д.). В этом случае динамичность экономических явлений и процессов вступает в противоречие с инерционностью их развития. Так как анализируемые экономические ряды динамики нередко относительно короткие, то и временной горизонт экстраполяции не может быть бесконечным. Поэтому чем короче *срок экстраполяции* (*период упреждения*), тем более надежные и точные результаты (при прочих равных условиях) дает прогноз. За короткий период не успевают сильно измениться условия развития явления и характер его динамики.

Экстраполяцию в общем виде можно представить формулой:

$$\hat{y}_{i+T} = f(y_i, T, a_j), \tag{10.52}$$

где $\hat{y}_{i+T}$ – прогнозируемый уровень;
y_i – текущий уровень прогнозируемого ряда;
T – период упреждения;
a_j – параметр уравнения тренда.

В зависимости от того, какие принципы и исходные данные положены в основу прогноза, выделяют следующие элементарные методы экстраполяции:

- средний абсолютный прирост;
- средний темп роста;
- экстраполяцию на основе выравнивания рядов по какой-либо аналитической формуле.

Прогнозирование по *среднему абсолютному приросту* может быть выполнено в том случае, если есть уверенность считать общую тенденцию линейной, т.е. метод основан на предположении о равномерном изменении уровня (под равномерностью понимается стабильность абсолютных приростов).

Для нахождения интересующего нас аналитического выражения тенденции на любую дату *t* необходимо определить средний абсолютный прирост и последовательно прибавлять его к последнему уровню ряда столько раз, на сколько периодов экстраполируется ряд, т.е. экстраполяцию можно сделать по следующей формуле:

$$\hat{y}_{i+t} = y_i + \overline{\text{Д}} \cdot t, \tag{10.53}$$

где $\hat{y}_{i+t}$ – экстраполируемый уровень, (i+t) – номер этого уровня (года);
i – номер последнего уровня (года) исследуемого периода, за который рассчитан $\overline{\Delta}$;
t – срок прогноза (период упреждения);
$\overline{\Delta}$ – средний абсолютный прирост.

Однако следует иметь в виду, что использование среднего абсолютного прироста для прогноза возможно только при следующем условии:

$$\text{у}^2_{\text{ост}} \le \rho^2, \tag{10.54}$$

где $\rho^2 = \frac{1}{2} \cdot \frac{\sum \text{Д}_i^2}{n}$; $\text{у}^2_{\text{ост}} = \frac{\sum (y_i - \overline{y}_{\overline{\text{Д}}})^2}{n}$.

Пример. По данным об удельном весе прибытия воздушных судов, выполненных без опоздания по сравнению с расписанием за 1991–2001 гг. (см. табл. 10.10) экстраполируем ряд на 2002–2003 гг. Средний абсолютный прирост равен (–0,17%), остаточная дисперсия: $\sigma^2_{\text{ост}} = 0{,}003$; $\rho^2 = 0{,}02$. Следовательно, $\sigma^2_{\text{ост}} \le \rho^2$; основное условие выполняется, можно делать прогноз:

$$\hat{y}_{2002} = 89{,}9 + (-0{,}17) = 89{,}7\%;$$
$$\hat{y}_{2003} = 89{,}7 + (-0{,}17) = 89{,}5\%.$$

Прогнозирование по *среднему темпу роста* осуществляется в случае, когда есть основание считать, что общая тенденция ряда характеризуется показательной (экспоненциальной) кривой. Для нахождения тенденции необходимо определить средний коэффициент роста, возведенный в степень, соответствующую периоду экстраполяции, т.е. по формуле:

$$\hat{y}_{i+t} = y_i \cdot \bar{k}_p^t, \tag{10.55}$$

где y_i – последний уровень ряда динамики;
t – срок прогноза;
$\bar{k}_p$ – средний коэффициент роста.

Если же ряду динамики свойственна иная закономерность, то данные, полученные при экстраполяции на основе среднего темпа роста, будут отличаться от данных, рассчитанных другими способами экстраполяции.

Рассмотренные способы экстраполяции тренда, будучи простейшими, в то же время являются и самыми приближенными.

Поэтому наиболее распространенным методом прогнозирования считают *аналитическое выражение тренда*. При этом для выхода за границы исследуемого периода достаточно продолжить значения независимой переменной времени (t).

При таком подходе к прогнозированию предполагается, что размер уровня, характеризующего явление, формируется под воздействием множества факторов, причем не представляется возможным выделить отдельно их влияние. В связи с этим ход развития связывается не с какими-либо конкретными факторами, а с течением времени, т.е. $y = f(t)$.

Экстраполяция дает возможность получить точечное значение прогноза. Точное совпадение фактических данных и прогностических точечных оценок, полученных путем экстраполяции кривых, характеризующих тенденцию, имеет малую вероятность. Возникновение отклонений фактических уровней ряда динамики от выравненных по уравнению тренда объясняется следующими причинами:

- выбранная для прогнозирования кривая не является единственно возможной для описания тенденции. Можно подобрать такую кривую, которая дает более точные результаты;
- построение прогноза осуществляется на основе ограниченного числа исходных данных. Кроме того, каждый исходный уровень обладает еще случайной компонентой. Поэтому и кривая, по которой осуществляется экстраполяция, будет содержать случайную компоненту;

- тенденция характеризует лишь движение среднего уровня ряда динамики, поэтому отдельные наблюдения от него отклоняются. Если такие отклонения наблюдались в прошлом, то они будут наблюдаться и в будущем.

Любой статистический прогноз носит приближенный характер. Поэтому целесообразно определение доверительных интервалов прогноза.

Величина доверительного интервала определяется следующим образом:

$$\hat{y}_t \pm t_\alpha \cdot \text{у}_{\bar{y}_t},$$

где $\sigma_{\bar{y}_t}$ – средняя квадратическая ошибка тренда;
$\hat{y}_t$ – расчетное значение уровня;
t_α – доверительная величина.

Вместо t_α-критерия Е. М. Четыркин[1] предлагает брать коэффициент K.

Пример. Необходимо провести прогноз на 2002–2005 гг. по данным табл. 10.8 об урожайности зерновых культур в хозяйстве.

Таблица 10.20

Прогнозные значения урожайности зерновых культур в хозяйстве на 2002–2005 гг.

Год	t	y_t	K	$\text{у}_{\bar{y}_t} \cdot K$	$\hat{y}_{i+t} \pm \text{у}_{\hat{y}_t} \cdot K$
А	1	2	3	4	5
2002	8	16,2	2,0153	3,6	12,6–19,8
2003	9	16,3	2,0621	3,7	12,6–20,0
2004	10	16,5	2,1131	3,8	12,7–20,3
2005	11	16,7	2,1680	3,9	12,8–20,6

[1] Значения K взяты из книги: *Четыркин Е. М.* Статистические методы прогнозирования. – М.: Статистика, 1975. – С. 183.

Для экстраполяции используем уравнение тренда (см. табл. 10.11), полученное по прямой: $\bar{y}_t = 14{,}8 + 0{,}17t$. Подставив соответствующие значения t в наше уравнение, получим точечные прогнозы на 2002–2005 гг. (табл. 10.20 графа 2). Для построения интервальных прогнозов рассчитаем среднюю квадратическую ошибку тренда ($\sigma_{yt} = 1{,}797$) и значения k^*. Результаты прогноза представлены в табл. 10.20.

При анализе рядов динамики иногда приходится прибегать к определению некоторых неизвестных уровней внутри данного ряда динамики, т.е. к *интерполяции*.

Интерполяция может производиться на основе среднего абсолютного прироста, среднего темпа роста и с помощью аналитического выравнивания. Она также основана на том или ином предположении о тенденции изменения уровней, но характер этого прогноза несколько иной: здесь уже не приходится предполагать, что тенденция, характерная для прошлого, сохранится и в будущем.

При интерполяции считается, что ни выявленная тенденция, ни ее характер не претерпели существенных изменений в том промежутке времени, уровень (уровни) которого нам не известен. Такое предположение обычно является более обоснованным, чем предположение о будущей тенденции.

ОСНОВНЫЕ ПОНЯТИЯ

Абсолютный прирост измеряет абсолютную скорость роста (или снижения) уровня ряда за единицу времени (месяц, квартал, год и т.д.). Он показывает, на сколько единиц увеличился или уменьшился уровень ряда по сравнению с базисным, т.е. за тот или иной промежуток времени.

Абсолютный прирост скорости (замедления) или ускорения – абсолютный показатель, который определяет, на сколько данная скорость больше (меньше) предыдущей.

Абсолютный размер 1% прироста – абсолютный показатель, который определяет, какое содержание имеется в 1% прироста, сколько весом 1%.

Автокорреляция – корреляционная зависимость между последовательными (т.е. соседними) значениями уровней динамического ряда y_1 и y_2; y_2 и y_3 и т.д.

Авторегрессия – регрессия, учитывающая влияние предыдущих уровней ряда на последующие.

Аналитическое выравнивание динамического ряда проводится при помощи математической формулы, отражающей общую тенденцию ряда.

Интервальный ряд динамики – ряд числовых значений определенного статистического показателя, характеризующего размеры изучаемого явления за определенные промежутки (периоды, интервалы) времени.

Интерполяция – приближенный расчет уровней, лежащих внутри ряда динамики, но почему-либо неизвестных.

Коэффициент опережения (замедления) – относительный показатель, характеризующий сравнение динамических рядов, относящихся к двум пространственным объектам (странам, республикам и т.д.).

Лаг – промежуток времени отставания одного явления от другого, связанного с ним.

Механическое сглаживание – метод нахождения плавных уровней ряда динамики путем использования скользящих средних. Различают метод невзвешенных и взвешенных скользящих средних.

Моментный ряд динамики – ряд числовых значений определенного статистического показателя, характеризующего размеры изучаемого явления на определенные даты, моменты времени.

Основная тенденция (тренд) – достаточно плавное и устойчивое изменение уровня явления во времени, более или менее свободное от случайных колебаний. Основную тенденцию можно представить либо аналитически, в виде уравнения (модели) тренда, либо графически.

Ряд динамики – ряд числовых значений определенного статистического показателя в последовательные моменты или периоды времени.

Ряд Фурье дает возможность выделить периодические (сезонные) колебания, свойственные динамике многих экономических явлений.

Сезонная компонента ряда динамики – внутригодичные колебания, имеющие более или менее регулярный характер. Их мерой обычно является индекс сезонности.

Смыкание рядов динамики – один из методов приведения несопоставимых рядов к сопоставимым путем прямого пересчета уровней с помощью специальных коэффициентов или относительных величин.

Средний абсолютный прирост – показатель, характеризующий среднюю абсолютную скорость роста (или снижения) уровня за отдельные периоды времени. Он показывает, на сколько единиц увеличился (или уменьшился) уровень по сравнению с предыдущим в среднем за единицу времени (в среднем ежегодно, ежемесячно и т.д.).

Средний темп прироста – относительный показатель, выраженный в процентах и показывающий, на сколько увеличился (или уменьшился) уровень по сравнению с предыдущим в среднем за единицу времени (в среднем ежегодно, ежемесячно и т.п.).

Средний темп роста – относительный показатель, выраженный в форме коэффициента и показывающий, во сколько раз увеличился уровень по сравнению с предыдущим в среднем за единицу времени (в среднем ежегодно, ежеквартально и т.п.).

Средняя хронологическая интервального ряда исчисляется по формуле средней арифметической, причем при равных интервалах применяется средняя арифметическая простая, а при неравных – средняя арифметическая взвешенная.

Средняя хронологическая моментного ряда исчисляется как сумма всех уровней ряда, поделенного на число членов ряда без одного, причем первый и последний члены ряда берутся в половинном размере.

Темп прироста – относительный показатель, характеризующий величину прироста (снижения).

Темп роста – относительный показатель, характеризующий интенсивность роста (или снижения). Он показывает, сколько процентов составляет уровень данного периода по сравнению с базисным или предыдущим уровнем, т.е. характеризует относительную скорость изменения уровня ряда в единицу времени.

Уровень ряда динамики – абсолютная (относительная, средняя) величина каждого члена динамического ряда.

Хронологическая средняя – средняя, исчисленная из уровней динамического ряда.

Экстраполяция – нахождение уровней за пределами изучаемого ряда, т.е. продление ряда на основе выявленной закономерности изменения уровней в изучаемый отрезок времени.

ТЕСТЫ

1. Ряд динамики характеризует:

а) структуру совокупности по какому-либо признаку;

б) изменение характеристики совокупности в пространстве;

в) изменение характеристики совокупности во времени.

2. Уровень ряда динамики – это:

а) определенное значение варьирующего признака в совокупности;

б) величина показателя на определенную дату или момент времени;

в) величина показателя за определенный период времени.

3. Средний уровень интервального ряда динамики определяется как:

а) средняя арифметическая;

б) средняя гармоническая;

в) средняя хронологическая.

4. Средний уровень моментного ряда исчисляется как средняя арифметическая взвешенная при:

а) равноотстоящих уровнях между датами;

б) неравноотстоящих уровнях между датами.

5. Средний уровень моментного ряда исчисляется как средняя хронологическая при:

а) равноотстоящих уровнях между датами;

б) неравноотстоящих уровнях между датами.

6. Если сравниваются смежные уровни ряда динамики, показатели называются:

а) ценными;

б) базисными.

7. Если все уровни ряда динамики сравниваются с одним и тем же уровнем, показатели называются:

а) цепными;

б) базисными.

8. Абсолютный прирост исчисляется как:

а) отношение уровней;

б) разность уровней ряда.

9. Темп роста исчисляется как:

а) отношение уровней ряда;

б) разность уровней ряда.

10. Основная тенденция представляет собой изменение ряда динамики:

а) равномерно повторяющееся через определенные промежутки времени внутри ряда;

б) определяющее какое-то общее направление развития.

11. Сезонные колебания представляют собой изменения ряда динамики, равномерно повторяющиеся:

а) через определенные промежутки времени с годичным интервалом;

б) внутри года.

12. Для выявления основной тенденции развития используются:

а) метод укрупнения интервалов;

б) метод скользящей средней;

в) метод аналитического выравнивания;

г) ряд Фурье.

13. При сравнении динамики взаимосвязанных показателей применяются приемы:

а) приведения рядов динамики к одному основанию;

б) смыкания динамических рядов.

14. С целью приведения несопоставимых уровней ряда динамики к сопоставимому виду применяются приемы:

а) приведения рядов динамики к одному основанию;

б) смыкания динамических рядов.

15. Индексы сезонности можно рассчитать как отношение фактического уровня за тот или иной месяц к:

а) среднемесячному уровню за год;

б) выравненному уровню за тот же месяц;

в) среднемесячному выравненному уровню за год.

16. Можно ли изучить взаимосвязи социально-экономических явлений по данным рядов динамики?

а) да;

б) нет.

17. Влияет ли автокорреляция на результаты измерения связи?

в) да;

г) нет.

ЛИТЕРАТУРА

Андерсен Т. Статистический анализ временных рядов. – М.: Мир, 1976. – 155 с.

Венецкий И.Г., Венецкая В.И. Основные математико-статистические понятия и формулы в экономическом анализе. – М.: Статистика, 1979. – 477 с.

Вайну Я.Я. Корреляция рядов динамики. – М.: Статистика, 1977. – 119 с.

Громыко Г.Л. Теория статистики. – М.: Изд-во Инфра-М, 2000. – 413 с.

Езекиэл М., Фокс К. Методы корреляции и регрессии. – М.: Статистика, 1966. – 558 с.

Закс Л. Статистическое оценивание. – М.: Статистика, 1976. – 598 с.

Иващенко Г.А., Кильдишев Г.С., Шмойлова Р.А. Статистическое изучение основной тенденции и взаимосвязи в рядах динамики. – Томск: Изд-во Томского ун-та, 1985. – 168 с.

Казинец Л.С. Темпы роста и абсолютные приросты. – М.: Статистика, 1975. – 191 с.

Кильдишев Г.С., Френкель А.А. Анализ временных рядов и прогнозирование. – М.: Статистика, 1973. – 103 с.

Ковалева Л.Н. Многофакторное прогнозирование на основе рядов динамики. – М.: Статистика, 1980. – 103 с.

Половников В.А. Анализ и прогнозирование транспортной работы морского флота. – М.: Транспорт, 1983. – 222 с.

Тинтнер Г. Введение в эконометрию. – М.: Статистика, 1965. – 361 с.

Ферстер Э., Ренц Б. Методы корреляционного и регрессионного анализа. – М.: Финансы и статистика, 1983. – 302 с.

Хан Г., Шапиро С. Статистические модели в инженерных задачах. – М.: Мир, 1969. – 395 с.

Четыркин Е.М. Статистические методы прогнозирования. – М.: Статистика, 1975. – 183 с.

ГЛАВА 11
СТАТИСТИЧЕСКИЙ АНАЛИЗ СТРУКТУРЫ

11.1 ПОНЯТИЕ И ВИДЫ СТРУКТУРЫ СОЦИАЛЬНО-ЭКОНОМИЧЕСКИХ ЯВЛЕНИЙ

Структура (лат. structure – строение, расположение, порядок) в наиболее общем смысле представляет собой совокупность устойчивых внутренних связей объекта, обеспечивающих его целостность и тождественность самому себе, т.е. сохранение основных свойств при различных внешних и внутренних изменениях. В статистике под *структурой* понимают совокупность элементов социально-экономических явлений, обладающих определенной устойчивостью внутригрупповых связей, при сохранении основных свойств, характеризующих эту совокупность как целое. Статистическая структура – это распределение различных частей в пределах общего для них качества, распределение составляющих совокупность единиц по количественному или качественному признаку. Подавляющая масса изучаемых статистикой сложных объектов, процессов или явлений в сфере промышленного или сельскохозяйственного производства, финансов, коммерции, демографии, в социальной и других областях может быть исследована с точки зрения их внутренней структуры по тому или иному признаку.

Статистический анализ структуры непосредственно связан с группировкой данных. Если основанием структуры выступает качественный признак, то процесс группировки, как правило, не вызывает затруднений. Группировка по количественному признаку обычно сложнее, так как требует обоснованного установления границ перехода одного качества в другое. Анализ структуры совокупности одновременно по нескольким количественным признакам обычно проводится на основе методов многомерной классификации.

Статистические приемы и методы анализа позволяют проводить исследование конкретных социально-экономических структур в определенных условиях места и времени, которое заключается, прежде всего, в точном количественном измерении и соизмерении, выявлении пропорций и закономерностей. Структура сложного социально-экономического явления всегда обладает той или иной степенью подвижности, имеет свойство меняться с течением времени как в количественном, так и в качественном отношении. Поэтому большое практическое значение имеют изучение структуры в динамике, оценка структурных сдвигов, выявление и характеристика основных тенденций развития. Углубленный анализ структуры, требующий применения корреляционно-регрессионного анализа и индексного метода, предполагает также изучение факторов, воздействующих на структуру, и оценку влияния структуры на взаимосвязанные с ней результативные показатели.

Классификация структур предполагает в первую очередь их разделение на два основных вида по временному фактору, на моментные структуры и интервальные. *Моментные структуры* характеризуют строение социально-экономических явлений по состоянию на определенные моменты времени и отображаются посредством моментных относительных показателей (как правило, на начало или на конец периода). Моментными являются, например, такие структуры, как население по полу, возрасту, уровню образования; основные фонды по отраслям народного хозяйства и по формам собственности; парк транспортных средств и т.п.

Интервальные структуры характеризуют строение социально-экономических явлений за определенные периоды времени (дни, недели, месяцы, кварталы, годы). Они отображаются интервальными относительными показателями. Например, такие структуры, как экспорт и импорт, товарооборот, материальные затраты.

Статистика имеет дело как с *фактическими*, реально существующими структурами, так и со структурами *перспективными*, *прогнозными*, *оптимальными* и *стандартизованными*. Последние представляют собой какие-либо гипотетические (условные) или фактические структуры, принятые в качестве эталонных для расчета и сравнения стандартизованных показателей. Например, для сравнения уровней рождаемости, смертности, заболеваемости и т.п. по двум и более регионам рассчитывают стандартизованные коэффициенты на основе некоторой стандартизованной структуры, в качестве которой может использоваться возрастная структура населения в целом по стране.

11.2
ПОКАЗАТЕЛИ СТРУКТУРЫ И СТРУКТУРНЫХ СДВИГОВ

Различают две формы выражения относительных показателей структуры (глава 6): долю и удельный вес, который представляет собой долю, выраженную в процентах. В дальнейшем изложении удельный вес (долю) i-й части совокупности ($i = \overline{l,k}$) в j-й период времени или по состоянию на j-й момент времени обозначим как d_{ij}, при этом все расчеты будем производить над величинами, выраженными в процентах.

Рассмотрим показатели, характеризующие изменения структуры, или структурные сдвиги. Отметим прежде всего, что термин *структурные сдвиги* применим лишь при исследовании структурных различий во времени. При территориальных же сравнениях структуры, а также при сравнении фактической структуры со стандартизованной более корректным является термин «структурные различия». Рассматриваемые в данной главе методы динамических сравнений большей частью применимы и для сравнений территориальных.

Для статистической оценки структурных сдвигов за два или более периодов используются *две группы показателей*:

- показатели, основывающиеся на разностях между удельными весами одноименных частей совокупности;
- показатели, базирующиеся на отношениях удельных весов одноименных частей совокупности.

Исходным в 1-й группе является *показатель абсолютного прироста удельного веса* i-й части совокупности (d_i), показывающий, на какую величину в долях единицы или процентах возросла или уменьшилась данная структурная часть в j-й период по сравнению с (j–1) периодом [1]. Этот показатель рассчитывается по следующей формуле:

$$\Delta d_i = d_{ij} - d_{ij-1}. \tag{11.1}$$

[1] Здесь и далее при исследовании моментных структур под периодами будут подразумеваться моменты времени; термин «абсолютный» при исследовании приростов удельных весов принято брать в кавычки, так как абсолютными эти показатели являются по методологии расчета, но не по единицам измерения.

Знак прироста показывает направление изменения удельного веса данной структурной части («+» – увеличение, «–» – уменьшение), а его величина – конкретное значение этого изменения в процентных пунктах.

Так как сумма удельных весов всех частей совокупности в любой момент времени всегда равна строго 100%, то при каких-либо изменениях в структуре одна часть приростов удельных весов всегда будет иметь положительный знак, а другая – отрицательный. Сумма же всех k приростов для совокупности в целом всегда равна нулю:

$$\sum_{i=1}^{k} \Delta d_i = \sum_{i=1}^{k} (d_{ij} - d_{ij-1}) = \sum_{i=1}^{k} d_{ij} - \sum_{i=1}^{k} \Delta d_{ij-1} = 100\% - 100\%.$$

Пример. Рассмотрим динамику структуры внешнеторгового оборота за 1997–2000 гг. (табл. 11.1).

Таблица 11.1

Структура внешнеторгового оборота РФ в 1997–2000 гг., %

Составляющие внешнеторгового оборота	Удельный вес, %				2000 г. по сравнению с 1999 г.	
	1997	1998	1999	2000	прирост удельного веса, проц. пунктов (гр. 4 – гр. 3)	рост удельного веса (гр.4 : гр.3)
А	1	2	3	4	5	6
Экспорт	55,3	56,4	65,7	70,1	4,4	1,067
Импорт	44,7	43,6	34,3	29,9	–4,4	0,872
Всего	100	100	100	100	0	х

Для характеристики структурных сдвигов в 2000 г. по сравнению с 1999 г. по формуле (11.1) рассчитаны «абсолютные» приросты удельного веса (графа 5): экспорт возрос на 4,4 проц. пункта и соответственно на эту же величину снизился удельный вес импорта.

Для 2-й группы показателей структурных сдвигов исходным является *темп роста удельного веса* (Tp_{di}), представляющий собой отношение удельного веса i-й части в j-й период времени к удельному весу этой же части в предшествующий период:

$$Tp_{d_i} = \frac{d_{ij}}{d_{ij-1}}. \tag{11.2}$$

Темп роста удельного веса всегда является положительной величиной. Однако если в совокупности произошли какие-либо структурные изменения, часть темпов роста будет больше единицы, а часть – меньше. В то же время их среднее значение, взвешенное по базисным удельным весам, всегда строго равно единице:

$$\overline{T}p_d = \frac{\sum_{i=1}^{k} Tp_{d_i} \cdot d_{ij-1}}{\sum_{i=1}^{k} d_{ij-1}} = \frac{\sum_{i=1}^{k} d_{ij}}{\sum_{i=1}^{k} d_{ij-1}} = \frac{100\%}{100\%}.$$

Рассчитанные темпы роста удельных весов структурных частей внешнеторгового оборота приведены в графе 6 табл. 11.1. Проверим правильность их вычисления расчетом средневзвешенного значения:

$$\overline{T}p_d = \frac{1,067 \cdot 0,657 + 0,872 \cdot 0,343}{0,657 + 0,343} = 1.$$

Мы рассмотрели показатели структурных сдвигов за один интервал между двумя периодами. Если же изучаемая структура представлена данными за три и более периода, появляется необходимость в динамическом осреднении приведенных выше показателей.

Средний «абсолютный» прирост удельного веса ($\overline{\Delta} d_i$) i-й структурной части за n периодов определяется по формуле:

$$\overline{\Delta}d_i = \frac{d_{in} - d_{i1}}{n-1}, \qquad (11.3)$$

где i – структурная часть ($i = \overline{1,k}$);
j – временной интервал ($j = \overline{1,n}$).

При отсутствии в расчетах ошибок и соблюдении достаточной точности вычислений сумма средних «абсолютных» приростов удельных весов всех k структурных частей совокупности, так же как и сумма их приростов за один временной интервал, должна быть равна нулю:

$$\sum_{i=1}^{k} \overline{\Delta}d_i = \sum_{i=1}^{k} \frac{d_{in} - d_{i1}}{n-1} = \frac{1}{n-1}\left(\sum_{i=1}^{k} d_{in} - \sum_{i=1}^{k} d_{i1}\right) = \frac{1}{n-1}(100\% - 100\%).$$

Анализируя структуру внешнеторгового оборота за все четыре года, мы можем определить годовые и общие за весь период приросты удельного веса каждой структурной части, кроме того, на основе формулы (11.3) найти средний годовой прирост, представляющий собой общий прирост, деленный на число интервалов между рассматриваемыми периодами (между четырьмя годами три интервала). Например, средний «абсолютный» прирост удельного веса экспорта равен:

$$\overline{\Delta}d_1 = \frac{70{,}1 - 55{,}3}{3} = 4{,}93 \text{ проц. пункта.}$$

Средний «абсолютный» прирост удельного веса импорта составит:

$$\overline{\Delta}d_2 = \frac{29{,}9 - 44{,}7}{3} = -4{,}93 \text{ проц. пункта.}$$

Таким образом, удельный вес экспорта ежегодно увеличивался в среднем почти на 5 процентных пунктов, а импорт на эту же величину снижался.

Определяющим показателем среднего «абсолютного» прироста удельного веса является общий прирост удельного веса каждой i-й части. Например, если мы заменим сумму всех индивидуальных приростов удельного веса экспорта суммой средних приростов, общий прирост останется без изменений:

$$1{,}1 + 9{,}3 + 4{,}4 = 4{,}93 + 4{,}93 + 4{,}93 = 14{,}8.$$

Относительным показателем, характеризующим изменение удельного веса i-й структурной части за n периодов, является *средний темп роста удельного веса*. При расчете этого показателя используется формула средней геометрической:

$$\overline{T}p_{d_i} = \sqrt[n-1]{Tp_{d_{i1}} \cdot Tp_{d_{i2}} \cdot Tp_{d_{i3}} \cdot \ldots \cdot Tp_{d_{in-1}}}. \tag{11.4}$$

Подкоренное выражение этой формулы представляет собой последовательное произведение индивидуальных темпов роста удельного веса за все временные интервалы. После проведения несложных алгебраических преобразований формула (11.4) примет следующий вид:

$$\overline{T}p_{d_i} = \sqrt[n-1]{\frac{d_{in}}{d_{i1}}}. \tag{11.5}$$

Используя эту формулу, за рассматриваемый период определим средний годовой темп роста удельного веса экспорта:

$$\overline{T}p_d = \sqrt[3]{\frac{70{,}1}{55{,}3}}. = \sqrt[3]{1{,}268} = 1{,}082$$

и импорта

$$\overline{T}p_d = \sqrt[3]{\frac{29{,}9}{44{,}7}}. = \sqrt[3]{0{,}669} = 0{,}875.$$

Итак, экспорт ежегодно увеличивался в среднем в 1,08 раза, а импорт снижался на 0,12 своей величины (1–0,88).

Определяющим показателем при расчете среднего темпа роста удельного веса i-й структурной части является общий темп роста ее удельного веса за весь рассматриваемый период. Это означает, что если заменить произведение индивидуальных темпов роста произведением средних значений, то общий прирост удельного веса данной части останется без изменений.

Проиллюстрируем этот тезис, используя темпы роста удельного веса экспорта:

$$Tp_{d\,21} = \frac{56,4}{55,3} = 1,020;$$

$$Tp_{d\,22} = \frac{65,7}{56,4} = 1,165;$$

$$Tp_{d\,23} = \frac{70,1}{65,7} = 1,067;$$

$$1,020 \cdot 1,165 \cdot 1,067 = 1,082 \cdot 1,082 \cdot 1,082.$$

При анализе структуры исследуемого объекта или явления за ряд периодов также можно определить средний, или общий (в данном случае одно и то же), удельный вес каждой i-й части за весь рассматриваемый временной интервал. Однако для этого одних лишь относительных данных об удельных весах структурных частей недостаточно, необходимо располагать исходными данными о размерах этих частей в абсолютном выражении (табл. 11.2).

Используя эти данные, *средний удельный вес* любой i-й структурной части можно определить по формуле:

$$\bar{d}_i = \frac{\sum_{j=1}^{n} x_{ij}}{\sum_{j=1}^{n} \sum_{i=1}^{k} x_{ij}} \cdot 100\%. \quad (11.6)$$

Рассмотрим табл. 11.3.

Таблица 11.2

Исходные данные для определения удельных весов

Структурная часть	Период				
	1	2	3	…	n
1	x_{11}	x_{12}	x_{13}	…	x_{1n}
2	x_{21}	x_{22}	x_{23}	…	x_{2n}
3	x_{31}	x_{32}	x_{33}	…	x_{3n}
.	.	.	.	…	.
.	.	.	.	…	.
k	x_{k1}	x_{k2}	x_{k3}	…	x_{kn}
Итого	$\sum_{i=1}^{k} x_{i1}$	$\sum_{i=1}^{k} x_{i2}$	$\sum_{i=1}^{k} x_{i3}$	…	$\sum_{i=1}^{k} x_{in}$

Таблица 11.3

Производство продукции промышленным предприятием

Вид продукции	Квартал							
	I		II		III		IV	
	млн руб.	%	млн руб.	%	млн руб.	%	млн руб.	%
А	80	20	800	80	100	20	1200	80
В	320	80	200	20	400	80	300	20
Итого	400	100	1000	100	500	100	1500	100

Располагая лишь относительными показателями структуры, можно предположить, что в среднем за год средний удельный вес объема продукции двух видов был равен:

$$\bar{d}_A = \frac{20+80+20+80}{4} = 50\%;$$

$$\bar{d}_B = \frac{80+20+80+20}{4} = 50\%.$$

Однако при этом мы не учли, что 1% в каждом квартале соответствует различная в абсолютном выражении величина. Рассчитаем истинный средний удельный вес по обоим видам продукции, используя формулу (11.6):

$$\bar{d}_A = \frac{80+800+100+1200}{400+1000+500+1500} \cdot 100\% = 64{,}1\%;$$

$$\bar{d}_B = \frac{320+200+400+300}{400+1000+500+1500} \cdot 100\% = 35{,}9\%.$$

Полученные значения существенно отличаются от предполагаемых.

Если в абсолютном выражении известен только общий объем признака за каждый период, а также удельные веса структурных частей, для расчета среднего удельного веса используют среднюю арифметическую взвешенную.

$$\bar{d}_i = \frac{\sum_{j=1}^{n}\left(d_{ij} \cdot \sum_{i=1}^{k} x_{ij}\right)}{\sum_{j=1}^{n}\sum_{i=1}^{k} x_{ij}} \cdot 100\%. \qquad (11.7)$$

На основе этой формулы для нашего примера мы получим те же самые значения средних удельных весов:

$$\bar{d}_A = \frac{20 \cdot 400 + 80 \cdot 100 + 20 \cdot 500 + 80 \cdot 1500}{400+1000+500+1500} = 64{,}1\%;$$

$$\bar{d}_B = \frac{80 \cdot 400 + 20 \cdot 1000 + 80 \cdot 500 + 20 \cdot 1500}{400 + 1000 + 500 + 1500} = 35,9\%.$$

Пример. Определим средний удельный вес экспорта в общем объеме внешнеторгового оборота РФ за 1997–2000 гг. Для этого приведенных в табл. 11.1 относительных показателей недостаточно. Необходимо привлечь данные об объемах внешнеторгового оборота за рассматриваемый период в стоимостном выражении: 1997 г. – 161,0 млрд долл., 1998 г. – 132,9 млрд долл., 1999 г. – 115,2 млрд долл., 2000 г. – 150,4 млрд долл. Произведем расчет с учетом этой дополнительной информации:

$$\bar{d}_э = \frac{55,3 \cdot 161,0 + 56,4 \cdot 132,9 + 65,7 \cdot 115,2 + 70,1 \cdot 150,4}{161,0 + 132,9 + 115,2 + 150,4} = 61,7\%.$$

Таким образом, за четыре года средний удельный вес экспорта составил 61,7%, а импорта соответственно 38,3%.

11.3 СВОДНАЯ ОЦЕНКА СТРУКТУРНЫХ ИЗМЕНЕНИЙ ВО ВРЕМЕНИ И ПРОСТРАНСТВЕ

В предшествующем разделе 11.2 были рассмотрены показатели, позволяющие измерить те количественные изменения, которые произошли с каждой отдельно взятой частью изучаемой совокупности. В ряде случаев перед исследователем встает задача оценить в целом структурные изменения в изучаемом социально-экономическом явлении за определенный временной интервал, характеризующие подвижность или, наоборот, стабильность, устойчивость данной структуры. Как правило, это необходимо для сравнения динамики одной и той же структуры в различные периоды или нескольких структур, относящихся к разным объектам. Во втором случае число структурных частей у разных объектов необязательно должно совпадать.

Среди предлагаемых для этих целей обобщающих показателей наиболее легко интерпретируется *линейный коэффициент «абсолют-*

ных» структурных сдвигов ($\overline{\Delta}_{d_1 - d_0}$), представляющий собой сумму приростов удельных весов, взятых без учета знака, деленную на число структурных частей:

$$\overline{\Delta}_{d_1-d_0} = \frac{\sum_{i=1}^{k} \left| d_{ij} - d_{ij-1} \right|}{k}. \tag{11.8}$$

Этот показатель отражает то среднее изменение удельного веса (в процентных пунктах), которое имело место за рассматриваемый временной интервал.

Для решения данной задачи применяется также *квадратический коэффициент «абсолютных» структурных сдвигов* ($\sigma_{d_1-d_0}$), который резче реагирует на происходящие в совокупности структурные изменения:

$$\text{у}_{d_1-d_0} = \sqrt{\frac{\sum_{i=1}^{k} \left(d_{ij} - d_{ij-1} \right)^2}{k}}. \tag{11.9}$$

Линейный и квадратический коэффициенты «абсолютных» структурных сдвигов позволяют получить сводную оценку скорости изменения удельных весов отдельных частей совокупности. Для сводной характеристики интенсивности изменения удельных весов используется *квадратический коэффициент относительных структурных сдвигов* (σ_{d_1/d_0})**:**

$$\text{у}_{\frac{d_1}{d_0}} = \sqrt{\sum_{i=1}^{k} \frac{(d_{ij} - d_{ij-1})^2}{d_{ij-1}}} \cdot 100. \tag{11.10}$$

Данный показатель отражает тот средний относительный прирост удельного веса (в процентах), который наблюдался за рассматриваемый период.

Пример. Рассмотрим применение формул (11.8) – (11.10), используя структуру потребительских расходов населения РФ (табл. 11.4).

Для расчета линейного коэффициента «абсолютных» структурных сдвигов за I период (с 1998 по 1999 г.) и за II период (с 1999 по 2000 г.) воспользуемся итогами в графах 4 и 7 табл. 11.4.

$$\bar{\Delta}^{(\mathrm{I})}_{d_1-d_0} = \frac{2,0}{4} = 0,5 \text{ проц. пункта;}$$

$$\bar{\Delta}^{(\mathrm{II})}_{d_1-d_0} = \frac{8,6}{4} = 2,2 \text{ проц. пункта.}$$

Итак, с 1998 по 1999 г. удельный вес отдельных статей потребительских расходов населения изменился несущественно – в среднем на 0,5 проц. пункта. В течение же следующего года «абсолютные» структурные сдвиги увеличились и структура потребительских расходов претерпела более сильные изменения. Эти выводы подтверждаются квадратическими коэффициентами «абсолютных» структурных сдвигов (необходимые промежуточные расчеты выполнены в графах 5 и 8 табл. 11.4).

$$у^{(\mathrm{I})}_{d_1-d_0} = \sqrt{\frac{1,34}{4}} = 0,6 \text{ проц. пункта;}$$

$$у^{(\mathrm{II})}_{d_1-d_0} = \sqrt{\frac{31,38}{4}} = 2,8 \text{ проц. пункта.}$$

Теперь определим величину квадратических коэффициентов относительных структурных сдвигов, воспользовавшись итогами в графах 6 и 9 табл. 11.4:

$$у^{(\mathrm{I})}_{\frac{d_1}{d_0}} = \sqrt{0,077 \cdot 100} = 2,8\%;$$

$$у^{(\mathrm{II})}_{\frac{d_1}{d_0}} = \sqrt{0,791 \cdot 100} = 8,9\%.$$

Таблица 11.4

Структура потребительских расходов домашних хозяйств

Статья расходов	Удельный вес, % к итогу			Расчетные графы						
	1998	1999	2000	$\lvert d_{i2}-d_{i1}\rvert$	$(d_{i2}-d_{i1})^2$	$\frac{(d_{i2}-d_{i1})^2}{d_{i1}}$	$\lvert d_{i3}-d_{i2}\rvert$	$(d_{i3}-d_{i2})^2$	$\frac{(d_{i3}-d_{i2})^2}{d_{i2}}$	$\lvert d_{i3}-d_{i1}\rvert$
А	1	2	3	4	5	6	7	8	9	10
Домашнее и общественное питание	53,3	53,7	49,4	0,4	0,16	0,003	4,3	18,49	0,344	3,9
Непродовольственные товары	30,2	30,8	34,3	0,6	0,36	0,012	3,5	12,25	0,398	4,1
Алкоголь	2,6	2,5	2,5	0,1	0,01	0,004	–	–	–	0,1
Услуги	13,9	13,0	13,8	0,9	0,81	0,058	0,8	0,64	0,049	0,1
Все расходы	100,0	100,0	100,0	2,0	1,34	0,077	8,6	31,38	0,791	8,2

Расчеты показывают, что если за первый период удельный вес каждой статьи расходов в среднем изменился почти на 3% своей величины, то в следующем периоде относительные структурные изменения были в 3 раза сильнее.

Для сводной оценки структурных изменений в исследуемой совокупности в целом за рассматриваемый временной интервал, охватывающий несколько дней, месяцев, кварталов или лет, наиболее удобным является *линейный коэффициент «абсолютных» структурных сдвигов за n периодов* $\left(\overline{\Delta}^{(n)}_{d_1-d_0}\right)$:

$$\overline{\Delta}^{(n)}_{d_1-d_0} = \frac{\sum_{i=1}^{k}\left|d_{in}-d_{i1}\right|}{k(n-1)}. \qquad (11.11)$$

Используя итог графы 10 табл. 11.4, получим:

$$\overline{\Delta}^{(n)}_{d_1-d_0} = \frac{8,2}{4\cdot 2} = 1,0 \text{ проц. пункта.}$$

Таким образом, за рассматриваемый период среднее годовое изменение по всем статьям расходов составило 1 проц. пункт.

Необходимо отметить, что показатель (11.11) может использоваться как для сравнения динамики двух и более структур, так и для анализа динамики одной и той же структуры за разные по продолжительности периоды времени.

11.4
СТАТИСТИЧЕСКИЕ ПОКАЗАТЕЛИ КОНЦЕНТРАЦИИ И ЦЕНТРАЛИЗАЦИИ

Одна из задач статистического анализа структуры заключается в определении степени концентрации изучаемого признака по единицам совокупности или в оценке неравномерности его распределения. Такая неравномерность может иметь место в распределении доходов по группам населения, жилой площади по группам семей, прибыли по группам предприятий, капитала по группам банков и т.д. При исследовании неравномерности распределения изучаемого признака по территории понятие «концентрация» обычно заменяется понятием «локализация».

Оценка степени концентрации наиболее часто осуществляется по *кривой концентрации (Лоренца)* и рассчитываемым на ее основе характеристикам. Для построения кривой концентрации необходимо иметь частотное распределение единиц исследуемой совокупности и взаимосвязанное с ним частотное распределение изучаемого признака. При этом для удобства вычисления и повышения аналитичности данных единицы совокупности обычно разбиваются на равные группы: 10 групп – по 10% единиц в каждой, 5 групп – по 20% единиц и т.д.

Пример. Проведем сравнительный анализ концентрации доходов населения условного региона в 2001 и 2002 гг. Для этого используем приведенную в табл. 11.5 группировку, в которой все население разбито

Таблица 11.5

Распределение доходов населения региона, %

Группы населения, ранжированные по уровню среднедушевого дохода (по 10% от общей численности населения)	2001		2002	
	Удельный вес в совокупном доходе	Накопленная частота	Удельный вес в совокупном доходе	Накопленная частота
1	4,3	4,3	3,2	3,2
2	6,1	10,4	4,8	8,0
3	7,1	17,5	6,1	14,1
4	8,1	25,6	7,2	21,3
5	9,1	34,7	8,4	29,7
6	10,1	44,8	9,7	39,4
7	11,2	56,0	11,3	50,7
8	12,6	68,6	13,2	63,9
9	14,3	82,9	15,8	79,7
10	17,1	100,0	20,3	100,0
Итого	100,0	444,8	100,0	410,0

на 10 равных групп таким образом, что 1-я группа объединяет 10% населения с наименьшими доходами, 2-я группа – следующие по уровню доходов 10% населения и так далее до последней, 10-й группы, объединяющей 10% населения с наибольшими доходами.

Предварительный анализ этих данных позволяет заключить, что концентрация доходов, или дифференциация населения по уровню доходов, за рассматриваемый период усилилась. Так, если в 2001 г. 10% беднейших слоев населения располагали 4,3% совокупного дохода, а 10% наиболее обеспеченных людей – 17,1% дохода, то в 2002 г. эти показатели составили соответственно 3,2 и 20,3%. Если в 2001 г. доходы менее обеспеченной половины населения составляли 34,7% общей величины, то в 2002 г. эта доля снизилась до 29,7%.

Кривая Лоренца строится в прямоугольной системе координат. На оси абсцисс откладываются накопленные частоты объема совокупности, а на оси ординат – накопленные частоты объема признака. Полученная при соединении точек кривая линия и будет характеризовать степень концентрации.

Если распределение является строго равномерным, то первые 10% единиц обладают 10% объема признака, первые 20% – соответственно 20% объема признака и т.д. Такое распределение отображается прямой, проходящей из нижнего левого угла графика к верхнему правому углу, и называется *линией равномерного распределения.* Чем сильнее концентрация изучаемого признака, тем заметнее кривая Лоренца отклоняется от линии равномерного распределения, и наоборот, чем слабее концентрация, тем ближе будет кривая к прямой.

Построенные по данным табл. 11.5 кривые концентрации доходов населения представлены на рис. 11.1.

Следует отметить, что кривая концентрации может сколько угодно приближаться к линии равномерного распределения, но никогда не пересечет ее. В нашем примере это объясняется следующим образом. Так как население ранжировано по уровню индивидуальных доходов от минимального значения к максимальному, то суммарный доход первых 10% населения не может превышать 10% общего совокупного дохода (иначе какая-либо другая 10%-ная группа с более высокими доходами должна была бы иметь менее 10% общего совокупного дохода, что невозможно). Суммарный доход первых 20% населения по той же причине не может превышать 20% общего совокупного дохода и т.д.

Степень концентрации определяется площадью фигуры А, ограниченной линией равномерного распределения и кривой концентрации. Чем больше площадь А и чем меньше площадь В, тем степень концентрации выше.

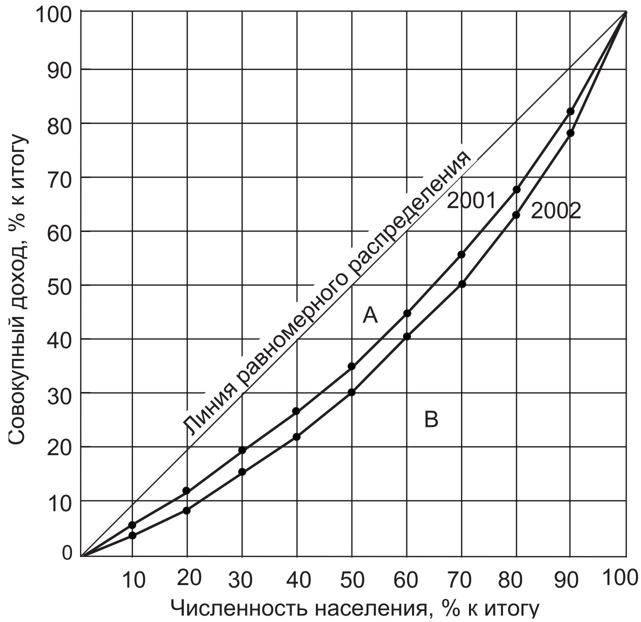

Рис. 11.1. Кривые концентрации доходов населения региона в 2001–2002 гг.

На сравнении площади А с площадью треугольника, расположенного ниже линии равномерного распределения, основан *коэффициент Джини* (G):

$$G = 1 - 2\sum_{i=1}^{k} d_{x_i} d_{y_i}^{H} + \sum_{i=1}^{k} d_{x_i} d_{y_i}, \qquad (11.12)$$

где d_{x_i} – доля i-й группы в общем объеме совокупности (в нашем примере в численности населения);

d_{y_i} – доля i-й группы в общем объеме признака (в нашем примере в доходах);

$d_{y_i}^{H}$ – накопленная доля i-й группы в общем объеме признака.

Если исследуемая совокупность разделена на 10 равновеликих групп и частоты выражены в процентах, то данный коэффициент принимает следующий вид:

$$G = 110 - 0,2\sum_{i=1}^{k} d_{y_i}^{H}. \qquad (11.13)$$

Используем итоги в графах 2 и 4 табл. 11.5 для наших распределений:

$$G_{01} = 110 - 0,2 \cdot 444,8 = 21,0\%;$$

$$G_{02} = 110 - 0,2 \cdot 410,0 = 28,0\%.$$

Итак, если в 2001 г. концентрация доходов населения составляла 21%, то за год она возросла на 7 проц. пунктов.

В последние годы Госкомстат РФ при характеристике степени дифференциации доходов использует 20%-е группы населения.

Пример. Расслоение населения страны по доходам, сложившееся в 2000 г., отражают данные табл. 11.6.

Таблица 11.6

Распределение общего объема денежных доходов населения РФ в 2000 г.

20%-е группы населения	Удельный вес в совокупном доходе, %	Накопленная частота, %
1-я (с наименьшими доходами)	6,0	6,0
2-я	10,4	16,4
3-я	14,8	31,2
4-я	21,2	52,4
5-я (с наивысшими доходами)	47,6	100,0
Итого	100,0	206,0

Для оценки уровня дифференциации в этом случае удобнее использовать следующее преобразование коэффициента Джини:

$$G = 120 - 0,4\sum_{i=1}^{k} d_{y_i}^{H}. \qquad (11.14)$$

Используя данные последней графы табл. 11.6, получим:

$$G = 120 - 0{,}4 \cdot 206 = 37{,}6\%.$$

Для сравнения отметим, что в 1991 г. коэффициент Джини составлял 26,0%.

Для построения кривой концентрации и расчета показателей концентрации необязательно иметь группы с равной численностью единиц.

Пример. Рассмотрим решение этой задачи на основе структурной группировки с различными объемами групп (табл. 11.7).

Построенная по этим данным кривая концентрации будет иметь следующий вид (рис. 11.2). Из графика видно, что в данном случае кривая Лоренца вогнута значительно сильнее, чем в предыдущем примере, и, следовательно, концентрация изучаемого признака выше. Это подтверждает и коэффициент Джини, рассчитанный по формуле (11.12):

$$G = 1 - 2 \cdot 0{,}2537 + 0{,}1685 = 0{,}6611\ (66{,}11\%).$$

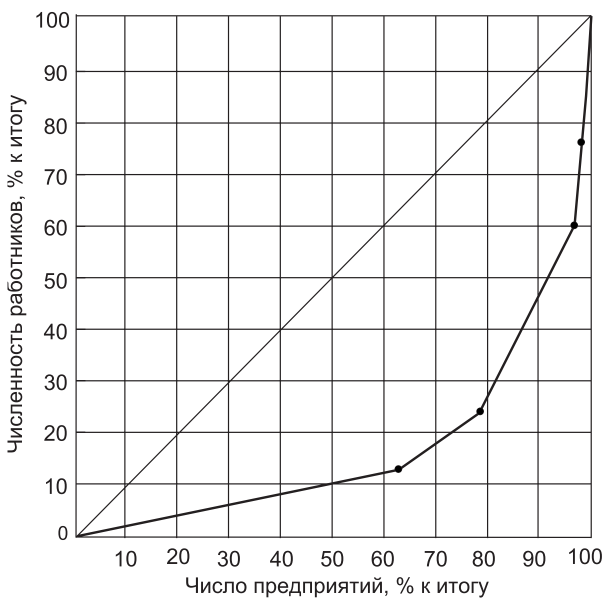

Рис. 11.2. Кривая концентрации работников на приватизированных предприятиях

Таблица 11.7

Численность занятых на приватизированных промышленных предприятиях

Группы предприятий по численности занятых, чел.	Число предприятий			Численность занятых			$d_{x_i}d_{y_i}$	$d_{x_i}d^{н}_{y_i}$
	в абсолютных единицах	доля в общем итоге d_{x_i}	накопленная доля $d^{н}_{x_i}$	млн чел.	доля в общем итоге d_{yi}	накопленная доля $d^{н}_{y_i}$		
1 – 499	4941	0,632	0,632	0,99	0,125	0,125	0,0790	0,0790
500 – 999	1173	0,150	0,782	0,84	0,106	0,231	0,0159	0,0346
1000 – 4999	1408	0,180	0,962	2,92	0,369	0,600	0,0664	0,1080
5000 – 9999	202	0,026	0,988	1,36	0,172	0,772	0,0045	0,0201
10 000 и более	94	0,012	1,000	1,81	0,228	1,000	0,0027	0,0120
Всего	7 818	1,000	Х	7,92	1,000	Х	0,1685	0,2537

Получить количественную оценку степени концентрации можно и на основе сравнения по группам долей или удельных весов объема совокупности (d_{xi}) с долями или удельными весами объема признака (d_{yi}). По данным табл. 11.7 рассчитаем сумму абсолютных разностей удельных весов:

$$|63{,}2 - 12{,}5| + |15{,}0 - 10{,}6| + |18{,}0 - 36{,}9| + |2{,}6 - 17{,}2| + \\ + |1{,}2 - 22{,}8| = 110{,}2\%.$$

При равномерном распределении удельные веса должны совпадать. Так, при равномерном распределении или близком к равномерному распределению работников по предприятиям на первых 63,2% предприятий было бы занято примерно 63,2% работников, на следующих 15,0% предприятий – примерно 15,0% и т.д.

При максимально возможном неравномерном распределении на предприятиях, входящих в первые четыре группы, было бы занято 0 рабочих, а на предприятиях последней, 5-й группы – все 100%. Тогда суммарная величина абсолютных разностей была бы следующей:

$$|63{,}2 - 0| + |15{,}0 - 0| + |18{,}0 - 0| + |2{,}6 - 0| + |1{,}2 - 100{,}0| = 197{,}6\%.$$

Теоретически возможное максимальное значение этой суммы при неравномерном распределении равно 200%. На соотношении фактической суммы с максимально возможной основан *коэффициент Лоренца (L),* также характеризующий степень концентрации:

$$L = \frac{\sum_{i=1}^{k}\left|d_{x_i} - d_{y_i}\right|}{200} \cdot 100\% = \frac{\sum_{i=1}^{k}\left|d_{x_i} - d_{y_i}\right|}{2}. \qquad (11.15)$$

Для нашего примера получим:

$$L = \frac{110{,}2}{2} = 55{,}1\%.$$

Рассчитанное значение данного коэффициента, как и полученное ранее значение коэффициента Джини, свидетельствует о высокой степени концентрации работников на приватизированных предприятиях.

Если под *концентрацией* понимается степень неравномерности распределения изучаемого признака, не связанная ни с общим объемом совокупности, ни с численностью отдельных групп, то *централизация* означает сосредоточение объема признака у отдельных единиц (объема продукции данного вида на отдельных предприятиях, капитала в отдельных банках и т.п.). Обобщающий показатель централизации (I_z), или коэффициент Герфиндаля, имеет следующий вид:

$$I_z = \sum_{i=1}^{n} \left(\frac{x_i}{\sum\limits_{i-1}^{n} x_i} \right), \tag{11.16}$$

где x_i – значение признака i-й единицы совокупности;

$\sum\limits_{i=1}^{n} x_i$ – объем признака всей совокупности;

n – объем совокупности (число входящих в нее единиц).

Максимального значения, равного 1, данный коэффициент достигает лишь в том случае, когда совокупность состоит только из одной единицы, обладающей всем объемом признака. Минимальное значение коэффициента приближается к нулю, но никогда его не достигает.

Пример. Рассмотрим следующие условные данные, приведенные в табл. 11.8.

Используя формулы (11.15) и (11.16), определим степень концентрации и централизации производства данного вида продукции в регионе A:

$$L_A = \frac{|25-25| + |25-25| + |50-50|}{2} = 0;$$

$$I_{z_A} = (0,25)^2 + (0,25)^2 + (0,25)^2 \cdot 2 = 0,25.$$

Таблица 11.8

Группировка предприятий отрасли по форме собственности

Форма собственности	Регион *А*						Регион *Б*					
	Число предприятий		Объем продукции				Число предприятий		Объем продукции			
	ед.	%	млн руб.	%	млн руб. на 1 предприятие (гр. 3 : гр. 1)	доля в общем объеме (гр. 5 : Σгр. 3)	ед.	%	млн руб.	%	млн руб. на 1 предприятие (гр. 9 : гр. 7)	доля в общем объеме (гр. 11 : Σгр.9)
А	1	2	3	4	5	6	7	8	9	10	11	12
Федеральная	1	25	200	25	200	0,25	5	25	1000	25	200	0,05
Субъекта РФ	1	25	200	25	200	0,25	5	25	2000	50	400	0,10
Муниципальная	2	50	400	50	200	0,25	10	50	1000	25	100	0,025
Итого	4	100	800	100	X	X	20	100	4000	100	X	X

В этом регионе мы имеем равномерное распределение производимой продукции в полном соответствии с числом предприятий различных форм собственности, поэтому степень концентрации нулевая. В то же время (так как все производство сосредоточено всего на четырех предприятиях) имеет место сравнительно высокая централизация. Выполним аналогичные расчеты по региону Б:

$$L_{Б} = \frac{|25-25|+|25-50|+|50-25|}{2} = 25\%;$$

$$I_{z_{Б}} = (0,25)^2 \cdot 5 + (0,10)^2 \cdot 5 + (0,025)^2 \cdot 10 = 0,07.$$

В данном регионе наблюдается концентрация производства во второй группе предприятий, вследствие этого соответствующий коэффициент увеличился до 25%. Но общее число предприятий в регионе *Б* в 5 раз больше, что обусловливает небольшую степень централизации. Отметим, что оценка степени централизации представляет практический интерес только при очень небольших объемах совокупности.

ОСНОВНЫЕ ПОНЯТИЯ

Интервальная структура – структура, характеризующая строение социально-экономических явлений за определенные периоды времени (дни, недели, месяцы, кварталы, годы).

Концентрация – неравномерность распределения изучаемого признака внутри совокупности, не связанная с общим ее объемом.

Моментная структура – структура, характеризующая строение социально-экономических явлений по состоянию на определенные моменты времени (на определенную дату, начало или конец периода).

Структура – совокупность элементов социально-экономических явлений, обладающих определенной устойчивостью внутригрупповых связей, при сохранении основных свойств, характеризующих эту совокупность как целое.

Для статистического анализа динамики структуры используются две группы показателей: показатели, базирующиеся на приростах удельных весов одноименных частей совокупности, и показатели, основу которых составляют темпы роста удельных весов. Каждая из названных групп объединяет индивидуальные показатели, рассчитываемые для отдельных структурных частей, и обобщающие показатели, рассчитываемые в целом по совокупности. Как те, так и другие могут быть исчислены за один период или как средние за несколько периодов.

Централизация – сосредоточение объема признака у отдельных единиц или неравномерность его распределения с учетом объема совокупности. Понятие «централизация» признака отличается от понятия «концентрация» признака.

Так, при нулевой концентрации вполне возможна сильная централизация и, наоборот, на фоне слабой централизации допустима высокая концентрация.

ТЕСТЫ

1. К какому виду по временному фактору относится половозрастная структура населения:

а) моментная;

б) интервальная.

2. Может ли темп роста удельного веса быть отрицательной величиной:

а) не может;

б) может в случае снижения удельного веса.

3. Средний «абсолютный» прирост удельного веса определяется по формуле:

а) средней арифметической;

б) средней геометрической.

4. Должна ли сумма средних темпов роста всех структурных частей исследуемой совокупности быть строго равной 100%:

а) должна;

б) не должна.

5. Должна ли сумма средних удельных весов всех структурных частей исследуемой совокупности быть строго равной 100%:

а) должна;

б) не должна.

6. Могут ли линейные и квадратические обобщающие коэффициенты, рассчитанные для сравнения структурных изменений в двух совокупностях, привести к противоположным выводам:

а) могут;

б) не могут.

7. Может ли кривая концентрации частично совпадать с линией равномерного распределения:

а) может;

б) не может.

8. Коэффициент Лоренца позволяет получить сводную оценку структурных сдвигов:

а) в процентных пунктах;

б) в процентах к удельному весу наибольшей структурной части;

в) в процентах к предельной величине структурных сдвигов.

9. Может ли высокая концентрация сопровождаться сильной централизацией:

а) может;

б) не может.

10. В каких границах может находиться коэффициент централизации в том случае, если все производство сосредоточено только на двух предприятиях:

а) от 0 до 0,5;

б) от 0,5 до 1,0.

ЛИТЕРАТУРА

Венецкий И.Г., Венецкая В.И. Основные математико-статистические понятия и формулы в экономическом анализе. – М.: Статистика, 1979. – 477 с.

Казинец Л.С. Измерение структурных сдвигов в экономике. – М.: Экономика, 1969. – 164 с.

Казинец Л.С. Темпы роста структурных сдвигов в экономике. – М.: Экономика, 1981. – 184 с.

Популярный экономико-статистический словарь-справочник / Под ред. И.И. Елисеевой. – М.: Финансы и статистика, 1993. – 192 с.

Рабинович П.М. Некоторые вопросы статистического исследования структуры социально-экономических явлений // Вестник статистики. – 1975. – № 10. – С. 18–24.

Статистический словарь/ Под ред. М.А. Королева. – М.: Финансы и статистика, 1989. – 623 с.

ГЛАВА 12
ЭКОНОМИЧЕСКИЕ ИНДЕКСЫ

12.1 ПОНЯТИЕ ЭКОНОМИЧЕСКИХ ИНДЕКСОВ. КЛАССИФИКАЦИЯ ИНДЕКСОВ

Индексы относятся к важнейшим обобщающим показателям. Слово «индекс» имеет несколько значений: показатель, указатель, опись, реестр. Оно используется как понятие в математике, экономике, в метеорологии и других науках.

В статистике под *индексом* понимается относительный показатель, который выражает соотношение величин какого-либо явления во времени, в пространстве или дает сравнение фактических данных с любым эталоном (план, прогноз, норматив и т.д.).

В международной практике индексы принято обозначать символами i и 1 (начальная буква латинского слова *index*). Буквой «i» обозначаются индивидуальные (частные) индексы, буквой «1» –общие индексы. Знак внизу справа означает период: 0 – базисный; 1 – отчетный.

Используются определенные символы для обозначения индексируемых показателей:

- q – количество (объем) произведенной продукции (или количество проданного товара) данного вида в натуральном выражении;
- p – цена единицы продукции или товара;
- z – себестоимость единицы продукции;
- t – затраты рабочего времени (труда) на производство единицы продукции данного вида, т.е. трудоемкость единицы изделия;
- T – общие затраты рабочего времени (труда) на производство продукции данного вида или численность работников предприятия, фирмы и т.д.;
- $w = q : T$ – производство продукции данного вида в единицу времени или в расчете на одного рабочего, т.е. уровень производительности труда в стоимостном выражении;
- v – выработка продукции в натуральном выражении на одного рабочего или в единицу времени;
- $F = zq$ – общие затраты на производство продукции данного вида;
- $Q = pq$ – общая стоимость произведенной продукции данного вида или товарооборот.

Все экономические индексы можно классифицировать по следующим признакам (рис. 12.1):

Рис. 12.1. Классификация экономических индексов

- степень охвата явления;
- база сравнения;
- вид весов (соизмерителя);
- форма построения;
- характер объекта исследования:
- объект исследования;
- состав явления;
- период исчисления.

По степени охвата явления индексы бывают индивидуальные и сводные. *Индивидуальные индексы* служат для характеристики изменения отдельных элементов сложного явления, например изменения объема производства отдельных видов продукции (телевизоров, электроэнергии и т.д.), а также цен на акции какого-либо предприятия. Для измерения динамики сложного явления, составные части которого непосредственно несоизмеримы (изменения физического объема продукции, включающей разноименные товары, индекса цен акций предприятий региона и т.п.), рассчитывают *сводные*, или *общие, индексы.*

Если индексы охватывают не все элементы сложного явления, а только часть их, то такие индексы называются *групповыми*, или *субиндексами*, например индексы физического объема продукции по отдельным отраслям промышленности, индексы цен по группам продовольственных и непродовольственных товаров. Групповые индексы отражают закономерности в развитии отдельных частей изучаемых явлений. В таких индексах проявляется их связь с методом группировок.

По базе сравнения все индексы можно разделить на две группы: *динамические* и *территориальные*. Первая группа индексов отражает изменение явления во времени. Например, индекс цен на продукцию в 2002 г. по сравнению с предыдущим годом; индекс стоимости потребительской корзины в августе по сравнению с июлем 2002 г.

При исчислении динамических индексов происходит сравнение значения показателя в отчетный период со значением этого же показателя за предыдущий период, который называют *базисным.* Однако в качестве последнего могут быть использованы и прогнозные, и плановые показатели.

Динамические индексы бывают базисными и цепными.

Вторая группа индексов (*территориальные*) применяется для межрегиональных сравнений. Большое значение эти индексы имеют в международной статистике при сопоставлении показателей социально-экономического развития различных стран. Например, индекс цен на автомобили в США по сравнению с Японией, индекс стоимости потребительской корзины в Москве по сравнению с Санкт-Петербургом.

По виду весов индексы бывают с постоянными и переменными весами.

В зависимости от формы построения различаются индексы агрегатные и средние. Последние делятся на арифметические и гармонические. Агрегатная форма общих индексов является основной формой экономических индексов. Средние индексы – производные, они получаются в результате преобразования агрегатных индексов.

По характеру объекта исследования общие индексы подразделяются на индексы количественных (объемных) и качественных показателей. В основе такого деления индексов лежит вид индексируемой величины. К первой группе индексов относятся, например, индексы объема продаж долларов США на Московской межбанковской валютной бирже, а ко второй – индекс курса немецкой марки.

По объекту исследования индексы бывают: производительности труда, себестоимости, физического объема продукции, стоимости продукции и т.д.

По составу явления можно выделить две группы индексов: постоянного (фиксированного) состава и переменного состава. Деление индексов на эти две группы используется для анализа динамики средних показателей.

По периоду исчисления индексы подразделяются на годовые, квартальные, месячные, недельные.

С помощью экономических индексов *решаются следующие задачи:*

- измерение динамики социально-экономического явления за два и более периодов времени;
- измерение динамики среднего экономического показателя;
- измерение соотношения показателей по разным регионам;
- определение степени влияния изменений значений одних показателей на динамику других;
- пересчет значения макроэкономических показателей из фактических цен в сопоставимые.

Каждая из этих задач решается с помощью различных индексов.

12.2
ИНДИВИДУАЛЬНЫЕ И ОБЩИЕ ИНДЕКСЫ

Индивидуальные индексы. Эти индексы получают в результате сравнения однотоварных явлений. Например, индекс цен на растительное масло определяется как отношение цены на этот товар в текущем периоде к цене базисного периода.

Индивидуальные индексы представляют собой относительные величины динамики, выполнения плана, сравнения, и их расчет не требует знания специальных правил.

В зависимости от экономического назначения индивидуальные индексы бывают физического объема продукции, себестоимости, цен, трудоемкости и т.д.

Индекс физического объема продукции i_q рассчитывается по формуле:

$$i_q = \frac{q_1}{q_0}. \qquad (12.1)$$

Этот индекс показывает, во сколько раз возрос (уменьшился) выпуск какого-либо одного товара в отчетном периоде по сравнению с базисным, или сколько процентов составляет рост (снижение) выпуска товара. Если из значения индекса, выраженного в процентах, вычесть 100%, то полученная величина покажет, на сколько процентов возрос (уменьшился) выпуск продукции. В знаменателе может быть не только количество продукции, произведенной за какой-то предыдущий период, но и плановое значение ($q_{\text{пл}}$), нормативное ($q_{\text{н}}$) или эталонное значение, принятое за базу сравнения ($q_{\text{э}}$). Тогда формула (12.1) примет соответственно следующий вид:

$$i_q = \frac{q_1}{q_{\text{пл}}}; \qquad (12.2)$$

$$i_q = \frac{q_1}{q_{\text{н}}}; \qquad (12.3)$$

$$i_q = \frac{q_1}{q_{\text{э}}}. \qquad (12.4)$$

Индексы других показателей строятся аналогично. Индивидуальный индекс цен:

$$i_p = \frac{p_1}{p_0}; \qquad (12.5)$$

характеризует изменение цены одного определенного товара в текущем периоде по сравнению с базисным.

Индивидуальный индекс себестоимости единицы продукции:

$$i_z = \frac{z_1}{z_0}; \qquad (12.6)$$

Он показывает изменение себестоимости единицы продукции в текущем периоде по сравнению с базисным.

Производительность труда может быть измерена количеством продукции, производимой в единицу времени (v), или затратами рабочего времени на производство единицы продукции (t). Поэтому можно построить:

- индекс количества продукции, произведенной в единицу времени:

$$i_v = \frac{v_1}{v_0} = \frac{q_1}{T_1} : \frac{q_0}{T_0}; \qquad (12.7)$$

- индекс производительности труда по трудовым затратам:

$$i_v = \frac{t_0}{t_1}. \qquad (12.8)$$

Так как между количеством продукции, произведенной в единицу времени, и затратами рабочего времени на производство единицы продукции существует обратно пропорциональная зависимость, т.е.:

$$t = \frac{1}{v}, \qquad (12.9)$$

то индекс (12.8) получится в результате деления величины показателя в базисном периоде на величину в текущем периоде.

Для характеристики производительности труда часто используется индивидуальный индекс выработки продукции в стоимостном выражении на одного рабочего:

$$i_w = \frac{w_1}{w_0} = \frac{q_1 p}{T_1} : \frac{q_0 p}{T_0}, \qquad (12.10)$$

где p – сопоставимые цены.

Индивидуальные индексы (12.7 и 12.10) показывают, во сколько раз производительность труда в базисном периоде выше (ниже), чем в отчетном.

Индекс, исчисленный по формуле (12.8), показывает, во сколько раз производительность труда в базисном периоде выше (ниже), чем в отчетном.

Индивидуальный индекс стоимости продукции отражает, во сколько раз изменилась стоимость какого-либо товара в текущем периоде по сравнению с базисным, или сколько процентов составляет рост (снижение) стоимости товара, и определяется по формуле:

$$i_{pq} = \frac{p_1 q_1}{p_0 q_0}. \qquad (12.11)$$

Индивидуальный индекс численности рабочих можно рассчитать следующим образом:

$$i_T = \frac{T_1}{T_0} = \frac{t_1 q_1}{t_0 q_0}. \qquad (12.12)$$

Он показывает, во сколько раз изменилась численность рабочих в текущем периоде по сравнению с базисным, или сколько процентов составляет рост (снижение) численности рабочих.

По данным о цене, количестве и стоимости проданных товаров (табл. 12.1) рассчитаем индивидуальные индексы. Результаты расчетов, представленные в графах 7–9 этой таблицы, показывают, что боль-

Таблица 12.1

Цена и количество проданных товаров в оном из супермаркетов г. Москвы в январе–феврале 2002 г.

Товар	Цена, руб. за 1 кг		Количество проданных товаров		Стоимость проданной продукции, тыс. руб.		Индивидуальный индекс, %			Стоимость продукции, проданной в феврале, в ценах января, тыс. руб. p_0q_1	$i_q \cdot p_0q_0$	$\frac{p_1q_1}{i_p}$	Стоимость продукции, проданной в январе, в ценах февраля, тыс. руб. p_1q_0
	январь p_0	февраль p_1	январь q_0	февраль q_1	январь p_0q_0	февраль p_1q_1	цен $i_p = \frac{p_1}{p_0}$	физического объема продукции $i_q = \frac{q_1}{q_0}$	стоимости $i_{pq} = \frac{p_1q_1}{p_0q_0}$				
А	1	2	3	4	5	6	7=2:1	8=4:3	9=6:5	10=1·4	11=8·5	12=6:7	13=2·3
Колбаса вареная, кг	95,9	117,9	1 904	2 017	182,6	237,8	122,94	105,93	130,23	193,43	193,43	193,43	244,5
Масло животное, кг	71,9	74,8	370	414	26,6	30,9	104,03	111,89	116,17	29,77	29,77	29,77	27,7
Масло растительное, л	28,8	30,9	539	566	15,5	17,5	107,29	105,01	112,90	16,30	16,30	16,30	16,7
Итого	–	–	–	–	224,7	286,2	–	–	–	239,5	239,5	239,5	288,9

ше всего возросли цены за январь–февраль 2002 г. на колбасу вареную – приблизительно на 123%. Индекс физического объема продукции в феврале по сравнению с январем 2002 г. самый высокий наблюдался по маслу животному – 111,9%. Индекс товарооборота также превысил по колбасным изделиям все рассматриваемые и составил 130,23%.

Общие индексы. В экономических расчетах чаще всего используются общие индексы, которые характеризуют изменение совокупности в целом. Построение этих индексов и является содержанием индексной методологии. В индексной теории сложились две концепции: синтетическая и аналитическая. Они по-разному интерпретируют общие индексы.

Согласно *синтетической концепции* особенность общих индексов состоит в том, что они выражают относительное изменение сложных (разнотоварных) явлений, отдельные части или элементы которых непосредственно несоизмеримы, и поэтому индексы – показатели синтетические. Например, промышленные предприятия производят несколько видов продукции, имеющей различное назначение. Следовательно, путем суммирования количества произведенных товаров различных видов нельзя получить показатель физического объема продукции. Методология построения общих индексов предусматривает прежде всего приведение разнотоварных явлений к соизмеримому виду.

В *аналитической теории* индексы трактуются как показатели, необходимые для измерения влияния изменения составных частей, компонентов, факторов сложного явления на изменение уровня этого явления. Например, изменение общей величины товарооборота в текущем периоде по сравнению с базисным связано с изменением как физического объема продаж товаров, так и цен по каждому виду товаров. Поэтому индексная методология предусматривает определение влияния каждого из факторов путем элиминирования влияния других факторов на уровень изучаемого явления.

Таким образом, общие индексы являются синтетическими и аналитическими показателями.

Общие индексы строят для количественных (объемных) и качественных показателей. В зависимости от цели исследования и наличия исходных данных используют различную форму построения общих индексов: агрегатную или средневзвешенную.

12.3 АГРЕГАТНЫЙ ИНДЕКС КАК ИСХОДНАЯ ФОРМА ИНДЕКСА

Агрегатный индекс – сложный относительный показатель, который характеризует среднее изменение социально-экономического явления, состоящего из несоизмеримых элементов.

Агрегат (лат. aggregatus) означает складываемый, суммируемый. Особенность этой формы индекса состоит в том, что в агрегатной форме непосредственно сравниваются две суммы одноименных показателей. В настоящее время это наиболее распространенная форма индексов, используемая в практической статистике многих стран мира.

Числитель и знаменатель агрегатного индекса представляют собой сумму произведений двух величин, одна из которых меняется (индексируемая величина), а другая остается неизменной в числителе и знаменателе (вес индекса).

Индексируемой величиной называется признак, изменение которого изучается (цена товаров, курс акций, затраты рабочего времени на производство продукции, количество проданных товаров и т.д.). *Вес индекса* – это величина, служащая для целей соизмерения индексируемых величин.

За каждым экономическим индексом стоят определенные экономические категории. Экономическое содержание индекса предопределяет методику его расчета.

Методика построения агрегатного индекса предусматривает ответ на три вопроса:

- какая величина будет индексируемой;
- по какому составу разнородных элементов явления необходимо исчислить индекс;
- что будет служить весом при расчете индекса.

При выборе веса индекса принято руководствоваться следующим правилом: если строится индекс количественного показателя, то веса берутся за базисный период; при построении индекса качественного показателя используются веса отчетного периода.

Построим три индекса – стоимости продукции, физического объема продукции и цен.

Стоимость продукции – это произведение количества продукции в натуральном выражении (q) на ее цену (p).

Индекс стоимости продукции, или *товарооборота* (I_{pq}), представляет собой отношение стоимости продукции текущего периода ($\Sigma p_1 q_1$) к стоимости продукции в базисном периоде ($\Sigma p_0 q_0$) и определяется по формуле:

$$I_{pq} = \frac{\sum p_1 q_1}{\sum p_0 q_0}. \tag{12.13}$$

Такой индекс показывает, во сколько раз возросла (уменьшилась) стоимость продукции (товарооборота) отчетного периода по сравнению с базисным, или сколько процентов составляет рост (снижение) стоимости продукции. Если из значения индекса стоимости (12.13) вычесть 100% (Ipq – 100), то разность покажет, на сколько процентов возросла (уменьшилась) стоимость продукции в текущем периоде по сравнению с базисным. Разность числителя и знаменателя ($\Sigma p_1 q_1 – \Sigma p_0 q_0$) показывает, на сколько рублей увеличилась (уменьшилась) стоимость продукции в текущем периоде по сравнению с базисным. Аналогично строятся индексы для показателей, которые являются произведением двух сомножителей: издержек производства (произведение себестоимости единицы продукции на количество продукции); затрат времени на производство всей продукции (произведение затрат времени на производство единицы продукции на количество выработанной продукции).

Пример. Рассчитаем индекс стоимости (товарооборота) продукции по данным табл. 12.1.

$$I_{pq} = \frac{286{,}2}{224{,}7} = 1{,}274, \text{ или } 127{,}4\%.$$

Следовательно, стоимость продукции (товарооборота) супермаркета в феврале 2002 г. по сравнению с январем возросла почти в 1,3 раза (рост составил 127,4%). Стоимость продукции увеличилась на 27,4% (127,4% – 100% = 27,4%) или на 61,5 тыс. руб. (286,2 – 224,7 = 61,5 тыс. руб.). Значение индекса стоимости продукции (товарооборота) зависит от двух факторов: изменения количества продукции и цен, что обусловливает возможность и необходимость построения еще двух индексов: физического объема продукции и цен.

Индекс физического объема продукции – это индекс количественного показателя. В этом индексе индексируемой величиной будет количество продукции в натуральном выражении, а весом – цена. Только умножив несоизмеримые между собой количества разнородной продукции на их цены, можно перейти к стоимостям продукции, которые будут уже величинами соизмеримыми. Так как индекс физического объема – индекс количественного показателя, то весами будут цены базисного периода. Тогда формула индекса примет следующий вид:

$$I_q = \frac{\sum q_1 p_0}{\sum q_0 p_0}, \qquad (12.14)$$

где в числителе дроби – условная стоимость произведенных в текущем периоде товаров в ценах базисного периода, а в знаменателе – фактическая стоимость товаров, произведенных в базисном периоде. Если объектом исследования является отдельное предприятие, то индекс определяется по совокупности произведенных товаров; когда объект исследования – отрасль промышленности, индекс рассчитывается по совокупности всех товаров, произведенных в отрасли, или отдельным их группам в зависимости от цели анализа. Если же объектом исследования является какой-либо регион, то индекс рассчитывается по товарам, произведенным предприятиями региона.

Индекс физического объема продукции (12.14) показывает, во сколько раз возросла (уменьшилась) стоимость продукции из-за роста (снижения) объема ее производства или сколько процентов составляет рост (снижение) стоимости продукции в результате изменения физического объема ее производства. Если из значения индекса физического объема продукции (12.14) вычесть 100% (I_q – 100), то разность покажет, на сколько процентов возросла (уменьшилась) стоимость продукции в текущем периоде по сравнению с базисным из-за роста (снижения) объема ее производства. Разность числителя и знаменателя ($\Sigma q_1 p_0 -$ $- \Sigma q_0 p_0$) показывает, на сколько рублей изменилась стоимость продукции в результате роста (уменьшения) ее объема. Изменение цен на продукцию в текущем периоде по сравнению с базисным не влияет на величину индекса. Для расчета индекса воспользуемся данными табл. 12.1.

$$I_q = \frac{239{,}5}{224{,}7} = 1{,}0659, \text{ или } 106{,}59\%.$$

Следовательно, стоимость продукции в феврале по сравнению с январем увеличилась в 1,06 раза (или рост стоимости составил 106,59%). Вычитая из числителя индекса знаменатель, получим:

$$239{,}5 - 224{,}7 = 14{,}8 \text{ тыс. руб.,}$$

т.е. за счет увеличения объема продукции на 6,59% (106,59% – 100% = = 6,59%) ее стоимость в абсолютном выражении увеличилась на 14,8 тыс. руб.

При построении агрегатного индекса цен, который в условиях рыночной экономики является наиболее широко распространенным показателем инфляции, исходят из тех же предпосылок, что и при построении индекса физического объема продукции.

Индекс цен – это индекс качественного показателя. Индексируемой величиной будет цена товара, так как этот индекс характеризует изменение цен. Весом будет выступать количество произведенных товаров. Умножив цену товара на его количество, получаем величину, которую можно суммировать и которая представляет собой показатель, соизмеримый с другими подобными ему величинами.

Индекс цен определяется по следующей формуле:

$$I_{\text{p}} = \frac{\sum p_1 q_1}{\sum p_0 q_1}, \tag{12.15}$$

где в числителе дроби – фактическая стоимость продукции текущего периода, а в знаменателе – условная стоимость тех же товаров в ценах базисного периода.

Индекс показывает, во сколько раз возросла (уменьшилась) стоимость продукции из-за изменения цен, или сколько процентов составляет рост (снижение) стоимости продукции в результате изменения цен. Если из значения индекса (12.15) вычесть 100% (Ip – 100%), то разность покажет, на сколько процентов возросла (уменьшилась) стоимость продукции из-за изменения цен, а разность числителя и знаменателя ($\Sigma p_1 q_1 - \Sigma p_0 q_1$) – на сколько рублей изменилась стоимость продукции в результате роста (снижения) цен. Изменение количества произведенной продукции в текущем периоде по сравнению с базисным не влияет на величину индекса.

Пример. Определим индекс цен по данным табл. 12.1:

$$I_p = \frac{286,2}{239,5} = 1,1950, \text{ или } 119,50\%.$$

Таким образом, в среднем по трем товарам цены выросли в 1,195 раза (или рост цен составил 119,5%). В результате увеличения цен на 19,5% (119,5% – 100% = 19,5%) покупатели заплатили на 46,7 тыс. руб. больше в феврале, чем в январе (286,2 – 239,5 = 46,7 тыс. руб.).

Как уже отмечалось выше, стоимость продукции можно представить как произведение количества товара на его цену. Точно такая же связь существует и между индексами стоимости, физического объема и цен, т.е.:

$$I_{pq} = I_p \cdot Iq \tag{12.16}$$

или

$$\frac{\sum p_1 q_1}{\sum p_0 q_0} = \frac{\sum p_1 q_1}{\sum p_0 q_1} \cdot \frac{\sum q_1 p_0}{\sum q_0 p_0}. \tag{12.17}$$

Разность числителя и знаменателя каждого индекса-сомножителя выражает размер изменения общей абсолютной величины под влиянием изменения одного фактора. Алгебраическая сумма этих разностей равна разности числителя и знаменателя индекса стоимости продукции:

$$\begin{gathered}\left(\sum q_1 p_0 - \sum q_0 p_0\right) + \left(\sum q_1 p_1 - \sum q_1 p_0\right) = \\ = \sum q_1 p_1 - \sum q_0 p_0.\end{gathered} \tag{12.18}$$

Равенства (12.16–12.18) выполняются в том случае, если при исчислении индекса объемного показателя веса были зафиксированы на уровне базисного периода, а при расчете индекса качественного показателя – на уровне отчетного периода.

Пример. Для нашего примера (см. табл. 12.1) индекс стоимости, исчисленный по формуле (12.16), равен:

$$1{,}0659 \cdot 1{,}1950 = 1{,}274, \text{ или } 127{,}4\%,$$

а алгебраическая сумма этих разностей, определенная по формуле (12.18), равна:

$$14{,}8 + 46{,}7 = 61{,}5 \text{ тыс. руб.}$$

Формулы для расчета общих индексов других показателей приведены в табл. 12.2.

12.4 СРЕДНИЕ ИНДЕКСЫ

Помимо агрегатных индексов в статистике применяется другая их форма – средневзвешенные индексы. К их исчислению прибегают тогда, когда имеющаяся в распоряжении информация не позволяет рассчитать общий агрегатный индекс. Так, если отсутствуют данные о ценах, но имеется информация о стоимости продукции в текущем периоде и известны индивидуальные индексы цен по каждому товару, то общий индекс цен как агрегатный определить нельзя, однако возможно исчислить его как средний из индивидуальных. Точно так же, если не известны количества произведенных отдельных видов продукции, но известны индивидуальные индексы и стоимость продукции базисного периода, то можно определить общий индекс физического объема продукции как средневзвешенную величину.

Средний индекс – это индекс, вычисленный как средняя величина из индивидуальных индексов. Агрегатный индекс является основной формой общего индекса, поэтому средний индекс должен быть тождествен агрегатному индексу. При исчислении средних индексов используются две формы средних: арифметическая и гармоническая.

Средний арифметический индекс тождествен агрегатному индексу, если весами индивидуальных индексов будут слагаемые знаменателя агрегатного индекса. Только в этом случае величина индекса, рассчитанного по формуле средней арифметической, будет равна агрегатному индексу.

Таблица 12.2

Основные формулы исчисления сводных, или общих, индексов

Наименование	Формула расчета	Что показывает индекс	Что показывает значение индекса, уменьшенное на 100%, т.е. I–100	Что показывает разность числителя и знаменателя
Индекс физического объема продукции	$I_q = \frac{\sum q_1 p_0}{\sum q_0 p_0}$	Во сколько раз изменилась стоимость продукции в результате изменения ее объема, или сколько процентов составил рост (снижение) стоимости продукции из-за изменения ее физического объема	На сколько процентов изменилась стоимость продукции в результате изменения ее объема	На сколько рублей изменилась стоимость продукции в результате роста (уменьшения) ее объема
Индекс цен	$I_p = \frac{\sum p_1 q_1}{\sum p_0 q_1}$	Во сколько раз изменилась стоимость продукции в результате изменения цен, или сколько процентов составил рост (снижение) стоимости продукции из-за изменения цен	На сколько процентов изменилась стоимость продукции в результате изменения цен	На сколько рублей изменилась стоимость продукции в результате роста (уменьшения) цен

Продолжение

Наименование	Формула расчета	Что показывает индекс	Что показывает значение индекса, уменьшенное на 100%, т.е. I–100	Что показывает разность числителя и знаменателя
Индекс стоимости продукции (товарооборота)	$I_{pq} = \frac{\sum p_1 q_1}{\sum p_0 q_0}$	Во сколько раз возросла (уменьшилась) стоимость продукции, или сколько процентов составил рост (снижение) стоимости продукции в текущем периоде по сравнению с базисным	На сколько процентов возросла (уменьшилась) стоимость продукции в текущем периоде по сравнению с базисным	На сколько рублей увеличилась (уменьшилась) стоимость продукции в текущем периоде по сравнению с базисным
Индекс физического объема	$I_q = \frac{\sum q_1 z_0}{\sum q_0 z_0}$	Во сколько раз изменились издержки производства продукции в результате изменения объема ее производства, или сколько процентов составил рост (снижение) издержек производства продукции из-за изменения физического объема ее производства	На сколько процентов изменились издержки производства продукции в результате изменения объема ее производства	На сколько рублей изменились издержки производства продукции в результате роста (уменьшения) объема ее производства

Индекс себестоимости продукции	$I_z = \frac{\sum z_1 q_1}{\sum z_0 q_1}$	Во сколько раз изменились издержки производства продукции в результате изменения себестоимости продукции, или сколько процентов составил рост (снижение) издержек производства продукции из-за изменения ее себестоимости	На сколько процентов изменились издержки производства продукции в результате изменения ее себестоимости	На сколько рублей изменились издержки производства в результате роста (уменьшения) себестоимости продукции
Индекс издержек производства	$I_{zq} = \frac{\sum z_1 q_1}{\sum z_0 q_0}$	Во сколько раз возросли (уменьшились) издержки производства продукции, или сколько процентов составил рост (снижение) издержек производства продукции в текущем периоде по сравнению с базисным	На сколько процентов возросли (уменьшились) издержки производства продукции в текущем периоде по сравнению с базисным	На сколько рублей увеличились (уменьшились) издержки производства продукции в текущем периоде по сравнению с базисным
Индекс физического объема продукции	$I_q = \frac{\sum q_1 t_0}{\sum q_0 t_0}$	Во сколько раз изменились затраты времени на производство продукции в результате изменения объема ее производства, или сколько процентов составил рост (снижение) затрат времени на производство продукции из-за изменения физического объема ее производства	На сколько процентов изменились затраты времени на производство продукции в результате изменения объема ее производства	На сколько человеко-часов возросли (уменьшились) затраты времени на производство продукции в результате роста (уменьшения) объема производства продукции

Продолжение

Наименование	Формула расчета	Что показывает индекс	Что показывает значение индекса, уменьшенное на 100%, т.е. I–100	Что показывает разность числителя и знаменателя
Индекс производительности труда по трудовым затратам	$I_t = \frac{\sum t_0 q_1}{\sum t_1 q_1}$	Во сколько раз увеличилась (уменьшилась) производительность труда, или сколько процентов составило снижение (рост) производительности труда в текущем периоде по сравнению с базисным	На сколько процентов изменилась производительность труда в текущем периоде по сравнению с базисным	Абсолютный размер экономии (перерасхода) затрат живого труда в связи с ростом (уменьшением) его производительности
Индекс затрат времени на производство продукции	$I_{tq} = \frac{\sum t_1 q_1}{\sum t_0 q_0}$	Во сколько раз изменились затраты времени на производство продукции, или сколько процентов составил рост (снижение) затрат времени на производство продукции в текущем периоде по сравнению с базисным	На сколько процентов изменились затраты времени на производство продукции в текущем периоде по сравнению с базисным	На сколько человеко-часов увеличились (уменьшились) затраты на производство продукции в текущем периоде по сравнению с базисным

Средний арифметический индекс физического объема продукции вычисляется по формуле:

$$I_q = \frac{\sum i_q p_0 q_0}{\sum p_0 q_0}. \qquad (12.19)$$

Так как $i_q \cdot q_0 = q_1$, то формула этого индекса легко преобразуется в формулу (12.14). Весами в формуле (12.19) является стоимость продукции базисного периода.

Средний арифметический индекс производительности труда определяется следующим образом:

$$I_t = \frac{\sum i_t t_1 q_1}{\sum t_1 q_1} = \frac{\sum i_t T_1}{\sum T_1}. \qquad (12.20)$$

Так как $i_t \cdot t_1 = t_0$, то формула этого индекса может быть преобразована в агрегатный индекс трудоемкости продукции. Весами являются общие затраты времени на производство продукции в текущем периоде.

В статистике широко известен и другой средний арифметический индекс, который используется при анализе производительности труда. Он носит название *индекса Струмилина* и определяется следующим образом:

$$I_v = \frac{\sum \left(\frac{q_1}{T_1} : \frac{q_0}{T_0} \right) \cdot T_1}{\sum T_1}. \qquad (12.21)$$

Индекс показывает, во сколько раз возросла (уменьшилась) производительность труда, или сколько процентов составил рост (снижение) производительности труда в среднем по всем единицам исследуемой совокупности.

Средние арифметические индексы чаще всего применяются на практике для расчета сводных индексов количественных показателей. При анализе качественных показателей данная форма индекса применяется для исчисления приведенных выше индексов (формулы (12.20) – (12.21)).

Индексы других качественных показателей (цен, себестоимости и т.д.) определяются по формуле средней гармонической взвешенной величины.

Средний гармонический индекс тождествен агрегатному, если индивидуальные индексы взвешены с помощью слагаемых числителя агрегатного индекса.

Например, индекс себестоимости можно исчислить так:

$$I_z = \frac{\sum z_1 q_1}{\sum \frac{z_1 q_1}{i_z}}, \tag{12.22}$$

а индекс цен:

$$I_p = \frac{\sum p_1 q_1}{\sum \frac{p_1 q_1}{i_p}}. \tag{12.23}$$

Таким образом, при определении среднего гармонического индекса себестоимости весами являются издержки производства текущего периода, а при вычислении индекса цен веса – стоимость продукции этого периода.

Рассчитаем средние индексы цен и физического объема продукции по данным табл. 12.1 (графы 11–12):

$$I_q = \frac{239,5}{224,7} = 1,0659, \text{ или } 106,59\%;$$

$$I_p = \frac{286,2}{239,5} = 1,1950, \text{ или } 119,50\%.$$

Этот же результат получился при расчете агрегатных индексов. Средние индексы широко используются для анализа рынка ценных бумаг. Наиболее известными являются индексы Доу-Джонса, Стэндарда и Пура.

Индекс Доу-Джонса (Dow Jones Industrial Average Index) определяется как средний арифметический индекс значений курсов акций, котирующихся на Нью-Йоркской фондовой бирже. Один сводный и три групповых индекса рассчитываются каждые полчаса, и ежедневно публикуется их значение на момент закрытия биржи. Групповые индексы определяются по ценам акций 30 промышленных, 20 транспортных и 15 компаний сферы услуг. Общий индекс рассчитывается по всем 65 компаниям. Их перечень был составлен в 1928 г. В качестве базисного выбран 1920 г. Первоначальная методика исчисления индекса была разработана основателем и редактором крупнейшей в США газеты «Уолл-стрит джорнел» Чарлзом Доу.

Индекс Стэндарда и Пура (Standard and Poor's 500 Stock Index) – индекс, рассчитываемый по курсам акций 500 крупнейших компаний Нью-Йоркской фондовой биржи как средний взвешенный показатель, учитывающий общее число выпущенных компанией акций. В число компаний, акции которых включены в индекс, входят 400 промышленных корпораций, 40 – финансовых, 20 – транспортных и 40 – сферы услуг.

12.5 ВЫБОР БАЗЫ И ВЕСОВ ИНДЕКСОВ

Выбор базы сравнения и весов индексов – это два важнейших методологических вопроса построения систем индексов. Система используется при изучении динамики социально-экономических явлений за некоторый интервал времени, включающий более двух периодов времени.

Системой индексов называется ряд последовательно построенных индексов. Такие системы характеризуют изменения, происходящие в изучаемом явлении в течение исследуемого периода времени.

В зависимости от базы сравнения системы индексов бывают базисными и цепными (рис. 12.2).

Система базисных индексов – это ряд последовательно вычисленных индексов одного и того же явления с постоянной базой сравнения, т.е. в знаменателе всех индексов находится индексируемая величина базисного периода.

Система цепных индексов – это ряд индексов одного и того же явления, вычисленных с меняющейся от индекса к индексу базой сравнения.

Рис. 12.2. Система индексов

В экономико-статистических исследованиях выбор системы индексов (базисные или цепные) проводится в зависимости от цели анализа. Базисные индексы дают более наглядную характеристику общей тенденции развития исследуемого явления, а цепные – четче отражают последовательность изменения уровней во времени.

Системы цепных и базисных индексов могут быть построены для индивидуальных и общих индексов. Системы индивидуальных индексов стоимости продукции, физического объема продукции и цен (табл. 12.3) просты по построению. Аналогично им строятся системы индивидуальных индексов и для других показателей.

Между цепными и базисными индексами существуют различные виды связи.

Если известны цепные индексы, то путем их последовательного перемножения можно получить базисные индексы. Например,

Таблица 12.3

Системы индивидуальных индексов

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

$$\frac{p_1}{p_0}\cdot\frac{p_2}{p_1}=\frac{p_2}{p_0} \qquad (12.24)$$

или

$$\frac{q_1}{q_0}\cdot\frac{q_2}{q_1}\cdot\frac{q_3}{q_2}=\frac{q_3}{q_0}. \qquad (12.25)$$

Зная последовательные значения базисных индексов, легко рассчитать на их основе цепные индексы:

$$\frac{p_2}{p_1}=\frac{p_2}{p_0}:\frac{p_1}{p_0} \qquad (12.26)$$

или

$$\frac{q_3}{q_2}=\frac{q_3}{q_0}:\frac{q_2}{q_0}. \qquad (12.27)$$

Системы базисных и цепных индексов могут быть построены для агрегатных индексов.

Система индексов стоимости имеет следующий вид:
цепные индексы:

$$\frac{\sum p_1q_1}{\sum p_0q_0};\ \frac{\sum p_2q_2}{\sum p_1q_1};...;\frac{\sum p_nq_n}{\sum p_{n-1}q_{n-1}};$$

базисные индексы:

$$\frac{\sum p_1q_1}{\sum p_0q_0};\ \frac{\sum p_2q_2}{\sum p_0q_0};...;\frac{\sum p_nq_n}{\sum p_0q_0}.$$

Формирование системы индексов, например цен или физического объема, отличается от уже рассмотренных в этом разделе систем индексов. Это связано с тем, что при построении систем этих индексов можно использовать постоянные и переменные веса.

Системой индексов с постоянными весами называется система сводных индексов одного и того же явления, вычисленных с весами, не меняющимися при переходе от одного индекса к другому. Постоянные веса позволяют исключить влияние изменения структуры на величину индекса. Например, система базисных индексов физического объема продукции с постоянными весами (p_0) имеет следующий вид (см. рис. 12.1):

$$\frac{\sum q_1p_0}{\sum q_0p_0};\ \frac{\sum q_2p_0}{\sum q_0p_0};...;\frac{\sum q_np_0}{\sum q_0p_0},$$

а систему цепных индексов с теми же постоянными весами можно представить так:

$$\frac{\sum q_1p_0}{\sum q_0p_0};\ \frac{\sum q_2p_0}{\sum q_1p_0};...;\frac{\sum q_np_0}{\sum q_{n-1}p_0}.$$

Система индексов с переменными весами представляет собой систему сводных индексов одного и того же явления, вычисленных с весами, последовательно меняющимися от одного индекса к другому (см. рис. 12.2). Переменные веса – это веса отчетного периода. Например, система базисных индексов цен с переменными весами следующая:

$$\frac{\sum p_1q_1}{\sum p_0q_1};\ \frac{\sum p_2q_2}{\sum p_0q_2};\ldots;\frac{\sum p_nq_n}{\sum p_0q_n}.$$

Элементами этой системы являются индексы-дефляторы, которые необходимы для пересчета стоимостных показателей системы национальных счетов в сопоставимые цены.

Система цепных индексов цен с переменными весами выглядит так:

$$\frac{\sum p_1q_1}{\sum p_0q_1};\ \frac{\sum p_2q_2}{\sum p_1q_2};\ldots;\frac{\sum p_nq_n}{\sum p_{n-1}q_n}.$$

Отдельные индексы этой системы используются для пересчета стоимостных показателей отчетного периода в цены предыдущего периода. Системы общих индексов других показателей строятся аналогично.

Системы агрегатных индексов обладают теми же свойствами, что и системы индивидуальных индексов, т.е. зная базисные индексы, можно рассчитать цепные; при наличии цепных индексов легко получить соответствующие им базисные. Например,

$$\frac{\sum p_1q_1}{\sum p_0q_0}\cdot\frac{\sum p_2q_2}{\sum p_1q_1}=\frac{\sum p_2q_2}{\sum p_0q_0} \tag{12.28}$$

или

$$\frac{\sum q_1p_0}{\sum q_0p_0}\cdot\frac{\sum q_2p_0}{\sum q_1p_0}\cdot\frac{\sum q_3p_0}{\sum q_2p_0}=\frac{\sum q_3p_0}{\sum q_0p_0} \tag{12.29}$$

или

$$\frac{\sum p_2q_2}{\sum p_0q_0}:\frac{\sum p_1q_1}{\sum p_0q_0}=\frac{\sum p_2q_2}{\sum p_1q_1} \tag{12.30}$$

или

$$\frac{\sum p_nq_n}{\sum p_0q_n}:\frac{\sum p_{n-1}q_n}{\sum p_0q_n}=\frac{\sum p_nq_n}{\sum p_{n-1}q_n}. \tag{12.31}$$

12.6
ИНДЕКСЫ СТРУКТУРНЫХ СДВИГОВ

При изучении динамики качественных показателей приходится определять изменение средней величины индексируемого показателя, которое обусловлено взаимодействием двух факторов – изменением значения индексируемого показателя у отдельных групп единиц и изменением структуры явления. Под *изменением структуры явления* понимается изменение доли отдельных групп единиц совокупности в общей их численности. Так, средняя заработная плата на предприятии может вырасти в результате роста оплаты труда работников или увеличения доли высокооплачиваемых сотрудников. Снижение трудоемкости производства единицы продукции по совокупности предприятий отрасли может быть обусловлено повышением производительности труда на предприятиях или концентрацией производства продукции на заводах с низкой трудоемкостью. Так как на изменение среднего значения показателя оказывают воздействие два фактора, возникает задача определить степень влияния каждого из факторов на общую динамику средней.

Эта задача решается с помощью индексного метода, т.е. путем построения системы взаимосвязанных индексов, в которую включаются три индекса: переменного состава, постоянного состава и структурных сдвигов.

Индексом переменного состава называется индекс, выражающий соотношение средних уровней изучаемого явления, относящихся к разным периодам времени. Например, индекс переменного состава себестоимости продукции одного и того же вида рассчитывается по формуле:

$$I_{\text{пс}} = \frac{\overline{Z}_1}{\overline{Z}_0} = \frac{\sum z_1 q_1}{\sum q_1} : \frac{\sum z_0 q_0}{\sum q_0}, \qquad (12.32)$$

где $I_{\text{пс}}$ – индекс переменного состава.

Индекс переменного состава отражает изменение не только индексируемой величины (в данном случае себестоимости), но и структуры совокупности (весов).

Индекс постоянного (фиксированного) состава – это индекс, исчисленный с весами, зафиксированными на уровне одного какого-либо периода, и показывающий изменение только индексируемой величины.

Индекс фиксированного состава определяется как агрегатный индекс. Так, индекс фиксированного состава себестоимости продукции рассчитывают по формуле:

$$I_{\text{фс}} = \frac{\sum z_1 q_1}{\sum q_1} : \frac{\sum z_0 q_1}{\sum q_1} = \frac{\sum z_1 q_1}{\sum z_0 q_1}, \qquad (12.33)$$

где $I_{\text{фс}}$ – индекс фиксированного состава.

Под *индексом структурных сдвигов* понимают индекс, характеризующий влияние изменения структуры изучаемого явления на динамику среднего уровня этого явления. Индекс определяется по формуле (при изучении изменения среднего уровня себестоимости):

$$I_{\text{сс}} = \frac{\sum z_0 q_1}{\sum q_1} : \frac{\sum z_0 q_0}{\sum q_0} = \frac{\sum z_0 q_1}{\sum z_0 q_0} : \frac{\sum q_1}{\sum q_0}, \qquad (12.34)$$

где $I_{\text{сс}}$ – индекс структурных сдвигов.

Система взаимосвязанных индексов при анализе динамики средней себестоимости имеет следующий вид:

$I_{\text{пс}}$	=	$I_{\text{фс}}$	·	$I_{\text{сс}}$	
Индекс переменного состава		Индекс фиксированного состава		Индекс структурных сдвигов	(12.35)

Рассмотрим применение такой системы.

Пример. Пусть имеются данные об объеме строительства коттеджей и себестоимости 1 кв. м жилья двух строительных фирм в мае–июне 2002 г. (табл. 12.4).

В июне по сравнению с маем 2002 г. себестоимость 1 кв. м коттеджного жилья в двух строительных фирмах возросла: в ООО «Скат» она увеличилась на 5,26%, а в ООО «Стройинвест» – на 16,67%. При этом изменилась структура рынка участников строительства: уменьшилась

Таблица 12.4

Построено коттеджного жилья и себестоимость 1 кв. м по двум строительным фирмам в мае–июне 2002 г.*

Строительная фирма	Построено коттеджного жилья				Себестоимость единицы продукции, долл. США		Индивидуальные индексы себестоимости, % $i_z = z_1 : z_0$	Общие затраты на производство жилья, тыс. долл. США		
	Всего, кв. м		% к итогу		базисный период (май) z_0	текущий период (июнь) z_1		Базисный период (май) z_0q_0	текущий период (июнь) z_1q_1	z_0q_1
	базисный период (май) q_0	текущий период (июнь) q_1	базисный период (май)	текущий период (июнь)						
А	1	2	3	4	5	6	7 = 6 : 5	8 = 5 · 1	9 = 6 · 2	10 = 5 · 2
ООО «Скат»	1620	1780	35,1	28,5	285	300	105,26	461,7	534,0	507,3
ООО «Стройинвест»	3000	4470	64,9	71,5	300	350	116,67	900,0	1564,5	1341,0
Итого	4620	6250	100,0	100,0	–	–	–	1361,7	2098,5	1848,3

* Цифры условные.

доля первой строительной фирмы (ООО «Скат») в общем объеме строительства коттеджей (с 35,1 до 28,5%) и возросла доля второй фирмы (ООО «Стройинвест») – с 64,9 до 71,5%).

Рассчитаем индекс переменного состава. Для этого сначала определим среднюю себестоимость единицы продукции в текущем и базисном периодах:

$$\bar{z}_0 = \frac{1361,7}{4620} = 0,2947 \text{ тыс. долл. США, или } 294,7 \text{ долл. США.}$$

$$\bar{z}_1 = \frac{2098,5}{6250} = 0,3358 \text{ тыс. долл. США, или } 335,8 \text{ долл. США.}$$

Тогда

$$I_{\text{п.с}} = 335,8 : 294,7 = 1,1395, \text{ или } 113,95\%.$$

Следовательно, средняя себестоимость по двум фирмам возросла в текущем периоде по сравнению с базисным на 13,95%, и в каждом из них в отдельности она возрастала. Это результат того, что исчисленный индекс учитывает влияние еще и структурного фактора.

Определим индекс себестоимости фиксированного состава:

$$I_{\text{фс}} = \frac{2098,5}{1848,3} = 1,135, \text{ или } 113,5\%.$$

Таким образом, себестоимость в текущем периоде по сравнению с базисным возросла в среднем на 13,5%.

Вычислим влияние изменения структуры на динамику средней себестоимости:

$$I_{\text{сс}} = \frac{1848,3}{1361,7} : \frac{6250}{4620} = 1,0034, \text{ или } 100,34\%.$$

Изменение доли строительных фирм в общем объеме построенного коттеджного жилья привело к увеличению себестоимости на 0,34%.

Аналогично строятся системы индексов для других показателей. Так, для показателя производительности труда можно построить систему индексов, в которой:

$$I_{\text{пс}} = \bar{t}_0 : \bar{t}_1 = \frac{\sum T_0}{\sum q_0} : \frac{\sum T_1}{\sum q_1}; \tag{12.36}$$

$$I_{\text{фс}} = \frac{\sum t_0 q_1}{\sum t_1 q_1}; \tag{12.37}$$

$$I_{cc} = \left[\frac{\sum T_0}{\sum q_0} : \frac{\sum T_1}{\sum q_1}\right] : \frac{\sum t_0 q_1}{\sum t_1 q_1} = \frac{\sum t_0 q_0}{\sum t_0 q_1} : \frac{\sum q_0}{\sum q_1}. \tag{12.38}$$

12.7 ИНДЕКСЫ ПРОСТРАНСТВЕННО-ТЕРРИТОРИАЛЬНОГО СОПОСТАВЛЕНИЯ

В статистической практике часто возникает потребность в сопоставлении уровней экономического явления в пространстве: по странам, экономическим районам, областям, т.е. в исчислении территориальных индексов. При построении территориальных индексов приходится решать вопрос, какие веса использовать при их исчислении. Например, если стоит задача сравнить цены двух регионов (*А* и *Б*), то можно построить два индекса:

$$I_{A/Б} = \frac{\sum p_{\text{А}} q_A}{\sum p_Б q_A} \tag{12.39}$$

и

$$I_{Б/A} = \frac{\sum p_Б q_Б}{\sum p_A q_Б}, \tag{12.40}$$

где $I_{А/Б}$ – индекс, в котором в качестве базы сравнения применяются данные по региону *А*;

$I_{Б/А}$ – индекс, используемый в качестве базы сравнения данных по региону *Б*.

Эти формулы могут дать совершенно различное представление о соотношении уровней явления. Например, при расчете по формуле (12.39) значение признака будет ниже в регионе *А*, а по формуле (12.40) – в регионе *Б*.

Пример. Рассмотрим табл. 12.5.

Рассчитав индексы по двум формулам, получаем:

$$I_{A/Б} = \frac{\sum p_A q_A}{\sum p_Б q_A} = \frac{236400}{243000} = 0{,}973, \text{ или } 97{,}3\%$$

и

$$I_{Б/A} = \frac{\sum p_Б q_Б}{\sum p_A q_Б} = \frac{186000}{180800} = 1{,}029, \text{ или } 102{,}9\%.$$

Индексы показывают, что при сравнении региона *А* с регионом *Б* цены ниже в регионе *А* на 2,7%, а при сравнении региона *Б* с регионом *А* оказывается, что цены несколько выше в регионе *Б*. Таким образом, расчет индексов не позволяет определить, в каком регионе выше цены. Причина заключается в резком различии структуры продаж в отдельных регионах.

В теории и практике статистики предлагаются различные методы построения территориальных индексов, в том числе *метод стандартных весов.* Этот метод заключается в том, что значения индексируемой величины взвешиваются не по весам какого-то одного региона, а по весам области, экономического района, республики, в которых находятся сравниваемые регионы.

В нашем примере в качестве весов можно использовать количество продукции, проданной в регионах *А* и *Б*, т.е.:

$$I_p = \frac{\sum p_A (q_A + q_Б)}{\sum p_Б (q_A + q_Б)}. \qquad (12.41)$$

Определим значение индекса по данным табл. 12.5:

$$I_p = \frac{417200}{429000} = 0{,}9725, \text{ или } 97{,}25\%.$$

Итак, цены в регионе *А* ниже, чем цены в регионе *Б*, в среднем на 2,75%.

Таблица 12.5

Цена на продукты питания и количество проданной продукции по двум регионам*

Наименование продукта	Регион *А*		Регион *Б*		Товарооборот		$p_Б q_А$	$p_А q_Б$	Количество продукции, проданной в регионах *А* и *Б* $q_А + q_Б$	$p_А(q_А + q_Б)$	$p_Б(q_А + q_Б)$
	Цена, руб., $p_А$	Количество проданного товара, $q_А$	Цена, руб., $p_Б$	Количество проданного товара, $q_Б$	Регион *А* $p_А q_А$	Регион *Б* $p_Б q_Б$					
А	1	2	3	4	5 = 1 · 2	6 = 3 · 4	7 = 3 · 2	8 = 1 · 4	9 = 2 + 4	10 = 1 · 9	11 = 3 · 9
Мясо и птица, кг	95	1700	100	1200	161500	120000	170000	114000	2900	275500	290000
Масло животное, кг	83	800	80	600	66400	48000	64000	49800	1400	116200	112000
Яйца, десяток	17	500	18	1000	8500	18000	9000	17000	1500	25500	27000
Итого	–	3000	–	2800	236400	186000	243000	180800	–	417200	429000

* Цифры условные.

12.8 ВАЖНЕЙШИЕ ЭКОНОМИЧЕСКИЕ ИНДЕКСЫ И ИХ ВЗАИМОСВЯЗИ

Между важнейшими индексами существуют взаимосвязи, позволяющие на основе одних индексов получить другие. Зная, например, значение цепных индексов за какой-либо период времени, можно рассчитать базисные индексы. И наоборот, если известны базисные индексы, то путем деления одного из них на другой можно получить цепные индексы.

Существующие взаимосвязи между важнейшими индексами позволяют выявить влияние различных факторов на изменение изучаемого явления, например связь между индексом стоимости продукции, физического объема продукции и цен (раздел 12.7). Другие индексы также связаны между собой. Так, индекс издержек производства – это произведение индекса себестоимости продукции и индекса физического объема продукции:

$$I_{zq} = I_z \cdot I_q \tag{12.42}$$

или

$$\frac{\sum z_1 q_1}{\sum z_0 q_0} = \frac{\sum z_1 q_1}{\sum z_0 q_1} \cdot \frac{\sum q_1 z_0}{\sum q_0 z_0}. \tag{12.43}$$

Отсюда если себестоимость увеличилась на 10%, а количество продукции снизилось на 8%, то индекс издержек на производство будет равен:

$$1{,}10 \cdot 0{,}92 = 1{,}012, \text{ или } 101{,}2\%.$$

Индекс затрат времени на производство продукции может быть получен в результате умножения индекса физического объема продукции и величины, обратной величине индекса трудоемкости, т.е. индекс производительности труда

$$I_{tq} = I_q \cdot \frac{1}{I_t} \tag{12.44}$$

или

$$\frac{\sum t_1 q_1}{\sum t_0 q_0} = \frac{\sum q_1 t_0}{\sum q_0 t_0} : \frac{\sum t_0 q_1}{\sum t_1 q_1}. \qquad (12.45)$$

При увеличении физического объема продукции в текущем периоде на 15% по сравнению с базисным производительность снизилась на 18%, поэтому индекс затрат времени на производство продукции будет равен:

1,15 : 0,82 = 1,402, или 140,2%.

Существует важная взаимосвязь между индексами физического объема продукции и индексами производительности труда.

Индекс производительности труда рассчитывается на основе следующей формулы:

$$I_w = \frac{\sum q_1 p_0}{\sum T_1} : \frac{\sum q_0 p_0}{\sum T_0}, \qquad (12.46)$$

т.е. представляет собой отношение средней выработки продукции (в сопоставимых ценах) в единицу времени (или на одного занятого) в текущем и базисном периодах.

Индекс физического объема продукции равен произведению индекса производительности труда на индекс затрат рабочего времени (или численности занятых):

$$\frac{\sum q_1 p_0}{\sum q_0 p_0} = \frac{\sum T_1}{\sum T_0} \cdot \left(\frac{\sum q_1 p_0}{\sum T_1} : \frac{\sum q_0 p_0}{\sum T_0} \right). \qquad (12.47)$$

Таким образом, если численность рабочих возросла на 12%, а производительность труда – на 7%, то индекс физического объема продукции будет равен:

1,12 · 1,07 = 1,20, или 120%.

Взаимосвязь между отдельными индексами может быть использована для выявления влияния отдельных факторов, оказывающих воздействие на изучаемое явление.

Пример. Пусть имеются следующие данные (табл. 12.6).

Таблица 12.6

Динамика отгруженной продукции, численности рабочих и производительности труда на предприятии в мае 2002 г.*

Наименование показателя	Значение показателя в периоде		Абсолютное изменение	Относительная величина динамики, %
	базисном	текущем		
А	1	2	3 = 2 – 1	4 = 2 : 1 · 100
Стоимость отгруженной продукции, млн руб.	450	610	+160	136
Численность рабочих, чел.	700	730	+30	104,3
Выработка продукции на одного рабочего, тыс. руб.	642,3	835,6	+193,3	130,1

* Цифры условные.

Из табл. 12.6 следует, что численность рабочих увеличилась на 4,3% (730 : 700 = 1,043, или 104,3%); производительность труда – на 30,1% (835,6 : 642,3 = 1,301, или 130,1%), а отгруженная продукция – на 36% (610 : 450 = 1,36, или 1,043 · 1,301 = 1,36, т. е. 136%).

Объем отгруженной продукции в текущем периоде по сравнению с базисным возрос на 160 млн руб. Это изменение является результатом роста численности рабочих и выработки продукции.

Для того чтобы определить влияние каждого из факторов, введем следующие обозначения. Обозначим численность рабочих через «a», а другой фактор – выработку продукции на одного рабочего – через «b». Тогда стоимость отгруженной продукции – это $a \cdot b$, а индекс стоимости продукции – это произведение индекса численности занятых на индекс производительности труда:

$$\frac{a_1 b_1}{a_0 b_0} = \frac{a_1}{a_0} \cdot \frac{b_1}{b_0}. \tag{12.48}$$

Чтобы выявить влияние каждого фактора на общее изменение стоимости отгруженной продукции, необходимо определить динамику одного из факторов, оставив значение другого неизменным. Тогда индекс (12.48) можно представить в двух вариантах:

$$1.\ \frac{a_1 b_1}{a_0 b_0} = \frac{a_1 b_0}{a_0 b_0} \cdot \frac{b_1 a_1}{b_0 a_1}; \tag{12.49}$$

$$2.\ \frac{a_1 b_1}{a_0 b_0} = \frac{a_1 b_1}{a_0 b_1} \cdot \frac{b_1 a_0}{b_0 a_0}. \tag{12.50}$$

Каждый из индексов-сомножителей (12.49 – 12.50) отражает влияние одного фактора на изменение стоимости отгруженной продукции при неизменном (базисном или отчетном) уровне другого фактора. Только в 1-м варианте изменение выработки продукции (b) определяется при численности занятых в текущем периоде, а во 2-м варианте – в базисном периоде. И наоборот, изменение численности занятых (a) в 1-м варианте определяется при базисном значении выработки продукции, а во 2-м варианте – при текущем его значении.

Рассчитаем влияние каждого из факторов по формуле (12.49):

$$\frac{610}{450} = \frac{730 \cdot 642{,}3}{700 \cdot 642{,}3} \cdot \frac{835{,}6 \cdot 730}{642{,}3 \cdot 730} = \frac{469}{450} \cdot \frac{610}{469}.$$

Следовательно,

1. Общий прирост продукции равен: 160 млн руб. [$a_1 b_1 - a_0 b_0 = 610 - 450$].

2. За счет увеличения численности рабочих стоимость продукции возросла на 19 млн руб. [$a_1 b_0 - a_0 b_0 = 469 - 450$].

3. За счет повышения производительности труда стоимость продукции увеличилась на 141 млн руб. [$a_1 b_1 - a_1 b_0 = 610 - 469$].

Теперь определим влияние каждого из факторов по формуле (12.50):

$$\frac{610}{450} = \frac{730 \cdot 835{,}6}{700 \cdot 835{,}6} \cdot \frac{700 \cdot 835{,}6}{700 \cdot 642{,}3} = \frac{610}{585} \cdot \frac{585}{450}.$$

Таким образом, общий прирост продукции равен 160 млн руб. [$a_1b_1 - a_0b_0 = 610 - 450$]. При этом за счет увеличения численности рабочих стоимость продукции возросла на 25 млн руб. [$a_1b_1 - a_0b_1 = 610 - 585$]. За счет повышения производительности труда стоимость продукции увеличилась на 135 млн руб. [$a_0b_1 - a_0b_0 = 610 - 450$].

Итак, мы получили различные результаты. Разницу в значениях отклонений, равную в нашем примере 6 млн руб. (141 – 135, или 25 – 19), называют изменением стоимости отгруженной продукции за счет совместного изменения обоих факторов, которая присоединяется то к одному, то к другому фактору. В экономических расчетах чаще применяется первый способ анализа влияния факторов (12.49), так как в этом случае индекс качественного признака определяется при неизменном значении количественного признака, зафиксированном на уровне текущего периода.

12.9 СВОЙСТВА ИНДЕКСОВ ЛАСПЕЙРЕСА И ПААШЕ

В рыночном хозяйстве особое место среди индексов качественных показателей отводится индексам цен. Основным назначением *индекса цен* является оценка динамики цен на товары производственного и непроизводственного потребления. Помимо этого, индекс цен выполняет роль общего измерителя инфляции при макроэкономических исследованиях; используется при корректировке законодательно устанавливаемого минимального размера оплаты труда, установлении ставок налогов.

Индексы цен нужны при разработке технико-экономических обоснований и проектов строительства новых предприятий. Без них нельзя обойтись при пересчете основных показателей системы национальных счетов (совокупного общественного продукта, национального дохода, капитальных вложений и т.д.) из фактически действовавших (текущих) цен в сопоставимые.

Таким образом, *индексы цен необходимы для решения двух задач:*

- отражения динамики инфляционных процессов в народном хозяйстве страны;
- пересчета важнейших стоимостных показателей СНС из фактических цен в сопоставимые при изучении динамики социально-экономических явлений.

Для реализации этих различных по содержанию задач служат два *типа индексов:*

- собственно индекс цен;
- индекс-дефлятор.

Рассмотрим подробнее первый индекс. Первая формула для расчета индекса цен была сформулирована в 1738 г. французским экономистом Дюто, предложившим вычислять обобщенный показатель изменения цен как отношение сумм цен на отдельные виды товаров в отчетном периоде к сумме цен на те же товары в базисном периоде. Эта формула имеет следующий вид:

$$I_p = \frac{\sum p_1}{\sum p_0}. \qquad (12.51)$$

В 1764 г. итальянец Карли предложил определять общий индекс цен как простую среднюю арифметическую величину из индивидуальных индексов цен:

$$I_p = \frac{\sum \frac{p_1}{p_0}}{n} = \frac{\sum i_p}{n}. \qquad (12.52)$$

И только в конце XIX в. были построены две формулы индекса цен, которые используются в качестве основных современной отечественной и зарубежной статистикой.

Автором первой формулы (12.15) является немецкий статистик Г. Пааше. Немецкий ученый Э. Ласпейрес предложил определять индекс цен следующим образом:

$$I_p = \frac{\sum p_1 q_0}{\sum p_0 q_0}. \qquad (12.53)$$

Индексируемой величиной обоих индексов являются цены. Весами же в индексе цен Пааше выступает количество продукции текущего периода, а в индексе цен Ласпейреса – количество продукции базисного периода.

Найдем значение индекса цен Ласпейреса по данным табл. 12.1. Он равен:

$$I_p = \frac{288,9}{224,7} = 1,2857, \text{ или } 128,57\%.$$

Следовательно, в среднем по всем товарам цены возросли на 28,57%. В результате роста цен стоимость товаров базисного периода увеличилась на 64,2 тыс. руб. (288,9 – 224,7 = 64,2 тыс. руб.).

Значения индексов цен Пааше и Ласпейреса не совпадают. Отличие значений объясняется тем, что индексы имеют различное экономическое содержание.

Индекс цен, исчисленный по формуле Пааше, дает ответ на вопрос, насколько товары в текущем периоде стали дороже (дешевле), чем в базисном. *Индекс цен Ласпейреса* показывает, во сколько бы раз товары базисного периода подорожали (подешевели) из-за изменения цен на них в отчетный период.

Согласно практике индекс цен, рассчитанный по формуле Пааше, имеет тенденцию некоторого занижения, а по формуле Ласпейреса – завышения темпов инфляции.

До начала 90-х гг. XX в. отечественная статистика отдавала предпочтение индексу цен Пааше. Сложность его расчета заключается в том, что взвешивание по весам отчетного периода требует ежегодного (ежеквартального, ежемесячного) сбора и обработки значительных объемов информации для формирования системы весов. А эта работа связана с большими затратами времени, материальных и трудовых ресурсов. Поэтому начиная с 1991 г. отечественные органы государственной статистики определяют индексы цен по формуле Ласпейреса, которой отдается предпочтение и в зарубежной статистике: Англии, Германии, США и др. При исчислении индекса цен по формуле Ласпейреса веса фиксируются на уровне базисного периода и остаются неизменными в течение некоторого промежутка времени, отсюда целью расчета индекса является измерение динамики стоимости базисного (неизменного) объема продукции.

Следует отметить, что индекс цен всегда имеет определенную степень условности. Это связано, прежде всего, с тем, что при его расчете учитываются изменения цен не по всей совокупности продукции, а по отдельным товарам-представителям, которые составляют так называемую товарную корзину. По мере отдаления от базисного года эта товарная корзина по видам, количеству и качеству вошедших в нее товаров-представителей все менее соответствует структуре и составу объема продукции текущего года. Поэтому состав товарной корзины, а следовательно, и система весов должны периодически пересматриваться. Только тогда они отражают современную структуру объема продукции. Особенно это важно в период резкого изменения экономических условий в народном хозяйстве страны.

При расчете индекса цен по формуле Ласпейреса необходимо *решить три вопроса*:

- выбор базисного года для постоянных весов;
- определение срока использования весовых коэффициентов без их пересмотра;
- увязку индекса, рассчитанного по новым весам (после их пересмотра), с ранее существующими динамическими рядами индексов цен.

Например, при выборе базисного года немецкие статистики ориентируются на *следующие критерии*:

- базисный год должен находиться в середине длительной фазы подъема (снижения) экономического развития;
- динамика цен в базисном году не должна быть ниже, чем в соседние с ним годы, но не должна быть и стабильной;
- год должен быть сравнительно «нормальным» для сельскохозяйственного производства, т.е. не выделяться по погодным условиям среди других лет.

В странах Европейского Союза принято пересматривать весовые коэффициенты по истечении пятилетнего срока. В последнее время в связи с существенными колебаниями цен, что ведет к изменению структуры объема продукции, во многих странах системы весов пересматривают ежегодно, т.е. используют подвижную систему весов. Эта система остается неизменной в течение года, и по ней ежемесячно или ежеквартально рассчитываются индексы цен. Когда наступает новый календарный год, веса корректируют. Именно этот подход используется в настоящее время при исчислении индексов цен органами российской государственной статистики.

Увязка индекса, рассчитанного по новым, измененным весам (после их пересмотра), осуществляется с помощью процедуры смыкания динамических рядов (глава 10).

Одним из важнейших показателей статистики цен, широко используемым в экономической и социальной политике государства, является *индекс потребительских цен* (ИПЦ). Он применяется для пересмотра правительственных социальных программ, служит основой для повышения минимального размера заработной платы, отражает реальную покупательную способность денег, которыми располагают различные слои населения для удовлетворения своих материальных, культурных и духовных потребностей.

Методология расчета этого показателя включает:

1) отбор товаров (услуг)-представителей и торговых предприятий, по которым производится регистрация цен. Для вычисления ежемесячного ИПЦ отбор товаров (услуг)-представителей производится в соответствии с Общероссийским классификатором экономической деятельности, продукции, услуг и вновь разработанным классификатором на платные услуги населению.

От него несколько отличается набор товаров (услуг)-представителей, используемый для исчисления еженедельного индекса цен, целью которого является наиболее полное отражение изменения цен в условиях нестабильности потребительского рынка;

2) формирование структуры весов по отдельным группам товаров и услуг для расчета сводного индекса потребительских цен.

Для этого используются данные о структуре потребительских расходов населения.

Методология исчисления ИПЦ предполагает расчет индекса для отдельных регионов, товарных групп и услуг, отдельных групп населения с различным уровнем доходов, а также федерального индекса цен.

Если подходить к классификации индексов с чисто математических (формальных) позиций, то *все индексы* (не только индексы цен) можно разделить *на две группы:*

- индексы, при исчислении которых использовались веса базисного периода (формула Ласпейреса);
- индексы, рассчитанные по весам отчетного периода (формула Пааше).

В табл. 12.7 приведены варианты определения агрегатных индексов физического объема и цен.

Эти индексы, а также индекс стоимости (12.13) находятся в определенных соотношениях, которые порой бывает полезно применять в индексных расчетах.

Рассмотрим более подробно *свойства индексов Ласпейреса и Пааше.* Для удобства изложения введем следующие обозначения:

$I_p^{\text{П}}$, $I_q^{\text{П}}$ – индекс цен, физического объема с текущими весами (индекс Пааше);

$I_p^{\text{Л}}$, $I_q^{\text{Л}}$ – индекс цен, физического объема с базисными весами (индекс Ласпейреса).

Таблица 12.7

Индекс Ласпейреса и Пааше

Наименование индекса	Формула индекса	
	Ласпейреса (индекс с базисными весами)	Пааше (индекс с отчетными весами)
Индекс физического объема	$\dfrac{\sum q_1 p_0}{\sum q_0 p_0}$	$\dfrac{\sum q_1 p_1}{\sum q_0 p_1}$
Индекс цен	$\dfrac{\sum p_1 q_0}{\sum p_0 q_0}$	$\dfrac{\sum p_1 q_1}{\sum p_0 q_1}$

Свойство 1:

$$I_{\text{р}}^{\text{П}} = \frac{\sum p_1 q_1}{\sum p_0 q_1} = \frac{I_{pq}}{I_q^{\text{Л}}} = \frac{\sum p_1 q_1}{\sum p_0 q_0} : \frac{\sum q_1 p_0}{\sum q_0 p_0}, \tag{12.54}$$

т.е. индекс цен в формуле Пааше равен отношению индекса стоимости продукции к индексу физического объема в формуле Ласпейреса.

Свойство 2:

$$I_{\text{р}}^{\text{Л}} \cdot I_q^{\text{П}} = I_p^{\text{П}} \cdot I_q^{\text{Л}} = I_{pq} \tag{12.55}$$

или

$$\frac{\sum p_1 q_0}{\sum p_0 q_0} \cdot \frac{\sum q_1 p_1}{\sum q_0 p_1} = \frac{\sum p_1 q_1}{\sum p_0 q_1} \cdot \frac{\sum q_1 p_0}{\sum q_0 p_0} = \frac{\sum p_1 q_1}{\sum p_0 q_0}. \tag{12.56}$$

Данное свойство позволяет сократить объем вычислительной работы. Действительно, для определения индекса цен и физического объема необходимо иметь две величины условной стоимости: $\Sigma p_1 q_0$ и

$\Sigma p_0 q_1$. Если рассчитать величину базисного объема товаров в текущих ценах ($p_1 q_0$), то можно исчислить сначала индекс цен по формуле Ласпейреса ($I_p^{\text{Л}}$), а затем, разделив этот индекс на индекс стоимости (I_{pq}), получить индекс физического объема по формуле Пааше ($I_p^{\text{П}}$).

Свойство 3:

$$I_p^{\text{Л}} = \frac{\sum p_1 q_0}{\sum p_0 q_0} = \frac{\sum p_1 q_0 \cdot (p_0 / p_0)}{\sum p_0 q_0} = \frac{\sum p_1 q_1 \cdot (p_1 / p_0)}{\sum p_0 q_0} = \frac{\sum p_0 q_0 \cdot (i_p)}{\sum p_0 q_0}. \quad (12.57)$$

Если имеются индивидуальные индексы цен, то индекс цен по формуле Ласпейреса может быть исчислен как средняя арифметическая величина, где в качестве весов используется стоимость продукции базисного периода $(p_0 q_0)$. Именно этот способ определения индекса цен наиболее часто используется на практике зарубежными статистиками.

Свойство 4:

$$I_q^{\text{Л}} = \frac{\sum q_1 p_0}{\sum q_0 p_0} = \frac{\sum q_1 p_0 \cdot (q_0 / q_0)}{\sum q_0 p_0} = \frac{\sum q_0 p_0 \cdot (q_1 / q_0)}{\sum q_0 p_0} = \frac{\sum q_0 p_0 \cdot (i_q)}{\sum q_0 p_0}. \quad (12.58)$$

Индекс физического объема по формуле Ласпейреса – это средняя арифметическая величина из индивидуальных индексов объема (i_q), взвешенных по стоимости базисного периода ($p_0 q_0$).

Свойство 5:

$$I_p^{\text{П}} = \frac{\sum p_1 q_1}{\sum p_0 q_1} = \frac{\sum p_1 q_1 \cdot (p_0 / p_0)}{\sum p_0 q_1} = \frac{\sum p_0 q_1 \cdot (p_1 / p_0)}{\sum p_0 q_1} = \frac{\sum p_0 q_1 \cdot (i_p)}{\sum p_0 q_1}. \quad (12.59)$$

В данном случае в качестве весов используется условная стоимость – ($p_0 q_1$).

Свойство 6:

$$I_q^{\text{П}} = \frac{\sum q_1 p_1}{\sum q_0 p_1} = \frac{\sum q_1 p_1 \cdot (q_0 / q_0)}{\sum q_0 p_1} = \frac{\sum q_0 p_1 \cdot (q_1 / q_0)}{\sum q_0 p_1} = \frac{\sum q_0 p_1 \cdot (i_q)}{\sum q_0 p_1}. \quad (12.60)$$

Весами служит стоимость продукции базисного периода, исчисленная в ценах отчетного периода ($p_1 q_0$).

При взвешивании индекса по величине стоимости продукции базисного периода (p_0q_0) возникает постоянная погрешность, причиной которой является тот факт, что цена входит как сомножитель в веса и между изменениями цен весов существует корреляция:

$$I_p^{\text{П}} : I_p^{\text{Л}} = 1 + r_{ip,iq} \cdot V_{ip} \cdot V_{iq}, \tag{12.61}$$

где $r_{ip,\,iq}$ – коэффициент корреляции между индивидуальными индексами физического объема и цен на отдельные виды продукции;

V_{iq} – коэффициент вариации индивидуальных индексов физического объема продукции;

V_{ip} – коэффициент вариации индивидуальных индексов цен.

Так как коэффициенты вариации всегда положительны, а величина коэффициента корреляции между изменениями цен и физического объема на товарном рынке обычно отрицательна, то значение индекса по формуле Пааше всегда меньше значения индекса по формуле Ласпейреса.

12.10
ИДЕАЛЬНЫЙ ИНДЕКС ФИШЕРА

Индекс цен американского экономиста И. Фишера представляет собой среднюю геометрическую из произведения двух агрегатных индексов цен Ласпейреса и Пааше:

$$I_p = \sqrt{\frac{\sum p_1 q_0}{\sum p_0 q_0} \cdot \frac{\sum p_1 q_1}{\sum p_0 q_1}}. \tag{12.62}$$

Формула, предложенная Фишером, может быть использована и для определения индекса физического объема:

$$I_q = \sqrt{\frac{\sum q_1 p_0}{\sum q_0 p_0} \cdot \frac{\sum q_1 p_1}{\sum q_0 p_1}}. \tag{12.63}$$

Геометрическая форма индексов имеет принципиальный недостаток: она лишена конкретного экономического содержания. Так, в отличие от агрегатного индекса Ласпейреса или Пааше, разность между числителем и знаменателем не покажет никакой реальной экономии (или потерь) из-за изменения цен или физического объема продукции.

И Фишер назвал эту формулу расчета индекса идеальной формулой. Идеальность формулы заключается, прежде всего, в том, что индекс является обратимым во времени, т.е. при перестановке базисного и отчетного периодов получается «обратный» индекс, т.е. величина, обратная величине первоначального индекса. Этому условию отвечает любой индивидуальный индекс. Например, индекс цен равен:

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

тогда обратный индекс цен определяется следующим образом:

$$\frac{1}{i_p} = \frac{p_0}{p_1}.$$

Если перемножить эти два индекса, то получится 1:

$$\frac{p_1}{p_0} \cdot \frac{p_0}{p_1} = 1.$$

Этому условию удовлетворяет идеальный индекс цен Фишера:

$$\frac{\sum p_1 q_1}{\sum p_0 q_0} \cdot \frac{\sum p_1 q_1}{\sum p_0 q_0} \cdot \sqrt{\frac{\sum p_0 q_0}{\sum p_1 q_0} \cdot \frac{\sum p_0 q_1}{\sum p_1 q_1}} = 1. \tag{12.64}$$

Индекс Фишера в силу сложности расчета и трудности экономической интерпретации на практике используется довольно редко. Чаще всего он применяется при исчислении индексов цен за длительный период времени для сглаживания тенденций в структуре и составе объема продукции, в которых происходят значительные изменения.

12.11 ИНДЕКСЫ-ДЕФЛЯТОРЫ

Пересчет важнейших стоимостных показателей системы национальных счетов (национальный доход, валовой национальный продукт и т.д.) из фактических цен в сопоставимые осуществляется с помощью индекса-дефлятора. *Дефлятор – это коэффициент, переводящий значение стоимостного показателя за отчетный период в стоимостные измерители базисного.* Например, индекс-дефлятор валового внутреннего продукта (ВВП) представляет собой индекс цен, применяемый для корректировки номинального объема ВВП с учетом инфляции и получения на этой основе реального его объема[1].

Индекс-дефлятор рассчитывается как отношение фактической стоимости продукции отчетного периода к стоимости продукции, структура которой аналогична структуре отчетного года, но определенная в ценах базисного года. В основе расчета индекса-дефлятора лежит формула Пааше – агрегатная формула индекса с текущими весами. Индекс-дефлятор для ВВП в 2001 г. определяется по формуле:

$$I_d = \frac{\sum p_{2001} \cdot q_{2001}}{\sum p_0 \cdot q_{2001}}, \tag{12.65}$$

где I_d – индекс-дефлятор;
q_{2001} – объем продукции в 2001 г.;
p_{2001}, p_0 – цены, фактически действовавшие в 2001 г. и базисном году соответственно.

Реальный ВВП за 2001 г. определяется по формуле:

$$R_{2001} = Q_{2001} : I_d, \tag{12.66}$$

где Q_{2001} – номинальный ВВП.

[1] Согласно экономической теории с помощью номинального ВВП измеряется объем продукции текущего года (q_1) в текущих ценах (p_1). С помощью реального ВВП измеряется объем продукции текущего года (q_1) в ценах, которые сложились в базисном году (p_0).

Формулу (12.65) можно представить и в следующем виде:

$$I_d = \frac{Q_{2001}}{R_{2001}}. \tag{12.67}$$

Индекс-дефлятор для 2002 г. может быть исчислен так:

$$I_d = \frac{\sum p_{2002} \cdot q_{2002}}{\sum p_0 \cdot q_{2002}}, \tag{12.68}$$

где q_{2002} – объем продукции в 2002 г.;
p_{2002} – цены, фактически действовавшие в 2002 г.

Сравним формулы (12.65) и (12.68). Легко заметить, что в них используются различные веса (q_{2001} и q_{2002}). Поэтому важной особенностью индекса-дефлятора является то, что он не может быть использован для сравнительной оценки динамики цен за два периода, в данном случае за 2001 и 2002 гг. Индексы-дефляторы дают представление только об отношении стоимости продукции в текущем периоде к ее стоимости в базисном периоде.

При этом не учитывается отличие состава и структуры продукции в базисный период по сравнению с отчетным. Таким образом, индекс-дефлятор – это самостоятельный показатель.

В статистической практике индексы-дефляторы определяются не только в целом по народному хозяйству; они исчисляются по отдельным регионам, различным товарным группам, каналам реализации потребительских благ, отраслям экономики и т.д.

ОСНОВНЫЕ ПОНЯТИЯ

Агрегатный индекс – сложный относительный показатель, который характеризует среднее изменение социально-экономического явления, состоящего из несоизмеримых элементов.

Вес индекса – величина, служащая для целей соизмерения индексируемых величин.

Индекс – относительный показатель, который выражает соотношение величин какого-либо явления во времени, в пространстве или сравнение фактических данных с любым эталоном (план, прогноз, норматив и т.д.).

Индекс-дефлятор – отношение фактической стоимости продукции отчетного периода к стоимости объема продукции, структура которой аналогична структуре отчетного года, но определенная в ценах базисного года.

Индексируемая величина – признак, изменение которого изучается.

Индекс переменного состава – индекс, выражающий отношение средних уровней изучаемого явления, относящихся к разным периодам времени.

Индекс постоянного (фиксированного) состава – индекс, исчисленный с весами, зафиксированными на уровне одного какого-либо периода, и показывающий изменение только индексируемой величины.

Индекс структурных сдвигов – индекс, характеризующий влияние изменения структуры изучаемого явления на динамику среднего уровня этого явления.

Индивидуальные индексы – относительные показатели, которые отражают результат сравнения однотоварных явлений.

Сводный, или общий, индекс – показатель, измеряющий динамику сложного явления, составные части которого непосредственно несоизмеримы.

Система базисных индексов – ряд последовательно вычисленных индексов одного и того же явления с постоянной базой сравнения.

Система индексов – ряд последовательно построенных индексов.

Система индексов с переменными весами – система сводных индексов одного и того же явления, вычисленных с весами, последовательно меняющимися от одного индекса к другому.

Система индексов с постоянными весами – система сводных индексов одного и того же явления, вычисленных с весами, не меняющимися при переходе от одного индекса к другому.

Система цепных индексов – ряд индексов одного и того же явления, вычисленных с меняющейся от индекса к индексу базой сравнения.

Средний индекс – индекс, вычисленный как средняя величина из индивидуальных индексов.

Территориальные индексы – индексы, которые отражают изменение явления во времени.

ТЕСТЫ

1. Индекс стоимости продукции исчисляется по формуле:

а) $\frac{\sum p_1 q_1}{\sum p_0 q_1}$; б) $\frac{\sum p_1 q_1}{\sum p_0 q_0}$; в) $\frac{\sum q_1 p_1}{\sum q_0 p_1}$.

2. $\frac{\sum p_1 q_1}{\sum p_0 q_0}$; $\frac{\sum p_2 q_2}{\sum p_1 q_1}$; ...; $\frac{\sum p_n q_n}{\sum p_{n-1} q_{n-1}}$ **– это система индексов стоимости:**

а) цепная;

б) базисная.

3. $\frac{\sum z_0 q_1}{\sum z_0 q_0} : \frac{\sum q_1}{\sum q_0}$ **– это:**

а) индекс переменного состава;

б) индекс постоянного состава;

в) индекс структурных сдвигов.

4. Индекс цен Ласпейреса определяется по формуле:

а) $\frac{\sum p_1 q_1}{\sum p_0 q_1}$; б) $\frac{\sum p_1 q_1}{\sum p_0 q_0}$; в) $\sqrt{\frac{\sum p_1 q_0}{\sum p_0 q_0} \cdot \frac{\sum p_1 q_1}{\sum p_0 q_1}}$.

5. Индекс количества продукции, произведенной в единицу времени, рассчитывается по формуле:

а) $\frac{q_1}{T_1} : \frac{q_0}{T_0}$; б) $\frac{t_0}{t_1}$; в) $\frac{q_1 p}{T_1} : \frac{q_0 p}{T_0}$; г) $\frac{t_1 q_1}{t_0 q_0}$.

6. Индекс Струмилина рассчитывается:

а) как средний арифметический индекс;

б) как средний гармонический индекс;

в) как средний геометрический индекс.

7. Система базисных индексов физического объема продукции с постоянными весами имеет следующий вид:

а) $\frac{\sum q_1 p_0}{\sum q_0 p_0}$; $\frac{\sum q_2 p_0}{\sum q_1 p_0}$; ...; $\frac{\sum q_n p_0}{\sum q_{n-1} p_0}$;

б) $\frac{\sum q_1 p_0}{\sum q_0 p_0}$; $\frac{\sum q_2 p_0}{\sum q_0 p_0}$; ...; $\frac{\sum q_n p_0}{\sum q_0 p_0}$;

в) $\frac{\sum q_1 p_1}{\sum q_0 p_0}$; $\frac{\sum q_2 p_2}{\sum q_0 p_0}$; ...; $\frac{\sum q_n p_n}{\sum q_0 p_0}$.

8. Если себестоимость увеличилась на 14%, а количество продукции снизилось на 6%, то индекс издержек производства будет равен:

а) 107;

б) 120;

в) 121.

9. Индекс-дефлятор – это индекс:

а) из системы цепных индексов цен с переменными весами;

б) из системы цепных индексов с постоянными весами;

в) из системы базисных индексов с переменными весами;

г) из системы базисных индексов с постоянными весами.

10. Если индекс переменного состава равен 118%, а индекс структурных сдвигов 107%, то индекс фиксированного состава равен:

а) 110;

б) 111;

в) 115.

ЛИТЕРАТУРА

Адамов В.Е. Факторный индексный анализ (методология и проблемы). – М.: Статистика, 1977. – 199 с.

Аллен Р. Экономические индексы. – М.: Статистика, 1980. – 256 с.

Бакланов Г.И. Некоторые вопросы индексного метода. – М.: Статистика, 1972. – 72 с.

Елисеева И.И., Юзбашев М.М. Общая теория статистики. – М.: Финансы и статистика, 1998. – 480 с.

Казинец Л.С. Теория индексов (основные вопросы). – М.: Госстатиздат, 1963. – 352 с.

Кевеш П. Теория индексов и практика экономического анализа. – М.: Финансы и статистика, 1990. – 303 с.

Ковалевский Г.В. Индексный метод в экономике. – М.: Финансы и статистика, 1989. – 238 с.

Перегудов В.Н. Теоретические вопросы индексного анализа. – М.: Госстатиздат, 1960. – 267 с.

Плошко Б.Г. Индексы. – Л.: Изд-во Ленинградского университета, 1958. – 91 с.

Стоимость жизни и ее измерение. – М.: Финансы и статистика, 1991. – 174 с.

Торвей Р. Индексы потребительских цен: Методическое руководство. – М.: Финансы и статистика, 1993. – 248 с.

ГЛАВА 13
ОБЩИЕ ВОПРОСЫ АНАЛИЗА И ОБОБЩЕНИЯ СТАТИСТИЧЕСКИХ ДАННЫХ

13.1 ПОНЯТИЕ И ОСНОВНЫЕ ПРИНЦИПЫ ЭКОНОМИКО-СТАТИСТИЧЕСКОГО АНАЛИЗА

Анализ и обобщение статистических данных – заключительный этап статистического исследования, конечной целью которого является получение теоретических выводов и практических заключений о тенденциях и закономерностях изучаемых социально-экономических явлений и процессов.

Анализ – это метод научного исследования объекта путем рассмотрения его отдельных сторон и составных частей.

Экономико-статистический анализ – это разработка методики, основанной на широком применении традиционных статистических и математико-статистических методов, с целью контроля адекватного отражения исследуемых явлений и процессов.

Задачами статистического анализа являются: определение и оценка специфики и особенностей изучаемых явлений и процессов, изучение их структуры, взаимосвязей и закономерностей их развития.

В *качестве этапов статистического анализа* выделяются:

- формулировка цели анализа;
- критическая оценка данных;
- сравнительная оценка и обеспечение сопоставимости данных;
- формирование обобщающих показателей;
- фиксация и обоснование существенных свойств, особенностей, сходств и различий, связей и закономерностей изучаемых явлений и процессов;
- формулировка заключений, выводов и практических предложений о резервах и перспективах развития изучаемого явления.

Методы анализа должны меняться в зависимости от характера изучаемых процессов, их специфики, особенностей и форм проявления.

Статистический анализ данных проводится в неразрывной связи теоретического, качественного анализа сущности исследуемых явлений и соответствующего количественного инструментария, изучения их структуры, связей и динамики.

Экономико-статистический анализ должен проводиться при строгом соблюдении следующих принципов, которые учитывают их экономическую и статистическую градацию.

Экономическими принципами являются:

- соответствие экономическим законам и положениям теории расширенного воспроизводства;
- адекватное отражение сущности экономической политики современного этапа общественно-экономического развития;
- ориентация на конечные экономические результаты;
- учет специфики изучаемого объекта, отрасли и т.д.;
- согласование интересов субъектов различных иерархических уровней как подразделений единого народнохозяйственного механизма.

К *статистическим принципам* относятся:

- четко определенная цель экономико-статистического исследования;
- согласованность систем по горизонтали и вертикали;
- сопоставимость во времени и пространстве;
- логическая взаимосвязь между показателями, характеризующими объект или явление;
- комплексность и полнота отображения объекта исследования в статистических показателях;
- максимальная степень аналитичности.

Соблюдение данных принципов наряду с предпосылками применения методологии статистического анализа позволяет осуществить научно обоснованное экономико-статистическое исследование субъектов экономики в соответствии с принятой международной методологией учета и статистики.

13.2 АПРИОРНЫЙ АНАЛИЗ И ЕГО РОЛЬ В ИССЛЕДОВАНИИ СОЦИАЛЬНО-ЭКОНОМИЧЕСКИХ ЯВЛЕНИЙ

Оценка эффективности и деловой активности субъектов экономического процесса и состояния социальной инфраструктуры общества во многом зависит от качества статистического анализа эмпирического материала, а также от того, насколько точно будут выявлены и научно обоснованы закономерности и тенденции развития. Основные трудности, связанные с применением количественных математико-статистических методов, заключаются в том, что они достаточно нейтральны к исследуемым социально-экономическим процессам. Поэтому важным этапом статистического исследования на информационной базе, характеризующей реальные социально-экономические явления, является критическая оценка исходных данных с точки зрения их достоверности и научной обоснованности.

Под *критической оценкой* статистического материала следует понимать полноту, качество и достоверность его соответствия целям и задачам исследования.

Надежность выводов и заключений по анализу статистических данных обеспечивается минимизацией в исходной информации пробелов, неточностей, несопоставимости, неопределенности и т.д.

Во время информационного бума, которым характеризуется современный этап общественного развития, необходимо больше внимания уделять критической оценке и априорному анализу исходной статистической информации. Развитие новых организационно-правовых форм во всех сферах общественной жизни, наличие коммерческой тайны и так далее увеличивают вероятность получения преднамеренно искаженных фактов, затушевывающих результаты производственно-хозяйственной деятельности фирм, банков и других структур.

Методы априорного анализа включают:

- выявление экономически обоснованных и существенных причинно-следственных связей между признаками и явлениями;
- оценку однородности исследуемой совокупности;
- анализ характера распределения совокупности по изучаемым признакам.

Понятия, используемые при проведении анализа статистическими методами, должны быть точно определены.

Например, при изучении производительности труда в строительстве необходимо установить, будет ли выработка рассчитываться на одного работника или одного рабочего, занятого в строительном производстве. Необходимо также четко определить, к какому моменту или периоду времени относится исследуемое явление или процесс.

Одной из основополагающих предпосылок проведения научно обоснованного статистического анализа, адекватно отражающего причинно-следственные связи и зависимости, тенденции развития реальных явлений и процессов в статике и динамике, является однородность статистической совокупности (глава 3).

Анализ *однородности статистической совокупности* целесообразно проводить в следующей последовательности:

- определение степени однородности всей совокупности по одному или нескольким существенным признакам;
- определение и анализ аномальных наблюдений;
- выбор оптимального варианта выделения однородных совокупностей.

В статистической теории и практике разработаны различные подходы к оценке степени однородности. Проблемой оценки однородности совокупности занимались такие известные ученые, как Ю. Аболенцев, Г. Кильдишев, В. Овсиенко и др.

Наиболее сложным и дискуссионным является вопрос о способах и критериях выделения однородных групп объектов в пределах исходной совокупности.

Важной предпосылкой получения научно обоснованных результатов статистического анализа является проверка и выполнение гипотезы о близости распределения эмпирических данных нормальному закону. Для нормального закона распределения характерно, что

$$\bar{X} \approx \mathrm{Mo} \approx \mathrm{Me}; \quad \mathrm{As} = 0, \; \mathrm{E\kappa} = 0.$$

Одним из недостатков данного подхода к оценке характера распределения считается наличие субъективности в анализе достаточности величины отклонения $\bar{X}$ от Ме и Мо. В данном случае на практике используются критерии Пирсона, Романовского, Ястремского, Колмогорова (глава 7).

Любая исследуемая совокупность наряду со значениями признаков, сложившихся под влиянием факторов, непосредственно характерных для анализируемой совокупности, может содержать и значения признаков, полученных под воздействием иных факторов, не характерных для основной совокупности. Такие значения резко выделяются, и, следовательно, использование методологии статистического анализа данной совокупности без предварительного анализа и изучения аномальных наблюдений приводит к серьезным ошибкам. Резко выделяющиеся из общей совокупности наблюдения требуют изучения.

Причины появления в совокупности аномальных наблюдений условно подразделяют следующим образом:

- внешние, возникающие в результате технических ошибок;
- внутренние, объективно существующие.

Такие наблюдения представляют интерес для исследователя, так как могут содержать за счет влияния неучтенных факторов особую информацию. На практике в зависимости от условий места и времени влияние одних факторов в каждый конкретный исследуемый момент или промежуток времени значительнее, чем других.

Выбор того или иного метода выявления, анализа аномальных наблюдений определяется объемом совокупности, характером исследуемых процессов и задач (одномерных и многомерных).

При реализации одномерных задач, при анализе как динамической, так статической информации наиболее широкое применение получил *метод выявления аномальных наблюдений*, основанный на определении q-статистики:

$$q_t = \frac{|y_t - \bar{y}|}{\text{у}_y}, \tag{13.1}$$

где y_t – отдельные уровни ряда;
$\bar{y}$ – средний уровень ряда;
σ_y – среднее квадратическое отклонение значений ряда от их среднего уровня.

Если для расчетного значения выполняется неравенство вида

$$q_t \geq q_{tkp}\ (p) \tag{13.2}$$

с заранее заданным уровнем вероятности, то данные наблюдения считаются аномальными и после логико-экономического анализа причин ошибок аномальности подлежат замене скорректированным значением (в случае ошибки «I») и не подлежат корректировке (в случае «II»).

Корректировка осуществляется по схеме.

1. Рассчитывается новое значение уровня ряда:

$$y_{i(1)}^{(1)} = q_{kp}(p)\text{у}_y + \bar{y}. \tag{13.3}$$

2. $y_{i(1)}$ заменяется на $y_i^{(1)}$.
3. Определяются новые характеристики ряда с $y_i^{(1)}$: $\bar{y}_i^{(1)}$ и $\sigma_y^{(1)}$.
4. $y_i^{(2)} = q_{kp}(\text{p})\sigma_{\text{у}}^{(1)} + \bar{y}^{(1)}$. (13.4)

5. Проверяется аномальность значения $y_i^{(1)}$:

$$\left|y_i^{(1)} - y_i^{(2)}\right| \leq е, \tag{13.5}$$

где ε – заданный уровень точности определения $y_i^{(k)}$.

Если данное условие выполняется, то значение $y_i^{(1)}$ является скорректированным, не аномальным значением, занимает место $y_{i(1)}$ в ряду и анализу подвергается $y_{i(2)}$.

Если условие не выполняется, то рекомендуется рассчитать $y_{i(2)}$ и проверить на аномальность. Процесс корректировки носит итерационный характер.

При исследовании динамики наибольшее распространение получил *метод Ирвина,* основанный на определении λ_i-статистики. При его использовании выявление аномальных наблюдений производится по схеме:

$$л_i = \frac{\left|y_i - y_{i-1}\right|}{у_y}. \tag{13.6}$$

Если расчетное значение превысит уровень критического (с заданным уровнем точности и числом наблюдений) (табл. 13.1), то оно признается аномальным.

Схема реализации данного метода аналогична предыдущей с той лишь разницей, что $\overline{y}$ заменяется на y_{i-1} (предыдущее значение ряда).

Способ, основанный на расчете *q-статистики*, применим для относительно стационарных рядов, так как при использовании для анализа динамических рядов, имеющих ярко выраженную тенденцию, он приводит к ошибкам. Более корректным является применение статистики, в которой определяются отклонения от теоретических значений, полученных по уравнению тренда:

$$q_i = \frac{\left|y_i - \overline{y}_t\right|}{у_y}. \tag{13.7}$$

В общем виде, систематизируя методику анализа аномальных наблюдений, описанную в статистической литературе, градацию статистических методов выявления аномальности в исходных данных можно представить следующей схемой (рис. 13.1).

Таблица 13.1

Табулированные значения λ

Число наблюдений	λ_{kp}	
	0,95	0,99
2	2,8	3,7
3	2,2	2,9
10	1,5	2,0
20	1,3	1,8
30	1,3	1,7
50	1,1	1,6
100	1,0	1,5

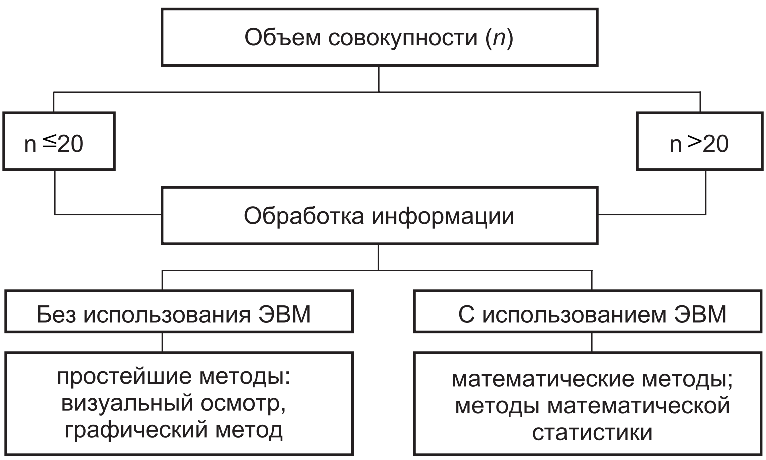

Рис. 13.1. Методы выявления аномальных наблюдений

Нецелесообразность исключения аномальных наблюдений из изучаемой совокупности реализуется широким использованием метода группировок.

Важной задачей статистических исследований на этапе априорного анализа является выделение однородных групп (даже аномальных). В данном случае в анализе эффективно применять сложные комбинационные группировки с развернутым сказуемым (глава 3).

Всесторонний качественный анализ исходных данных является залогом проведения научно обоснованного, логически выверенного экономико-статистического исследования социально-экономических явлений и процессов.

13.3 КОМПЛЕКСНОЕ ПРИМЕНЕНИЕ МАТЕМАТИКО-СТАТИСТИЧЕСКИХ МЕТОДОВ АНАЛИЗА ДАННЫХ

Наряду с традиционными статистическими методами анализа данных, подробно описанными в главах 3–12 данного учебника, при исследовании реальных социально-экономических явлений и процессов широко используются математико-статистические методы исходя из отечественной и зарубежной методологии.

Комплексность применения математико-статистических методов предполагает наиболее полное раскрытие сущности, закономерностей и тенденций развития конкретных явлений и процессов с целью более адекватного отражения их свойств и особенностей, резервов и перспектив развития и путей совершенствования.

Усложнение структуры социально-экономических явлений предполагает использование ряда методов классификации и выделение однородных групп, в основе построения которых лежат меры близости или метрики. Сущность заключается в том, что распределение исследуемых объектов или явлений в совокупности должно подчиняться нормальному закону распределения, с тем чтобы получить модели, которые действительно будут отражать качественно однородные группы.

Наибольшее распространение в практике анализа экономических явлений и процессов получили: кластерный анализ, метод главных компонент, факторный анализ.

Пусть имеется *n* объектов, каждый из которых характеризуется набором *k*-признаков. Требуется разбить эту совокупность на однородные группы. Полученные в результате разбиения группы называются *кластерами*, а метод их нахождения – *кластерным анализом*.

Наиболее трудным считается определение однородности объектов, которые задаются введением расстояния между объектами x_i и x_j ($\rho\,(x_i, x_j)$).

Объекты будут однородными в случае

$$\rho\,(x_i, x_j) \le \rho_{\text{пор}},$$

где $\rho_{\text{пор}}$ – заданное пороговое значение.

Выбор расстояния (ρ) является основным моментом исследования, от которого зависят окончательные варианты разбиения. Наиболее распространенными считаются процедуры «ближайшего соседа», основанные на близости объектов по совокупности рассматриваемых признаков, и «дальнего соседа».

В задачах кластерного анализа часто используют *Евклидово и Хемингово расстояния*.

Евклидово расстояние:

$$c_{\mathrm{E}}(x_i, x_j) = \sqrt{\left(x_i^{1,2,\dots,k} - x_j^{1,2,\dots,k}\right)^2};$$

сравнивается близость двух объектов по большому числу *признаков*.

Хемингово расстояние:

$$c_{\mathrm{X}}(x_i, x_j) = \sum_{i,j}^{1,2,\dots k} \left| x_i^{1,2,\dots k} - x_j^{1,2,\dots k} \right|;$$

используется как мера различия объектов, задаваемых атрибутивными признаками.

Выбор метрики-расстояния определяется структурой признакового пространства и целью классификации.

При использовании процедур кластерного анализа расчленение объектов совокупности на качественно однородные группы производится одновременно по большому числу признаков, но при соблюдении условия, что ни один признак не выделяется по своей значимости так, что группировка на его основе является главной. Особенностью кластерного анализа является то, что различия между единицами, входящими в выделенную группу, незначительны, а различия между группами существенны.

Пример. Провести классификацию шести объектов, каждый из которых характеризуется двумя признаками (табл. 13.2).

Таблица 13.2

№ п/п	1	2	3	4	5	6
x_1	2	4	5	12	14	15
x_2	8	10	7	6	6	4

где x_1 – объем выпускаемой продукции, млн руб.;
x_2 – среднегодовая стоимость основных промышленно-производственных фондов, млн руб.

Представим эти данные графически (рис. 13.2).

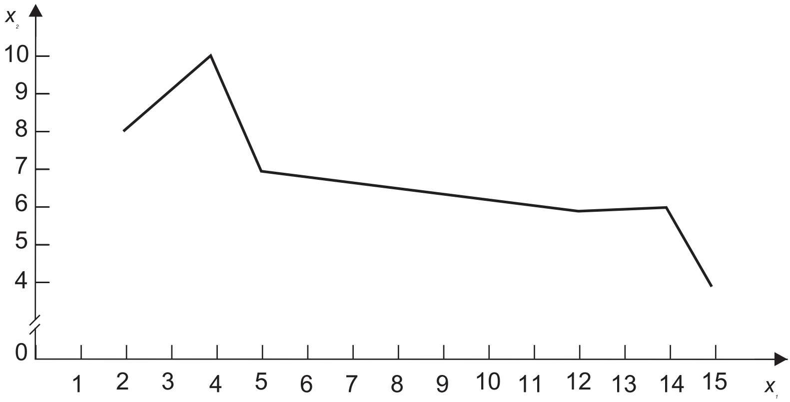

Рис. 13.2. Зависимость между объемом выпускаемой продукции (x_1) и среднегодовой стоимостью основных промышленно-производственных фондов (x_2)

$$\text{c}(x_i x_j) = \sqrt{\sum_{i=I}^{k}\left(x_{il} - x_{jl}\right)^2},$$

где l – признаки;
k – количество признаков.

$$\rho_{11} = 0;$$

$$\text{c}_{12} = \sqrt{(2-4)^2 + (8-10)^2} = \sqrt{8} \approx 2{,}83.$$

Расчеты последующих $\rho(x_i, x_j)$ аналогичны:

$\rho_{13} = 3{,}16$; $\rho_{14} = 10{,}19$; $\rho_{15} = 12{,}17$; $\rho_{16} = 13{,}60$;

$\rho_{23} = 3{,}16$; $\rho_{24} = 8{,}94$; $\rho_{25} = 10{,}77$; $\rho_{26} = 12{,}53$;

$\rho_{34} = 7{,}07$; $\rho_{35} = 9{,}06$; $\rho_{36} = 10{,}44$; $\rho_{45} = 2{,}00$;

$\rho_{46} = 3{,}61$; $\rho_{56} = 2{,}24$.

1. Принцип «ближайшего соседа».

$$R_1 = \begin{array}{c|c} & \begin{matrix} 1 & 2 & 3 & 4 & 5 & 6 \end{matrix} \\ \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \end{matrix} & \begin{pmatrix} 0 & 2{,}83 & 3{,}16 & 10{,}19 & 12{,}17 & 13{,}60 \\ & 0 & 3{,}16 & 8{,}94 & 10{,}77 & 12{,}53 \\ & & 0 & 7{,}07 & 9{,}06 & 10{,}44 \\ & & & 0 & \boxed{2} & 3{,}61 \\ & & & & 0 & 2{,}24 \\ & & & & & 0 \end{pmatrix} \end{array};$$

$$\rho_{\min} = \rho_{4,5} = 2(S_1;\ S_2;\ S_3;\ S_{4,5};\ S_6);$$

$$R_2 = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4{,}5 \\ 6 \end{array} \left(\begin{array}{ccccc} 1 & 2 & 3 & 4{,}5 & 6 \\ 0 & 2{,}83 & 3{,}16 & 10{,}19 & 13{,}60 \\ & 0 & 3{,}16 & 8{,}94 & 12{,}53 \\ & & 0 & 7{,}07 & 10{,}44 \\ & & & 0 & \boxed{2{,}24} \\ & & & & 0 \end{array} \right);$$

$\rho_{min} = \rho_{4,5,6} = 2{,}24(S_1; S_2; S_3; S_{4,5,6})$;

$$R_3 = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4{,}5{,}6 \end{array} \left(\begin{array}{cccc} 1 & 2 & 3 & 4{,}5{,}6 \\ 0 & \boxed{2{,}83} & 3{,}16 & 10{,}19 \\ & 0 & 3{,}16 & 8{,}94 \\ & & 0 & 7{,}07 \\ & & & 0 \end{array} \right);$$

$\rho_{min} = \rho_{1,2} = 2{,}83\ (S_{1,2}; S_3; S_{4,5,6})$;

$$R_4 = \begin{array}{c} \\ 1{,}2 \\ 3 \\ 4{,}5{,}6 \end{array} \left(\begin{array}{ccc} 1{,}2 & 3 & 4{,}5{,}6 \\ 0 & \boxed{3{,}16} & 8{,}94 \\ & 0 & 7{,}07 \\ & & 0 \end{array} \right);$$

$\rho_{min} = \rho_{1,2,3} = 3{,}16\ (S_{1,2,3}; S_{4,5,6})$;

$$R_5 = \begin{array}{c} \\ 1{,}2{,}3 \\ 4{,}5{,}6 \end{array} \left(\begin{array}{cc} 1{,}2{,}3 & 4{,}5{,}6 \\ 0 & 7{,}07 \\ & 0 \end{array} \right).$$

Таким образом, при проведении кластерного анализа по принципу «ближайшего соседа» получили два кластера (рис. 13.3).

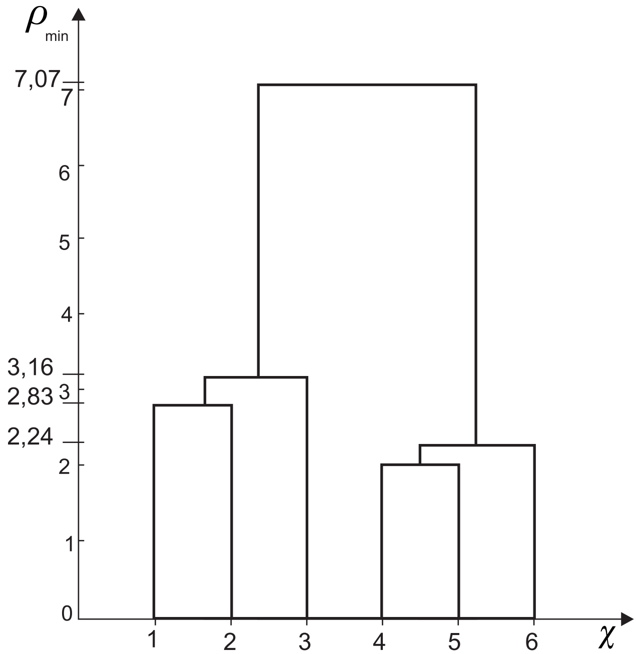

Рис. 13.3. Дендрограмма по принципу «ближайшего соседа»

2. Принцип «дальнего соседа».

$$R_1 = \begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ \\ \end{array} \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 0 & 2{,}83 & 3{,}16 & 10{,}19 & 12{,}17 & 13{,}60 \\ & 0 & 3{,}16 & 8{,}94 & 10{,}77 & 12{,}53 \\ & & 0 & 7{,}07 & 9{,}06 & 10{,}44 \\ & & & 0 & \boxed{2} & 3{,}61 \\ & & & & 0 & 2{,}24 \\ & & & & & 0 \end{pmatrix};$$

$\rho_{min} = \rho_{4,5} = 2$ (S_1; S_2; S_3; $S_{4,5}$; S_6);

$$R_2 = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4,5 \\ 6 \end{array} \begin{pmatrix} 1 & 2 & 3 & 4,5 & 6 \\ 0 & \boxed{2,83} & 3,16 & 12,17 & 13,60 \\ & 0 & 3,16 & 10,77 & 12,53 \\ & & 0 & 9,06 & 10,44 \\ & & & 0 & 3,61 \\ & & & & 0 \end{pmatrix};$$

$\rho_{min} = \rho_{1,2} = 2{,}83$ ($S_{1,2}$; S_3; $S_{4,5}$; S_6);

$$R_3 = \begin{array}{c} \\ 1,2 \\ 3 \\ 4,5 \\ 6 \end{array} \begin{pmatrix} 1,2 & 3 & 4,5 & 6 \\ 0 & \boxed{3,16} & 12,17 & 13,60 \\ & 0 & 9,06 & 10,44 \\ & & 0 & 3,61 \\ & & & 0 \end{pmatrix};$$

$\rho_{min} = \rho_{1,2,3} = 3{,}16$ ($S_{1,2,3}$; $S_{4,5}$; S_6);

$$R_4 = \begin{array}{c} \\ 1,2,3 \\ 4,5 \\ 6 \end{array} \begin{pmatrix} 1,2,3 & 4,5 & 6 \\ 0 & 12,17 & 13,60 \\ & 0 & \boxed{3,61} \\ & & 0 \end{pmatrix};$$

$\rho_{min} = \rho_{4,5,6} = 2{,}61$ ($S_{1,2,3}$; $S_{4,5,6}$);

$$R_5 = \begin{array}{c} \\ 1,2,3 \\ 4,5,6 \end{array} \begin{pmatrix} 1,2,3 & 4,5,6 \\ 0 & 13,60 \\ & 0 \end{pmatrix};$$

При проведении кластерного анализа по принципу «дальнего соседа» получили два кластера (рис. 13.4).

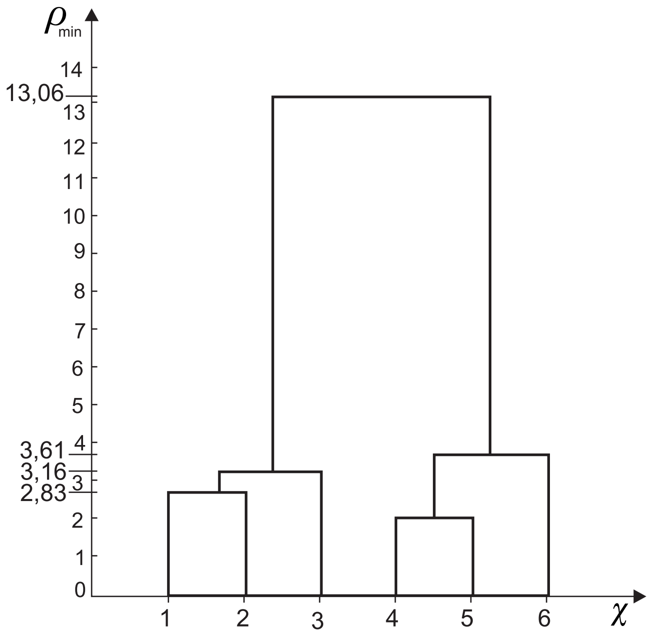

Рис. 13.4. Дендрограмма по принципу «дальнего соседа»

Модели на основе результатов кластерного анализа позволяют исследовать однородные по основным экономико-техническим характеристикам и параметрам деятельности социально-экономические объекты и процессы, а также степень их деловой активности.

По мере углубления анализа деятельности экономических структур – объектов и процессов – в рассмотрение включается все большее число признаков. При этом требуется обозримость. Закономерность распыляется на большое множество связей. Поэтому целесообразно осуществлять классификацию по нескольким обобщающим признакам, полученным с помощью *метода главных компонент* или *факторного анализа*.

Предпосылками, обусловливающими возможность применения этих методов, являются:

- наличие сильно коррелированных признаков, приводящих к дублированию информации;
- слабая информативность ряда факторных признаков;
- возможность и целесообразность агрегирования нескольких факторных признаков.

Итак, сокращение размерности методом главных компонент и факторного анализа предполагает переход от описания результативного признака меньшим числом наиболее информативных (с точки зрения их влияния на результативный), факторных признаков.

Метод главных компонент, как правило, рассматриваемый в качестве средства снижения размерности, также используется для проведения классификаций. Математической моделью, на которой основывается метод главных компонент, является следующая:

$$Y_j = V_{1j}F_1 + V_{2j}F_2 + \ldots + V_{pj}F_p = \sum_{p=1}^{n} V_{pj}F_p, \ \ j = 1, 2, \ldots, p,$$

где $V_{j1,2,\ldots,p}$ – весовой коэффициент общего результативного фактора (главной компоненты) на j-й переменной;

$F_{1,2,\ldots,p}$ – общий фактор (главная компонента).

Сущность метода заключается в выделении линейных комбинаций исходных факторных признаков, имеющих максимально возможную дисперсию. При этом первая главная компонента обладает максимальной дисперсией и является нормированной линейной комбинацией всех возможных исходных признаков, а вторая – учитывает максимальное значение оставшейся дисперсии и корреляционно не связана с первой компонентой.

В целом, если переходить к оценке степени и направления связей между анализируемыми признаками на основе совместного использования методов регрессионного анализа из главных компонент, уравнение главных компонент включает меньшее их число, чем количество факторных признаков, так как из дальнейшего рассмотрения должны быть исключены главные компоненты, вклад которых в общую дисперсию несуществен и составляет менее 10%.

Комплексное использование корреляционного анализа и метода главных компонент выражается в расчете парных коэффициентов корреляции между исходными (включенными в исследование) признаками и соответствующими главными компонентами. На основе общей оценки значения парного коэффициента корреляции (по абсолютной величине) и доли вклада (в процентном выражении) каждой компоненты в общую вариацию признака осуществляется отбор наиболее статистически существенных главных компонент.

К *достоинствам метода главных компонент* относят следующие:

- в матрице значений компоненты размещаются в убывающей последовательности собственных значений, что способствует классификации признаков;
- число компонент соответствует количеству исходных факторных признаков;
- главные компоненты не коррелированы между собой, что существенно при построении регрессионных моделей;
- главные компоненты полностью обусловливают вариацию исходных факторных признаков.

Факторный анализ заключается в переходе от исходной информации к обобщенным факторам, являющимся результатом их первоначальной агрегации и линейной комбинации.

Основная модель факторного анализа линейная и имеет вид:

$$Y_j = a_{1j}F_1 + a_{2j}F_2 + ... + a_{jp}F_p + d_jV_j \ ,$$

где $F_1, F_2, \ldots F_p$ – обобщенные факторы, обусловливающие систематическую вариацию и корреляционную связь между ними;

a_{ij} – факторные нагрузки;

V_j – характерные факторы, учитывающие вариацию, не объяснимую общими факторами.

Факторные нагрузки оценивают степень тесноты связи между исходными признаками $x_1, x_2, \ldots x_k$ и обобщенными факторами F_j. Связь считается существенной, если парный коэффициент корреляции больше или равен по абсолютному значению 0,5.

В практической деятельности вклад общего фактора в общую дисперсию составляет не менее 80–90%.

Совокупное использование методов факторного и регрессионного анализов невозможно без учета специфики и различий между ними.

При применении регрессионного анализа в модель не могут быть включены все переменные, влияющие на результативный признак, что ведет к некоторой потере информации. На основе реализации факторного анализа в модель включаются общие факторы, являющиеся реальным отображением ряда экономически связанных между собой исходных переменных, воспроизводящих и объясняющих их свойства.

Переход от большого числа факторных признаков к обобщенным факторам, или главным компонентам, не служит существенной потерей информации. В отдельных случаях общие факторы могут отражать свойства исходных факторных признаков, непосредственно статистически не измеряемых и влияния на результат не оказывающих.

Модели регрессии на обобщенные факторы и главные компоненты заведомо не содержат коллинеарно связанные признаки.

Однако на практике целесообразнее строить модели регрессии на исходные переменные, так как использование общих факторов и главных компонент сильно осложняет экономическую интерпретацию параметров модели.

Комплексность методологии статистического анализа в оценке статических и динамических совокупностей эмпирических данных осуществляется в следующей последовательности, которая может быть рассмотрена как задание для самостоятельной научно-исследовательской работы студентов.

А. Методика комплексного анализа статической информации и выявления причинно-следственных связей

I. Априорный анализ исходных статистических данных

1. Обобщение исходных данных:
- построение интервальных вариационных рядов по каждому из рассматриваемых показателей на базе предварительного определения целесообразного количества групп;
- графическое изображение полученных в п. 1 рядов распределения в виде гистограммы, полигона, кумуляты и огивы.

2. Оценка однородности совокупности на основе:
- метода группировок;
- показателей вариации;
- анализа аномальных наблюдений на основе λ- и q-статистик.

3. Оценка характера распределения совокупности исходных данных:
- вычисление и анализ средней, моды, медианы, показателей вариации.

Вывод о характере распределения: близки ли к нормальному распределению случайных величин эмпирические распределения, полученные в виде вариационных рядов. С этой целью могут быть использованы различные модификации соотношений средних величин и показателей вариации.

- проверка данных на базе одного из критериев: Пирсона, Ястремского, Романовского, Колмогорова.

II. Моделирование связи социально-экономических явлений.

1. Отбор факторных признаков:

- метод экспертных оценок и ранговые коэффициенты корреляции как инструмент анализа экспертной информации;
- графический метод как способ наглядного отображения зависимости результативного с каждым из факторных признаков;
- метод корреляционного анализа в оценке характера парных и множественных зависимостей между исходными признаками. Расчет парных, частных и множественных коэффициентов корреляции. Исследование связей на мультиколлинеарность.

2. Построение модели связи и оценка ее существенности:

- определение параметров модели методом наименьших квадратов;
- построение уравнения связи методом пошагового регрессионного анализа;
- проверка адекватности регрессионной модели исследуемому социально-экономическому явлению:

  проверка значимости коэффициентов регрессии при факторных признаках, вошедших в модель, на основе t-критерия Стьюдента;

  проверка значимости уравнения регрессии на основе F-критерия Фишера–Снедекора;

  расчет и анализ средней ошибки аппроксимации ($\bar{\varepsilon}$);

  анализ средней квадратической ошибки ($\sigma_{\text{ош}}$) и остаточной дисперсии ($\sigma^2_{\text{ост}}$).

3. Интерпретация модели связи (уравнения регрессии). С этой целью расчет и анализ:

- β-коэффициентов, построение модели связи в стандартизованном масштабе;
- частных коэффициентов эластичности (Э_{x_i});
- частных и множественного коэффициентов детерминации;
- Δ_{xj}-коэффицентов;
- Q_i-коэффицентов.

Б. Методика комплексного анализа и прогнозирования динамической информации

I. Анализ и прогнозирование тенденции.

1. Оценка аномальных наблюдений на основе λ- и q-статистик.

2. Расчет аналитических (Δ_i, $T_{\text{р}}$ и $T_{\text{пр}}$) и средних показателей рядов динамики и на их основе анализ тенденций и закономерностей развития социально-экономических явлений.

3. Определение наличия тенденции средних и дисперсии в рядах динамики на базе методов:
- Фостера–Стюарта;
- сравнения средних уровней ряда динамики.

4. Определение наличия тенденции автокорреляции (для связных рядов динамики) на основе:
- первого циклического коэффициента;
- критерия Неймана;
- критерия Дарбина–Уотсона.

5. Выявление основной тенденции динамического ряда методами:
- укрупнения интервалов;
- усреднения по левой и правой половине;
- простой и взвешенной скользящей средней;
- аналитического выравнивания и определения параметров на базе методов:
  конечных разностей;
  средних значений (линейных отклонений);
  наименьших квадратов.

6. Оценка адекватности выбранного уравнения тренда (в пункте 5 аналитического выравнивания) на основе:
- минимизации сумм квадратов отклонений эмпирических данных от теоретических (расчетных);
- средней квадратической ошибки;
- средней ошибки аппроксимации.

7. Корреляция рядов динамики. Модели, исключающие автокорреляцию, методами:
- последовательных или конечных разностей;
- отклонения эмпирических значений признаков от выравненных по тренду;
- Фриша-Воу.

8. Прогнозирование динамики на основе простейших методов:
- среднего уровня ряда;

- среднего абсолютного прироста;
- среднего темпа роста;
- линейного тренда.

II. Выявление периодической компоненты. Модели сезонных колебаний:

- графический анализ исходных данных;
- выявление тенденции средней и дисперсии;
- проверка ряда динамики на наличие сезонной компоненты на основе критерия «пиков» и «ям» и др.;
- расчет параметров уравнения тренда и определение теоретических уровней ряда динамики по тренду;
- определение абсолютных и относительных отклонений фактических уровней от выравненных по тренду. Графический метод в анализе амплитуды отклонений эмпирических и теоретических значений уровней ряда динамики;
- проверка абсолютных и относительных фактических уровней на наличие автокорреляции;
- построение модели сезонной волны по отклонениям фактических данных от тренда методами гармонического анализа. Определение гармоники Фурье, наилучшим образом отражающей периодичность изменения уровней ряда динамики на основе:
  минимизации суммы квадратов отклонений эмпирических данных от выравненных по гармонике;
  средней квадратической ошибки.

Изучение социально-экономических явлений и процессов на основе комплексной методики анализа, обобщения и прогнозирования на базе широкого применения традиционных статистических и математико-статистических методов позволит глубоко и досконально исследовать причинно-следственные связи и закономерности и показать природу изучаемого явления или процесса.

ОСНОВНЫЕ ПОНЯТИЯ

Заключительным этапом проведения статистического исследования является анализ статистических данных с целью получения теоретический выводов и практических заключений о тенденциях и закономерностях изучаемых социально-экономических явлений и процессов. Рассмотрим основные понятия данной главы учебника.

Анализ – метод научного исследования объекта путем рассмотрения его отдельных сторон и составных частей.

Аномальные наблюдения – наблюдения, характерные для нестабильных явлений и процессов. Выявляются и корректируются с помощью методов Ирвина и q-статистики.

Критическая оценка исходных данных – полнота, качество и достоверность соответствия эмпирического материала целям и задачам исследования.

Экономико-статистический анализ – разработка методики, основанной на широком применении традиционных статистических и математических методов, с целью контроля адекватного отражения исследуемых явлений и процессов.

ЛИТЕРАТУРА

Айвазян С.А., Енюков И.С., Мешалкин Л.Д. Прикладная статистика: Основы моделирования и первичной обработки данных. – М.: Финансы и статистика, 1983. – 471 с.

Бешелев С.Л., Гурвич Ф.Г. Математико-статистические методы экспертных оценок. – М.: Статистика, 1980. – 159 с.

Болч Б., Хуань К. Многомерные статистические методы для экономики. – М.: Статистика, 1979. – 316 с.

ПРИЛОЖЕНИЯ

Приложение 1

Нормальный закон распределения

Значение функции $\Phi(t) = P(\left| T \right|) \le t_{табл}$

Целые и десятые доли *t*	Сотые доли *t*									
	0	1	2	3	4	5	6	7	8	9
0,0	0,0000	0,0080	0,0160	0,0239	0,0319	0,0399	0,0478	0,0558	0,0638	0,0718
0,1	0797	0876	0955	1034	1114	1192	1271	1350	1428	1507
0,2	1585	1663	1741	1819	1897	1974	2051	2128	2205	2282
0,3	2358	2434	2510	2586	2661	2737	2812	2886	2961	3035
0,4	3108	3182	3255	3328	3401	3473	3545	3616	3688	3752
0,5	3829	3899	3969	4039	4108	4177	4245	4313	4381	4448
0,6	4515	4581	4647	4713	4778	4843	4909	4971	5035	5098
0,7	5161	5223	5285	5346	5467	5497	5527	5587	5646	5705
0,8	5763	5821	5878	5935	5991	6047	6102	6157	6211	6265
0,9	6319	6372	6424	6476	6528	6579	6629	6679	6729	6778

Продолжение

Целые и десятые доли *t*	Сотые доли *t*									
	0	1	2	3	4	5	6	7	8	9
1,0	0,6817	0,6875	0,6923	0,6970	0,7017	0,7063	0,7109	0,7154	0,7199	0,7243
1,1	7287	7330	7373	7415	7457	7499	7540	7580	7620	7660
1,2	7699	7737	7775	7813	7850	7887	7923	7959	7995	8030
1,3	8064	8098	8132	8165	8198	8230	8262	8293	8324	8355
1,4	8385	8415	8444	8473	8501	8529	8557	8584	8611	8638
1,5	8664	8690	8715	8740	8764	8788	8812	8836	8859	8882
1,6	8904	8926	8948	8969	8990	9011	9031	9051	9070	9089
1,7	9108	9127	9146	9164	9182	9199	9216	9233	9249	9265
1,8	9281	9297	9312	9327	9342	9357	9317	9385	9399	9412
1,9	9425	9438	9451	9464	9476	9488	9500	9512	9523	9534
2,0	0,9545	0,9556	0,9566	0,9576	0,9586	0,9596	0,9606	0,9616	0,9625	0,9634
2,1	9643	9651	9660	9668	9676	9684	9692	9700	9707	9715
2,2	9722	9729	9736	9743	9749	9756	9762	9768	9774	9780
2,3	9785	9791	9797	9802	9807	9812	9817	9822	9827	9832
2,4	9836	9841	9845	9849	9853	9857	9861	9865	9869	9872
2,5	9876	9879	9883	9886	9889	9892	9895	9898	9901	9904

2,6	9907	9910	9912	9915	9917	9920	9922	9924	9926	9928
2,7	9931	9933	9935	9937	9939	9940	9942	9944	9946	9947
2,8	9949	9951	9952	9953	9955	9956	9958	9959	9960	9961
2,9	9963	9964	9965	9966	9967	9968	9969	9970	9971	9972
3,0	0,9973	0,9974	0,9975	0,9976	0,9976	0,9977	0,9978	0,9979	0,9979	0,9980
3,1	9981	9981	9982	9983	9983	9984	9984	9985	9985	9986
3,5	9995	9996	9996	9996	9996	9996	9996	9996	9997	9997
3,6	9997	9997	9997	9997	9997	9997	9997	9998	9998	9998
3,7	9998	9998	9998	9998	9998	9998	9998	9998	9998	9998
3,8	9999	9999	9999	9999	9999	9999	9999	9999	9999	9999
3,9	9999	9999	9999	9999	9999	9999	9999	9999	9999	9999
4,0	0,999936	9999	9999	9999	9999	9999	9999	9999	9999	9999
4,5	0,999994	–	–	–	–	–	–	–	–	–
5,0	0,99999994	–	–	–	–	–	–	–	–	–

Распределение Стьюдента (*t*-распределение)

υ	Вероятность $\alpha = S_t(t) = P(\lvert T \rvert > t_{\text{табл}})$												
	0,9	0,8	0,7	0,6	0,5	0,4	0,3	0,2	0,1	0,05	0,02	0,01	0,001
1	0,158	0,325	0,510	0,727	1,000	1,376	1,963	3,078	6,314	12,706	31,821	63,657	636,619
2	0,142	0,289	0,445	0,617	0,816	1,061	1,386	1,886	2,920	4,303	6,965	9,925	31,598
3	0,137	0,277	0,424	0,584	0,765	0,978	1,250	1,638	2,353	3,182	4,541	5,841	12,941
4	0,134	0,271	0,414	0,569	0,741	0,941	1,190	1,533	2,132	2,776	3,747	4,604	8,610
5	0,132	0,267	0,408	0,559	0,727	0,920	1,156	1,476	2,015	2,571	3,365	4,032	6,859
6	0,131	0,265	0,404	0,553	0,718	0,906	1,134	1,440	1,943	2,447	3,143	3,707	5,959
7	0,130	0,263	0,402	0,549	0,711	0,896	1,119	1,415	1,895	2,365	2,998	3,499	5,405
8	0,130	0,262	0,399	0,546	0,706	0,889	1,108	1,397	1,860	2,306	2,896	3,355	5,041
9	0,129	0,261	0,398	0,543	0,703	0,883	1,100	1,383	1,833	2,262	2,821	3,250	4,781
10	0,129	0,260	0,327	0,542	0,700	0,879	1,093	1,372	1,812	2,228	2,764	3,169	4,583
11	0,129	0,260	0,396	0,540	0,697	0,876	1,088	1,363	1,796	2,201	2,718	3,106	4,437
12	0,128	0,259	0,395	0,539	0,695	0,873	1,083	1,356	1,782	2,179	2,681	3,055	4,318
13	0,128	0,259	0,394	0,538	0,694	0,870	1,079	1,350	1,771	2,160	2,650	3,012	4,221
14	0,128	0,258	0,393	0,537	0,692	0,868	1,076	1,345	1,761	2,145	2,624	2,977	4,140
15	0,128	0,258	0,393	0,536	0,691	0,866	1,074	1,341	1,753	2,131	2,602	2,947	4,073

16	0,128	0,258	0,392	0,535	0,690	0,865	1,071	1,337	1,746	2,120	2,583	2,921	4,015
17	0,128	0,257	0,392	0,534	0,689	0,863	1,069	1,333	1,740	2,110	2,567	2,898	3,965
18	0,127	0,257	0,392	0,534	0,688	0,862	1,067	1,330	1,734	2,101	2,552	2,878	3,922
19	0,127	0,257	0,391	0,533	0,688	0,861	1,066	1,328	1,729	2,093	2,539	2,861	3,833
20	0,127	0,257	0,391	0,533	0,687	0,860	1,064	1,325	1,725	2,086	2,528	2,845	3,850
21	0,127	0,257	0,391	0,532	0,686	0,859	1,063	1,323	1,721	2,080	2,518	2,831	3,819
22	0,127	0,256	0,390	0,532	0,686	0,858	1,061	1,321	1,717	2,074	2,508	2,819	3,792
23	0,127	0,256	0,390	0,532	0,685	0,868	1,060	1,319	1,714	2,069	2,500	2,807	3,767
24	0,127	0,256	0,390	0,531	0,685	0,857	1,059	1,318	1,711	2,064	2,402	2,797	3,745
25	0,127	0,256	0,390	0,531	0,684	0,856	1,058	1,316	1,708	2,060	2,485	2,787	3,725
26	0,127	0,256	0,390	0,531	0,684	0,856	1,058	1,315	1,706	2,056	2,479	2,779	3,707
27	0,127	0,256	0,389	0,531	0,684	0,855	1,057	1,314	1,703	2,052	2,473	2,771	3,690
28	0,127	0,256	0,389	0,530	0,683	0,855	1,056	1,313	1,701	2,048	2,467	2,763	3,674
29	0,127	0,256	0,389	0,530	0,683	0,854	1,055	1,311	1,699	2,045	2,462	2,756	3,659
30	0,127	0,256	0,389	0,530	0,683	0,854	1,055	1,310	1,697	2,042	2,457	2,750	3,646
40	0,126	0,255	0,388	0,529	0,681	0,851	1,050	1,303	1,684	2,021	2,423	2,704	3,551
60	0,126	0,254	0,387	0,527	0,679	0,848	1,046	1,296	1,671	2,000	2,390	2,660	3,460
120	0,126	0,254	0,386	0,526	0,677	0,845	1,041	1,289	1,658	1,980	2,358	2,617	3,373
∞	0,126	0,253	0,385	0,524	0,674	0,842	1,036	1,282	1,645	1,960	2,326	2,576	3,291

Приложение 3

Распределение Пирсона (χ^2-распределение)

Значения $\chi^2_{табл}$ для вероятностей P ($\chi^2>\chi^2_{табл}$)

υ	Вероятность										
	0,999	0,995	0,99	0,98	0,975	0,95	0,90	0,80	0,75	0,70	0,50
1	0,05157	0,04393	0,03157	0,03628	0,03982	0,00393	0,0158	0,0642	0,102	0,148	0,455
2	0,00200	0,0100	0,0201	0,0404	0,0506	0,103	0,211	0,446	0,575	0,713	1,386
3	0,0243	0,0717	0,115	0,185	0,216	0,352	0,584	1,005	1,213	1,424	2,366
4	0,0908	0,207	0,297	0,429	0,484	0,711	1,064	1,649	1,923	2,195	3,357
5	0,210	0,412	0,554	0,752	0,831	1,145	1,610	2,343	2,675	3,000	4,351
6	0,381	0,676	0,872	1,134	1,237	1,635	2,204	3,070	3,455	3,828	5,348
7	0,598	0,989	1,239	1,564	1,690	2,167	2,833	3,822	4,255	4,671	6,346
8	0,857	1,344	1,646	2,032	2,180	2,733	3,490	4,594	5,071	5,527	7,344
9	1,152	1,735	2,088	2,532	2,700	3,325	4,168	5,380	5,899	6,393	8,343
10	1,479	2,156	2,558	3,059	3,247	3,240	4,865	6,179	6,787	7,267	9,342
11	1,834	2,603	3,053	3,609	3,816	4,575	5,578	6,989	7,584	8,148	10,341
12	2,214	3,074	3,571	4,178	4,404	5,226	6,304	7,807	8,438	9,034	11,340

13	2,617	3,565	4,107	4,765	5,009	5,892	7,042	8,634	9,299	9,926	12,340
14	3,041	4,075	4,660	5,368	5,629	6,571	7,790	9,467	10,165	10,821	13,339
15	3,483	4,601	5,229	5,985	6,262	7,261	8,547	10,307	11,036	11,721	14,339
16	3,942	5,142	5,812	6,614	6,908	7,962	9,312	11,152	11,912	12,624	15,338
17	4,416	5,697	6,408	7,255	7,564	8,672	10,085	12,002	12,892	13,531	16,338
18	4,905	6,265	7,015	7,906	8,231	9,390	10,865	12,857	13,675	14,440	17,338
19	5,407	6,844	7,633	8,567	8,907	10,117	11,651	13,716	14,562	15,352	18,338
20	5,921	7,434	8,260	9,237	9,591	10,871	12,443	14,578	15,452	16,266	19,337
21	6,447	8,034	8,897	9,915	10,283	11,591	13,240	15,445	16,344	17,182	20,337
22	6,983	8,643	9,542	10,600	10,982	12,338	14,041	16,314	17,240	18,101	21,337
23	7,529	9,260	10,196	11,293	11,688	13,091	14,848	17,187	18,137	19,021	22,337
24	8,035	9,886	10,856	11,992	12,401	13,848	15,659	18,062	19,037	19,943	23,337
25	8,649	10,520	11,524	12,697	13,120	14,611	16,173	18,940	19,939	20,887	24,337
26	9,222	11,160	12,198	13,409	13,844	15,379	17,292	19,830	20,843	21,792	25,336
27	9,803	11,808	12,879	14,125	14,573	16,151	18,114	20,703	21,749	22,719	26,136
28	10,391	12,461	13,565	14,847	15,308	16,928	18,937	21,588	22,657	23,617	27,386
29	10,986	13,121	14,256	15,574	16,047	17,708	19,768	22,475	23,567	24,577	28,336
30	11,588	13,787	14,953	16,306	16,791	18,493	20,599	23,364	24,478	25,508	29,336

Продолжение

υ	Вероятность									
	0,30	0,25	0,20	0,10	0,05	0,025	0,02	0,01	0,005	0,001
1	1,074	1,323	1,642	2,706	3,841	5,024	5,412	6,635	7,879	10,827
2	2,408	2,773	3,219	4,605	5,991	7,378	7,824	9,210	10,597	13,815
3	3,665	4,108	4,642	6,251	7,815	9,348	9,837	11,345	12,838	16,268
4	4,878	5,385	5,989	7,779	9,488	11,143	11,668	13,277	14,860	18,465
5	6,064	6,626	7,289	9,236	11,070	12,839	13,388	15,086	16,750	20,517
6	7,231	7,841	8,558	10,645	12,592	14,449	15,033	16,812	18,548	22,457
7	8,383	9,037	9,803	12,017	14,067	16,013	16,622	18,475	20,278	24,322
8	9,524	10,219	11,030	13,362	15,507	17,535	18,168	20,090	21,955	26,125
9	10,656	11,389	12,242	14,684	16,919	19,023	19,679	21,666	23,589	27,877
10	11,781	12,549	13,412	15,987	18,307	20,483	21,161	23,209	25,188	29,588
11	12,899	13,701	14,631	17,275	19,675	21,920	22,618	24,725	26,757	31,264
12	14,011	14,845	15,812	18,549	21,026	23,337	24,054	26,217	28,300	32,909
13	15,119	15,984	16,985	19,812	22,362	24,736	25,472	27,688	29,819	34,528
14	16,222	17,117	18,151	21,064	23,685	26,119	26,873	29,141	31,319	36,123

15	17,322	18,245	19,311	22,307	24,996	27,488	28,259	30,578	32,801	37,697
16	18,418	19,369	20,465	23,542	26,296	28,845	29,633	32,000	34,267	39,252
17	19,511	20,489	21,615	24,769	27,587	30,191	30,995	33,409	35,718	40,790
18	20,601	21,605	22,760	25,989	28,869	31,526	32,346	34,805	37,156	42,312
19	21,689	22,718	23,900	27,204	30,144	32,852	33,687	36,191	38,582	43,820
20	22,775	23,828	25,038	28,412	31,410	34,170	35,020	37,566	39,997	45,315
21	23,858	24,935	26,171	29,615	32,671	35,479	36,343	38,932	41,401	46,797
22	24,939	26,039	27,301	30,813	33,924	36,781	37,659	40,289	42,796	48,268
23	26,018	27,141	28,429	32,007	35,172	38,076	38,968	41,638	44,181	49,728
24	27,096	28,241	29,553	33,196	36,415	39,364	40,270	42,980	45,558	51,170
25	28,172	29,339	30,675	34,382	37,652	40,046	41,566	44,314	46,928	52,620
26	29,246	30,434	31,795	35,563	38,885	41,923	42,856	45,642	48,290	54,052
27	30,319	31,528	32,912	36,741	40,113	43,194	44,140	46,963	49,645	55,476
28	31,391	32,620	34,027	37,916	41,337	44,461	45,419	48,278	50,993	56,893
29	32,461	33,711	35,139	39,087	42,557	45,722	46,693	49,588	52,336	58,302
30	33,530	34,800	36,250	40,256	43,773	46,979	47,962	50,892	53,672	59,703

Распределение Фишера–Снедекора (*F*-распределение)

Значения $F_{\text{табл}}$, удовлетворяющие условию $P(F > F_{\text{табл}})$.
Первое значение соответствует вероятности 0,05; второе – вероятности 0,01 и третье – вероятности 0,001; число степеней свободы v_1 – числителя; v_2 – знаменателя

υ_2 \ υ_1	1	2	3	4	5	6	8	12	24	∞	*t*
1	161,4 4052 406523	199,5 4999 500016	215,7 5403 536700	224,6 5625 562527	230,2 5764 576449	234,0 5859 585953	238,9 5981 598149	243,9 6106 610598	249,0 6234 623432	253,3 6366 636535	12,71 63,66 636,2
2	18,51 98,49 998,46	19,00 99,01 999,00	19,16 99,17 999,20	19,25 99,25 999,20	19,30 99,30 999,20	19,33 99,33 999,20	19,37 99,36 999,40	19,41 99,42 999,60	19,45 99,46 999,40	19,50 99,50 999,40	4,30 9,92 31,00
3	10,13 34,12 67,47	9,55 30,81 148,51	9,28 29,46 141,10	9,12 28,71 137,10	9,01 28,24 134,60	8,94 27,91 132,90	8,84 27,49 130,60	8,74 27,05 128,30	8,64 26,60 125,90	8,53 26,12 123,50	3,18 5,84 12,94
4	7,71 21,20 74,13	6,94 18,00 61,24	6,59 16,69 56,18	6,39 15,98 5343	6,26 15,52 51,71	6,16 15,21 50,52	6,04 14,80 49,00	5,91 14,37 47,41	5,77 13,93 45,77	5,63 13,46 44,05	2,78 4,60 8,61
5	6,61 16,26 47,04	5,79 13,27 36,61	5,41 12,06 33,20	5,19 11,39 31,09	5,05 10,97 20,75	4,95 10,67 28,83	4,82 10,27 27,64	4,68 9,89 26,42	4,53 9,47 25,14	4,36 9,02 23,78	2,57 4,03 6,86

6	5,99 13,74 35,51	5,14 10,92 26,99	4,76 9,78 23,70	4,53 9,15 21,90	4,39 8,75 20,81	4,28 8,47 20,03	4,15 8,10 19,03	4,00 7,72 17,99	3,84 7,31 16,89	3,67 6,88 15,75	2,45 3,71 5,96
7	5,59 12,25 29,22	4,74 9,55 21,69	4,35 8,45 18,77	4,12 7,85 17,19	3,97 7,46 16,21	3,87 7,19 15,52	3,73 6,84 14,63	3,57 6,47 13,71	3,41 6,07 12,73	3,23 5,65 11,70	2,36 3,50 5,40
8	5,32 11,26 25,42	4,46 8,65 18,49	4,07 7,59 15,83	3,84 7,10 14,39	3,69 6,63 13,49	3,58 6,37 12,86	3,44 6,03 12,04	3,28 5,67 11,19	3,12 5,28 10,30	2,99 4,86 9,35	2,31 3,36 5,04
9	5,12 10,56 22,86	4,26 8,02 16,39	3,86 6,99 13,90	3,63 6,42 12,56	3,48 6,06 11,71	3,37 5,80 11,13	3,23 5,47 10,37	3,07 5,11 9,57	2,90 4,73 8,72	2,71 4,31 7,81	2,26 3,25 4,78
10	4,96 10,04 21,04	4,10 7,56 14,91	3,71 6,55 12,55	3,48 5,99 11,28	3,33 5,64 10,48	3,22 5,39 9,92	3,07 5,06 9,20	2,91 4,71 8,45	2,74 4,33 7,64	2,54 3,91 6,77	2,23 3,17 4,59
11	4,84 9,65 19,69	3,98 7,20 13,81	3,59 6,22 11,56	3,36 5,67 10,35	3,20 5,32 9,58	3,09 5,07 9,05	2,95 4,74 8,35	2,79 4,40 7,62	2,61 4,02 6,85	2,40 3,60 6,00	2,20 3,11 4,49
12	4,75 9,33 18,64	3,88 6,93 12,98	3,49 5,95 10,81	3,26 5,41 9,63	3,11 5,06 8,89	3,00 4,82 8,38	2,85 4,50 7,71	2,69 4,16 7,00	2,50 3,78 6,25	2,30 3,36 5,42	2,18 3,06 4,32
13	4,67 9,07 17,81	3,80 6,70 12,31	3,41 5,74 10,21	3,18 5,20 9,07	3,02 4,86 8,35	2,92 4,62 7,86	2,77 4,30 7,21	2,60 3,96 6,52	2,42 3,59 5,78	2,21 3,16 4,97	2,16 3,01 4,12

Продолжение

υ_2 \ υ_1	1	2	3	4	5	6	8	12	24	∞	*t*
14	4,60 8,86 17,14	3,74 6,51 11,78	3,34 5,56 9,73	3,11 5,03 8,62	2,96 4,69 7,92	2,85 4,46 7,44	2,70 4,14 6,80	2,53 3,80 6,13	2,35 3,43 5,41	2,13 3,00 4,60	2,14 2,98 4,14
15	4,45 8,68 16,59	3,68 6,36 11,34	3,29 5,42 9,34	3,06 4,89 8,25	2,90 4,56 7,57	2,79 4,32 7,09	2,64 4,00 6,47	2,48 3,67 5,81	2,29 3,29 5,10	2,07 2,87 4,31	2,13 2,95 4,07
16	4,41 8,53 16,12	3,63 6,23 10,97	3,24 5,29 9,01	3,01 4,77 7,94	2,85 4,44 7,27	2,74 4,20 6,80	2,59 3,89 6,20	2,42 3,55 5,55	2,24 3,18 4,85	2,01 2,75 4,06	2,12 2,92 4,02
17	4,45 8,40 15,72	3,59 6,11 10,66	3,20 5,18 8,73	2,96 4,67 7,68	2,81 4,34 7,02	2,70 4,10 6,56	2,55 3,79 5,96	2,38 3,45 5,32	2,19 3,08 4,63	1,96 2,65 3,85	2,11 2,90 3,96
18	4,41 8,28 15,38	3,55 6,01 10,39	3,16 5,09 8,49	2,93 4,58 7,46	2,77 4,25 6,81	2,66 4,01 6,35	2,51 3,71 5,76	2,34 3,37 5,13	2,15 3,01 4,45	1,92 2,57 3,67	2,10 2,88 3,92
19	4,38 8,18 15,08	3,52 5,93 10,16	3,13 5,01 8,28	2,90 4,50 7,26	2,74 4,17 6,61	2,63 3,94 6,18	2,48 3,63 5,59	2,31 3,30 4,97	2,11 2,92 4,29	1,88 2,49 3,52	2,09 2,86 3,88
20	4,35 8,10 14,82	3,49 5,85 9,95	3,10 4,94 8,10	2,87 4,43 7,10	2,71 4,10 6,46	2,60 3,87 6,02	2,45 3,56 5,44	2,28 3,23 4,82	2,08 2,86 4,15	1,84 2,42 3,38	2,09 2,84 3,85

21	4,32 8,02 14,62	3,47 5,78 9,77	3,07 4,87 7,94	2,84 4,37 6,95	2,68 4,04 6,32	2,57 3,81 5,88	2,42 3,51 5,31	2,25 3,17 4,70	2,05 2,80 4,03	1,82 2,36 3,26	2,08 2,83 3,82
22	4,30 7,94 14,38	3,44 5,72 9,61	3,05 4,82 7,80	2,82 4,31 6,87	2,66 3,99 6,19	2,55 3,75 5,76	2,40 3,45 5,19	2,23 3,12 4,58	2,03 2,75 3,92	1,78 2,30 3,15	2,07 2,82 3,79
23	4,28 7,88 14,19	3,42 5,66 9,46	3,03 4,76 7,67	2,80 4,26 6,70	2,64 3,94 6,08	2,53 3,71 5,56	2,38 3,41 5,09	2,20 3,07 4,48	2,00 2,70 3,82	1,76 2,26 3,05	2,07 2,81 3,77
24	4,26 7,82 14,03	3,40 5,61 9,34	3,01 4,72 7,55	2,78 4,22 6,59	2,62 3,90 5,98	2,51 3,67 5,55	2,36 3,36 4,99	2,18 3,03 4,39	1,98 2,66 3,84	1,73 2,21 2,97	2,06 2,80 3,75
25	4,24 7,77 13,88	3,38 5,57 9,22	2,99 4,68 7,45	2,76 4,18 6,49	2,60 3,86 5,89	2,49 3,63 5,46	2,34 3,32 4,91	2,16 2,99 4,31	1,96 2,62 3,66	1,71 2,17 2,89	2,06 2,79 3,72
26	4,22 7,72 13,74	3,37 5,53 9,12	2,98 4,64 7,36	2,74 4,14 6,41	2,59 3,82 5,80	2,47 3,59 5,38	2,32 3,29 4,83	2,15 2,96 4,24	1,95 2,58 3,59	1,69 2,13 2,82	2,06 2,78 3,71
27	4,21 7,68 13,61	3,35 5,49 9,02	2,96 4,60 7,27	2,73 4,11 6,33	2,57 3,78 5,73	2,46 3,56 5,31	2,30 3,26 4,76	2,13 2,93 4,17	1,93 2,55 3,52	1,67 2,10 2,76	2,05 2,77 3,69
28	4,19 7,64 13,50	3,34 5,45 8,93	2,95 4,57 7,18	2,71 4,07 6,25	2,56 3,75 5,66	2,44 3,53 5,24	2,29 3,23 4,69	2,12 2,90 4,11	1,91 2,52 3,46	1,65 2,06 2,70	2,05 2,76 3,67

Продолжение

υ_2 \ υ_1	1	2	3	4	5	6	8	12	24	∞	t
29	4,18 7,60 13,39	3,33 5,42 8,85	2,93 4,54 7,12	2,70 4,04 6,19	2,54 3,73 5,59	2,43 3,50 5,18	2,28 3,20 4,65	2,10 2,87 4,05	1,90 2,49 3,41	1,64 2,03 2,64	2,05 2,76 3,66
30	4,17 7,56 13,29	3,32 5,39 8,77	2,92 4,51 7,05	2,69 4,02 6,12	2,53 3,70 5,53	2,42 3,47 5,12	2,27 3,17 4,58	2,09 2,84 4,00	1,89 2,47 3,36	1,62 2,01 2,59	2,04 2,75 3,64
60	4,00 7,08 11,97	3,15 4,98 7,76	2,76 4,13 6,17	2,52 3,65 5,31	2,37 3,34 4,76	2,25 3,12 4,37	2,10 2,82 3,87	1,92 2,50 3,31	1,70 2,12 2,76	1,39 1,60 1,90	2,00 2,66 3,36
∞	3,84 6,64 10,83	2,99 4,60 6,91	2,60 3,78 5,42	2,37 3,32 4,62	2,21 3,02 4,10	2,09 2,80 3,74	1,94 2,51 3,27	1,75 2,18 2,74	1,52 1,79 2,13	1,03 1,04 1,05	1,96 2,58 3,29

Таблица Фишера–Иейтса

Значения $r_{кр}$, найденные для уровня значимости α и чисел степеней свободы $\upsilon = n - 2$ в случае парной корреляции и $\upsilon = n - 1 - 2$, где 1 – число исключенных величин в случае частной корреляции

υ	Двусторонние границы				υ	Двусторонние границы			
	0,05	0,02	0,01	0,001		0,05	0,02	0,01	0,001
1	0,997	1,000	1,000	1,000	16	0,468	0,543	0,590	0,708
2	0,950	0,980	0,990	0,999	17	0,456	0,529	0,575	0,693
3	0,878	0,934	0,959	0,991	18	0,444	0,516	0,561	0,679
4	0,811	0,882	0,917	0,974	19	0,433	0,503	0,549	0,665
5	0,754	0,833	0,875	0,951	20	0,423	0,492	0,537	0,652
6	0,707	0,789	0,834	0,925	25	0,381	0,445	0,487	0,597
7	0,666	0,750	0,798	0,898	30	0,349	0,409	0,449	0,554
8	0,632	0,715	0,765	0,872	35	0,325	0,381	0,418	0,519
9	0,602	0,685	0,735	0,847	40	0,304	0,358	0,393	0,490
10	0,576	0,658	0,708	0,823	45	0,288	0,338	0,372	0,465
11	0,553	0,634	0,684	0,801	50	0,273	0,322	0,354	0,443
12	0,532	0,612	0,661	0,780	60	0,250	0,295	0,325	0,408
13	0,514	0,592	0,641	0,760	70	0,232	0,274	0,302	0,380
14	0,497	0,574	0,623	0,742	80	0,217	0,257	0,283	0,338
15	0,482	0,558	0,606	0,725	90	0,205	0,242	0,267	0,338
					100	0,195	0,230	0,254	0,321
υ	0,025	0,01	0,005	0,0005	υ	0,025	0,01	0,005	0,0005
	Односторонние границы					Односторонние границы			

Таблица Z-преобразования Фишера $Z = \frac{1}{2}\{\ln(1+r) - \ln(1-r)\}$

r	0	1	2	3	4	5	6	7	8	9
0,0	0,0000	0,0101	0,0200	0,0300	0,0400	0,0501	0,0601	0,0701	0,0802	0,0902
1	0,1003	0,1104	0,1206	0,1308	0,1409	0,1511	0,1614	0,1717	0,1820	0,1923
2	0,2027	0,2132	0,2237	0,2342	0,2448	0,2554	0,2661	0,2769	0,2877	0,2986
3	0,3095	0,3205	0,3316	0,3428	0,3541	0,3654	0,3767	0,3884	0,4001	0,4118
4	0,4236	0,4356	0,4477	0,4599	0,4722	0,4847	0,4973	0,5101	0,5230	0,5361
5	0,5493	0,5627	0,5764	0,5901	0,6042	0,6184	0,6328	0,6475	0,6625	0,6777
6	0,6932	0,7089	0,7250	0,7414	0,7582	0,7753	0,7928	0,8107	0,8291	0,8480
7	0,8673	0,8872	0,9077	0,9287	0,9505	0,9730	0,9962	1,0203	1,0454	1,0714
8	1,0986	1,1270	1,1568	1,1881	1,2212	1,2562	1,2933	1,3331	1,3758	1,4219
9	1,4722	1,5275	1,5890	1,6584	1,7381	1,8318	1,9459	2,0923	2,2976	2,6467
0,99	2,6466	2,6996	2,7587	2,8257	2,9031	2,9945	3,1063	3,2504	3,4534	3,8002

Приложение 7

Критерий А. Н. Колмогорова

Точные и асимптотические границы для верхней грани модуля разности истинной и эмпирической функции распределения

n	Уровень значимости 0,05			Уровень значимости 0,01		
	точная граница	асимптотическая граница	отношение	точная граница	асимптотическая граница	отношение
5	0,5633	0,6074	1,078	0,6685	0,7279	1,089
10	0,4087	0,4295	1,051	0,4864	0,5147	1,058
15	0,3375	0,3507	1,039	0,4042	0,4202	1,040
20	0,2939	0,3037	1,033	0,3524	0,3639	1,033
25	0,2639	0,2716	1,029	0,3165	0,3255	1,028
30	0,2417	0,2480	1,026	0,2898	0,2972	1,025
40	0,2101	0,2147	1,022	0,2521	0,2574	1,021
50	0,1884	0,1921	1,019	0,2260	0,2302	1,018
60	0,1723	0,1753	1,018	0,2067	0,2101	1,016
70	0,1597	0,1623	1,016	0,1917	0,1945	1,015
80	0,1496	0,1518	1,015	0,1795	0,1820	1,014
90	0,1412	0,1432	1,014			
100	0,1340	0,1358	1,013			

При $n > 100$ следует принять асимптотические границы $\bar{\varepsilon}_{0,05} = \frac{1,36}{\sqrt{n}}$ и $\bar{\varepsilon}_{0,01} = \frac{1,63}{\sqrt{n}}$, для которых истинные коэффициенты доверия несколько больше заданных величин 0,95 и 0,99 соответственно.

Значения плотности $f(t)=\frac{1}{\sqrt{2p}}e^{\frac{t^2}{2}}$ вероятности для нормированного нормального закона распределения $f(-t) = f(t)$

Целые и десятые доли *t*	Сотые доли *t*									
	0	1	2	3	4	5	6	7	8	9
0,0	0,3989	0,3989	0,3989	0,3988	0,3986	0,3984	0,3982	0,3980	0,3977	0,3973
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	3726	3712	3697
0,4	3683	3668	3653	3637	3621	3605	3589	3572	3555	3538
0,5	3525	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	2661	2637	2613	2589	2565	2541	2516	2492	2468	2444
1,0	0,2420	0,2396	0,2371	0,2347	0,2323	0,2299	0,2275	0,2251	0,2227	0,2203
1,1	2179	2155	2131	2107	2083	2059	2036	2012	1989	1965

1,2	1942	1919	1895	1972	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	0762	0748	0734	0721	0707	0694	0681	0669
1,9	0656	0644	0632	0620	0608	0596	0584	0573	0562	0551
2,0	0,0540	0,0529	0,0519	0,0508	0,0498	0,0488	0,0478	0,0468	0,0459	0,0449
2,1	0440	0431	0422	0413	0404	0396	0387	0379	0371	0363
2,2	0355	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	0208	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	0,0044	0,0043	0,0042	0,0040	0,0039	0,0038	0,0037	0,0036	0,0035	0,0034
3,1	0033	0032	0031	0030	0029	0028	0027	0026	0025	0025

Продолжение

Целые и десятые доли t	Сотые доли t									
	0	1	2	3	4	5	6	7	8	9
3,2	0024	0023	0022	0022	0021	0020	0020	0019	0018	0018
3,3	0017	0017	0016	0016	0015	0015	0014	0014	0013	0013
3,4	0012	0012	0012	0011	0011	0010	0010	0010	0009	0009
3,5	0009	0008	0008	0008	0008	0007	0007	0007	0007	0006
3,6	0006	0006	0006	0005	0005	0005	0005	0005	0005	0004
3,7	0004	0004	0004	0004	0004	0004	0003	0003	0003	0003
3,8	0003	0003	0003	0003	0003	0002	0002	0002	0002	0002
3,9	0002	0002	0002	0002	0002	0002	0002	0002	0001	0001
4,0	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001
4,1	0,0001338	–	–	–	–	–	–	–	–	–
4,5	0,000160	–	–	–	–	–	–	–	–	–
5,0	0,0000015	–	–	–	–	–	–	–	–	–

Приложение 9

Пяти- и однопроцентные пределы для отношения *G* наибольшей выборочной дисперсии к сумме *I* выборочных дисперсий, полученных из *I* независимых выборок объема *n*

Первое значение соответствует уровню значимости $\alpha = 0{,}05$, а второе – $\alpha = 0{,}01$

I \ $n-1$	1	2	3	4	5	6	7	8	9	10	16	36	144	∞
2	0,998	0,975	0,939	0,906	0,877	0,853	0,833	0,816	0,801	0,788	0,734	0,660	0,518	0,500
	0,999	0,995	0,979	0,959	0,937	0,917	0,809	0,882	0,867	0,854	0,795	0,700	0,606	0,500
3	0,967	0,871	0,798	0,746	0,707	0,677	0,653	0,633	0,617	0,603	.0,547	0,475	0,403	0,333
	0,993	0,942	0,883	0,834	0,903	0,761	0,734	0,711	0,691	0,674	0,606	0,515	0,423	0,333
4	0,906	0,768	0,684	0,629	0,590	0,560	0,537	0,518	0,502	0,488	0,437	0,372	0,309	0,250
	0,968	0,864	0,781	0,721	0,676	0,641	0,613	0,590	0,570	0,554	0,488	0,406	0,325	0,250
5	0,841	0,684	0,598	0,544	0,507	0,478	0,456	0,439	0,424	0,412	0,365	0,307	0,251	0,200
	0,928	0,789	0,696	0,633	0,588	0,553	0,526	0,504	0,485	0,470	0,409	0,335	0,264	0,200
6	0,781	0,616	0,532	0,480	0,445	0,418	0,398	0,382	0,368	0,357	0,314	0,261	0,212	0,167
	0,883	0,722	0,626	0,564	0,520	0,487	0,461	0,440	0,423	0,408	0,353	0,286	0,223	0,167
7	0,727	0,561	0,480	0,431	0,397	0,373	0,354	0,338	0,326	0,315	0,276	0,228	0,183	0,143
	0,838	0,664	0,569	0,508	0,466	0,435	0,411	0,391	0,375	0,362	0,311	0,249	0,193	0,143
8	0,680	0,516	0,438	0,391	0,360	0,336	0,319	0,304	0,293	0,283	0,246	0,202	0,162	0,125
	0,795	0,615	0,521	0,463	0,423	0,393	0,370	0,352	0,337	0,325	0,278	0,221	0,170	0,125
9	0,639	0,478	0,403	0,358	0,329	0,307	0,290	0,277	0,266	0,257	0,223	0,182	0,145	0,111
	0,754	0,573	0,481	0,425	0,387	0,359	0,338	0,321	0,307	0,295	0,251	0,199	0,152	0,111

Приложение 10

Квантили распределения выборочных характеристик эксцесса Ек и асимметрии As

Объем выборки n	Критические значения коэффициента					
	эксцесса Ек при 1–α				асимметрии As при 1–α	
	0,99	0,95	0,05	0,01	0,95	0,99
50	4,92	4,01	2,13	1,95	0,533	0,787
100	40	3,77	35	2,18	389	567
150	14	66	45	30	321	464
200	3,98	57	51	37	280	403
250	87	51	55	42	251	360
300	79	47	59	46	230	329
350	72	44	62	50	213	305
400	67	41	64	52	200	285
450	63	39	66	55	188	269
500	60	37	67	57	179	255
550	57	35	69	58	171	243
600	54	34	70	60	163	233
650	52	33	71	61	157	224
700	50	31	72	62	151	215
750	48	30	73	64	146	208
800	46	29	74	65	142	202
850	45	28	74	66	138	196
900	43	28	75	66	134	190
950	42	27	76	67	130	185
1000	41	26	76	68	127	180
1200	37	24	78	71	116	165
1400	34	22	80	72	107	152
1600	32	21	81	74	100	142
1800	30	20	82	76	095	134
2000	28	18	83	77	090	127
2500	25	16	85	79	080	114
3000	22	15	86	81	073	104
3500	21	14	87	82	068	096
4000	19	13	88	83	064	090
4500	18	12	88	84	060	085
5000	17	12	89	85	057	081

Критические значения корреляционного отношения η^2 и коэффициента детерминации R^2

а) уровень значимости $\alpha = 0,05$

k_2 \ k_1	1	2	3	4	5	6	8	10	20
3	0,771	865	903	924	938	947	959	967	983
4	658	776	832	865	887	902	924	937	967
5	569	699	764	806	835	854	885	904	948
6	500	632	704	751	785	811	847	871	928
7	444	575	651	702	739	768	810	839	908
8	399	527	604	657	697	729	775	807	887
9	362	488	563	628	659	692	742	777	867
10	332	451	527	582	624	659	711	749	847
11	306	420	495	550	593	628	682	722	828
12	283	394	466	521	564	600	655	696	809
14	247	345	417	471	514	550	607	650	773
16	219	312	378	429	477	507	564	609	740
18	197	283	348	394	435	470	527	573	709
20	179	259	318	364	404	432	495	540	680
22	164	238	294	339	377	410	466	511	653
24	151	221	273	316	353	385	440	484	628
26	140	206	256	297	332	363	417	461	605
28	130	193	240	279	314	344	396	439	583
30	122	182	227	264	297	326	373	419	563
32	115	171	214	250	282	310	360	401	544
34	108	162	203	238	268	296	344	384	526
36	102	153	102	226	256	282	329	368	509
38	097	146	184	218	245	271	316	355	493
40	093	139	176	207	234	259	304	342	479
50	075	113	143	170	194	216	254	288	416
60	063	195	121	144	165	184	218	249	368
80	047	072	093	110	127	142	170	196	298
100	038	058	075	090	103	116	140	161	251
120	032	049	063	075	087	098	119	137	217
200	019	030	038	046	053	060	073	086	139
400	010	015	019	023	027	031	038	044	074

Приложение 12

Таблица 5%-го и 1%-го уровней вероятности коэффициентов корреляции (r_a)

Размер выборки	Положительные значения r_a		Отрицательные значения r_a	
	5%-й уровень	1%-й уровень	5%-й уровень	1%-й уровень
5	0,253	0,297	–0,753	–0,798
6	0,354	0,447	–0,708	–0,863
7	0,370	0,510	–0,674	–0,799
8	0,371	0,531	–0,625	–0,764
9	0,366	0,533	–0,593	–0,737
10	0,360	0,525	–0,564	–0,705
11	0,353	0,515	–0,539	–0,679
12	0,348	0,505	–0,516	–0,655
13	0,341	0,495	–0,497	–0,634
14	0,335	0,485	–0,479	–0,615
15	0,328	0,475	–0,462	–0,597
20	0,299	0,432	–0,399	–0,524
25	0,276	0,398	–0,356	–0,473
30	0,257	0,370	–0,324	–0,433
35	0,242	0,347	–0,300	–0,401
40	0,229	0,329	–0,279	–0,376
45	0,218	0,313	–0,262	–0,256
50	0,208	0,301	–0,248	–0,339

Приложение 13

Значения средней μ и стандартных ошибок $\sigma_1\sigma_2$ для n от 10 до 50

n	μ	σ_1	σ_2
10	3,858	1,288	1,964
15	4,636	1,521	2,153
20	5,195	1,677	2,279
25	5,632	1,791	2,373
30	5,990	1,882	2,447
35	6,294	1,956	2,509
40	6,557	2,019	2,561
45	6,790	2,072	2,606
50	6,998	2,121	2,645

Приложение 14

Распределение критерия Дарбина–Уотсона для положительной автокорреляции

(для 5%-го уровня значимости)

n	$V = 1$		$V = 2$		$V = 3$		$V = 4$		$V = 5$	
	d_1	d_2	d_1	d_2	d_1	d_2	d_1	d_2	d_1	d_2
15	1,08	1,36	0,95	1,54	0,82	1,75	0,69	1,97	0,56	2,21
16	1,10	1,37	0,98	1,54	0,86	1,73	0,74	1,93	0,62	2,15
17	1,13	1,38	1,02	1,54	0,90	1,71	0,78	1,90	0,67	2,10
18	1,16	1,39	1,05	1,53	0,93	1,69	0,82	1,87	0,71	2,06
19	1,18	1,40	1,08	1,53	0,97	1,68	0,86	1,85	0,75	2,02
20	1,20	1,41	1,10	1,54	1,00	1,68	0,90	1,83	0,79	1,99
21	1,22	1,42	1,13	1,54	1,03	1,67	0,93	1,81	0,83	1,96
22	1,24	1,43	1,15	1,54	1,05	1,66	0,96	1,80	0,86	1,94
23	1,26	1,44	1,17	1,54	1,08	1,66	0,99	1,79	0,90	1,92
24	1,27	1,45	1,19	1,55	1,10	1,66	1,01	1,78	0,93	1,90
25	1,29	1,45	1,21	1,55	1,12	1,66	1,04	1,77	0,95	1,89
26	1,30	1,46	1,22	1,55	1,14	1,65	1,06	1,76	0,98	1,89
27	1,32	1,47	1,24	1,56	1,16	1,65	1,08	1,76	1,01	1,86
28	1,33	1,48	1,26	1,56	1,18	1,65	1,10	1,75	1,03	1,85
29	1,34	1,48	1,27	1,56	1,20	1,65	1,12	1,74	1,05	1,84
30	1,35	1,49	1,28	1,57	1,21	1,65	1,14	1,74	1,07	1,83
31	1,36	1,50	1,30	1,57	1,23	1,65	1,16	1,74	1,09	1,83
32	1,37	1,50	1,31	1,57	1,24	1,65	1,18	1,73	1,11	1,82
33	1,38	1,51	1,32	1,58	1,26	1,63	1,19	1,73	1,13	1,81
34	1,39	1,51	1,33	1,58	1,27	1,65	1,21	1,73	1,15	1,81
35	1,40	1,52	1,34	1,58	1,28	1,65	1,22	1,73	1,16	1,80
36	1,41	1,52	1,35	1,59	1,29	1,65	1,24	1,73	1,18	1,80
37	1,42	1,53	1,36	1,59	1,31	1,66	1,25	1,72	1,19	1,80
38	1,43	1,54	1,37	1,59	1,32	1,66	1,26	1,72	1,21	1,79
39	1,43	1,54	1,38	1,60	1,33	1,66	1,27	1,72	1,22	1,79

Продолжение

n	$V = 1$		$V = 2$		$V = 3$		$V = 4$		$V = 5$	
	d_1	d_2	d_1	d_2	d_1	d_2	d_1	d_2	d_1	d_2
40	1,44	1,54	1,39	1,60	1,34	1,66	1,29	1,72	1,23	1,79
45	1,48	1,57	1,43	1,62	1,38	1,67	1,34	1,72	1,29	1,78
50	1,50	1,59	1,46	1,63	1,42	1,67	1,38	1,72	1,34	1,77
55	1,53	1,60	1,49	1,64	1,45	1,68	1,41	1,72	1,38	1,77
60	1,55	1,62	1,51	1,65	1,48	1,69	1,44	1,73	1,41	1,77
65	1,57	1,63	1,54	1,66	1,50	1,70	1,47	1,73	1,44	1,77
70	1,58	1,64	1,55	1,67	1,52	1,70	1,49	1,74	1,46	1,77
75	1,60	1,65	1,57	1,68	1,54	1,71	1,51	1,74	1,49	1,77
80	1,61	1,66	1,59	1,69	1,56	1,72	1,53	1,74	1,51	1,77
85	1,62	1,67	1,60	1,70	1,57	1,72	1,55	1,75	1,52	1,77
90	1,63	1,68	1,61	1,70	1,59	1,73	1,57	1,75	1,54	1,78
95	1,64	1,69	1,62	1,71	1,60	1,73	1,58	1,75	1,56	1,78
100	1,65	1,69	1,63	1,72	1,61	1,74	1,59	1,76	1,57	1,78

Приложение 15

Таблица случайных чисел

5489	5583	3156	0835	1988	3912	0938	74609	0869	4420
3522	0935	7877	5665	7020	9555	7379	7124	7878	5544
7555	7579	2550	2487	9477	0864	2349	1012	8250	2633
5759	3554	5080	9074	7001	6249	3224	6868	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	3686	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	6316	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	5647
2042	1192	1175	8851	6432	4635	5757	6656	1660	5389
5470	7702	6958	9080	5925	8519	0127	9233	2452	7341
4504	1730	6005	1704	0345	3275	4738	4862	2556	8333
5850	1257	6163	4439	7276	6353	6912	0731	9033	5294
9083	4260	5277	4998	4298	5204	3965	4028	8936	5148

Продолжение

1762	8713	1189	1090	8989	7273	3213	1935	9321	4820
2023	2589	1740	0424	8294	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	1651	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	7960	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

ОСНОВНЫЕ ФОРМУЛЫ СТАТИСТИКИ

СТАТИСТИЧЕСКАЯ ГРУППИРОВКА

Формула Стерджесса (для определения оптимального числа групп):

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

где n – число групп;
N – число единиц совокупности.

Величина равного интервала:

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

где $R = x_{\max} - x_{\min}$, т. е. размах вариации;
$x_{\max}$ – наибольшее значение варьирующего признака;
$x_{\min}$ – наименьшее значение варьирующего признака.

Величина неравных интервалов:

- изменяющихся в арифметической прогрессии

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

- изменяющихся в геометрической прогрессии

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

где a – константа–число, которое будет положительным при прогрессивно возрастающих интервалах и отрицательным при прогрессивно убывающих интервалах;
q – константа–положительное число, которое при прогрессивно возрастающих интервалах будет больше 1, а при прогрессивно убывающих – меньше 1.

СРЕДНИЕ ВЕЛИЧИНЫ

Общая формула степенной средней:

- простая (невзвешенная)

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

где x_i – i-й вариант осредняемого признака;
n – число вариантов;
$\bar{x}$ – средняя величина признака;

- взвешенная

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

где f – частота (статистический вес) i-го варианта;
k – порядок средней.

При $k = 2$ получается средняя квадратическая ($\bar{x}_j$); при $k = 1$ – средняя арифметическая ($\bar{x}$); при $k = 0$ – средняя геометрическая ($\bar{x}$); при $k = -1$ – средняя гармоническая ($\bar{x}_n$).

Правило мажорантности: $\bar{x}_q > \bar{x} > \bar{x}_q > \bar{x}_n$.

Средняя арифметическая:

- простая (невзвешенная)

$$\bar{x} = \frac{\sum x_i}{n};$$

- взвешенная

$$\bar{x} = \frac{\sum x_i f_i}{\sum f_i}.$$

Средняя гармоническая:

- простая (невзвешенная)

$$\bar{x} = \frac{n}{\sum \frac{1}{x_i}};$$

- взвешенная

$$\bar{x} = \frac{\sum w_i}{\sum \frac{w_i}{x_i}},$$

где w_i – сложный вес; $w_i = x_i f_i$.

Средняя квадратическая:

- простая (невзвешенная)

$$\bar{x} = \sqrt{\frac{\sum x_i^2}{n}};$$

- взвешенная

$$\bar{x} = \sqrt[k]{\frac{\sum x_i^2 \cdot f_i}{\sum f_i}}.$$

Средняя геометрическая:

- простая (невзвешенная)

$$\bar{x} = \sqrt[n]{\prod x_i};$$

- взвешенная

$$\bar{x} = \sqrt[\sum f]{\prod (x_i)^{f_i}}.$$

СТРУКТУРНЫЕ (НЕПАРАМЕТРИЧЕСКИЕ) СРЕДНИЕ

Мода интервального ряда распределения:

$$\mathrm{Mo} = x_{\mathrm{o}} + i \cdot \frac{f_{\mathrm{Mo}} - f_{\mathrm{Mo}-1}}{(f_{\mathrm{Mo}} - f_{\mathrm{Mo}-1}) + (f_{\mathrm{Mo}} - f_{\mathrm{Mo}+1})},$$

где x_0 – нижняя граница модального интервала;
i – величина модального интервала;
f_{Mo} – частота модального интервала;
$f_{\mathrm{Mo}-1}$ – частота интервала, предшествующего модальному;
$f_{\mathrm{Mo}+1}$ – частота интервала, следующего за модальным.

Медиана интервального ряда распределения:

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

где x_0 – нижняя граница медианного интервала;
Σf – сумма частот;
$S_{\mathrm{Me}-1}$ – накопленная частота интервала, предшествующего медианному;
f_{Me} – частота медианного интервала.

Квартили:

первый

$$Q_1 = x_{Q1} + i \cdot \frac{\frac{1}{4}\sum f - S_{Q_1-1}}{fQ_1};$$

третий

$$Q_3 = x_{Q_3} + i \cdot \frac{\frac{3}{4}\sum f - S_{Q_3-1}}{fQ_3},$$

где x_{Q_1} – нижняя граница интервала, содержащего нижний квартиль;
x_{Q_3} – нижняя граница интервала, содержащего верхний квартиль;

S_{Q_1-1} – накопленная частота интервала, предшествующего интервалу,содержащему нижний квартиль;

S_{Q_3-1} – то же, для верхнего квартиля;

f_{Q1} – частота интервала, содержащего нижний квартиль;

f_{Q3} – то же, для верхнего квартиля.

Децили:

первый

$$d_1 = x_{d_1} + i \cdot \frac{\frac{1}{10}\sum f - S_{d_1-1}}{f_{d_1}};$$

второй

$$d_2 = x_{d_2} + i \cdot \frac{\frac{2}{10}\sum f - S_{d_2-1}}{f_{d_2}},$$

где x_{d1} – нижняя граница интервала, содержащего первый дециль;

x_{d2} – то же, для второго дециля;

i – величина интервала, содержащего первый (второй) дециль;

S_{d_1-1} и S_{d_2-1} – накопленные частоты интервалов, предшествующих интервалу, содержащему первый (второй) дециль;

f_{d1} и f_{d2} – частоты интервалов, содержащих первый (второй) дециль.

Остальные децили вычисляют по аналогичной схеме.

ПОКАЗАТЕЛИ ВАРИАЦИИ

Размах вариации:

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

Среднее линейное отклонение:

- невзвешенное

$$\bar{d} = \frac{\sum\left|x_i - \bar{x}\right|}{n};$$

- взвешенное

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

где $\left| x_i - \bar{x} \right|$ – абсолютное значение отклонений.

Дисперсия σ^2 – средний квадрат отклонений.

Дисперсия вариационного признака:

- невзвешенная

$$\text{у}^2 = \frac{\sum \left(x_i - \bar{x} \right)^2}{n};$$

- взвешенная

$$\text{у}^2 = \frac{\sum \left(x_i - \bar{x} \right)^2 f_i}{\sum f_i}.$$

Среднее квадратическое отклонение вариационного признака:

- невзвешенное

$$\text{у}^2 = \sqrt{\frac{\sum \left(x_i - \bar{x} \right)^2}{n}};$$

- взвешенное

$$\text{у}^2 = \sqrt{\frac{\sum \left(x_i - \bar{x} \right)^2 f_i}{\sum f_i}}.$$

Дисперсия альтернативного признака:

$$\text{у}_p^2 = pq,$$

где p – доля единиц в совокупности, обладающих данным признаком;
q – доля единиц, не обладающих данным признаком.

Среднее квадратическоc отклонение альтернативного признака:

$$\text{у}_p = \sqrt{pq}.$$

Коэффициент осцилляции:

$$V_R = \frac{R}{\overline{x}} \cdot 100\%.$$

Линейный коэффициент вариации:

$$V_{\overline{d}} = \frac{\overline{d}}{\overline{x}} \cdot 100\%.$$

Коэффициент вариации:

$$V_{\text{у}} = \frac{\text{у}}{\overline{x}} \cdot 100\%.$$

Показатель энтропии:

$$H(x) = -\sum p_i \cdot \log\ p_i,$$

где p_i – вероятности различных значений случайной величины.

Правило сложения дисперсий для средней величины признака:

$$\text{у}^2 = \overline{\sigma_i^2} + \text{д}_x^2,$$

где σ^2 – общая дисперсия;

$\overline{\sigma_i^2}$ – средняя из внутригрупповых дисперсий;

δ_x^2 – межгрупповая дисперсия.

Средняя из внутригрупповых дисперсий:

$$\overline{\text{у}}_i^2 = \frac{\sum \text{у}_i^2 n_i}{\sum n_i},$$

где σ_i^2 – групповые дисперсии;

n_i – число единиц в группах.

Межгрупповая дисперсия:

$$\text{д}_x^2 = \frac{\sum (\overline{x}_i - \overline{x})^2 n_i}{\sum n_i},$$

где $\overline{x}_i$ – групповые средние;
$\overline{x}$ – общая средняя.

Эмпирический коэффициент детерминации:

$$\text{з}^2 = \frac{\text{д}_x^2}{\text{у}^2}.$$

Эмпирическое корреляционное отношение:

$$\text{з} = \sqrt{\frac{\delta_x^2}{\text{у}^2}}.$$

Правило сложения дисперсий для доли признака:

$$\sigma_{\overline{p}}^2 = \overline{\sigma_{p_i}^2} + \delta_{pi}^2,$$

где $\sigma_{\overline{p}}^2$ – общая дисперсия доли;
$\overline{\sigma_{p_i}^2}$ – средняя из внутригрупповых дисперсий доли;
δ_{pi}^2 – межгрупповая дисперсия доли.

Общая дисперсия доли:

$$\text{у}_{\overline{p}}^2 = \overline{p}(1 - \overline{p}),$$

где $\overline{p}$ – доля изучаемого признака во всей совокупности, определяемая по формуле:

$$\overline{p} = \frac{\sum p_i n_i}{\sum n_i} \ (p_i - \text{групповые доли}).$$

Средняя из внутригрупповых дисперсий доли:

$$\overline{\sigma^2_{\bar{p}_i}} = \overline{p_i(1-p_i)} = \frac{\sum p_i(1-p_i)n_i}{\sum n_i}.$$

Межгрупповая дисперсия доли:

$$\delta^2_{p_i} = \frac{\sum (p_i - \bar{p})^2 n_i}{\sum n_i}.$$

Коэффициент асимметрии:

$$As = \frac{\mu_3}{\sigma^3},$$

где μ_3 – центральный момент 3-го порядка, определяемый по формуле:

$$\mu_3 = \frac{\sum (x_i - \bar{x})^3 f_i}{\sum f_i}.$$

Эксцесс:

$$E\kappa = \frac{\mu_4}{\sigma^4} - 3,$$

где μ_4 – центральный момент четвертого порядка, определяемый по формуле:

$$\mu_4 = \frac{\sum (x_i - x)^4 f_i}{\sum f_i}.$$

Нормальное распределение:

$$y_t = \frac{1}{\sqrt{2\text{р}}} e^{-\frac{1}{2}t^2},$$

где y_t – ордината кривой нормального распределения;

$t = \frac{x_i - \bar{x}}{\text{у}}$ – стандартизированная (нормированная) величина;

е и π – математические постоянные.

КРИТЕРИИ СОГЛАСИЯ

Пирсона:

$$\text{ч}^2 = \sum \frac{(f_\text{э} - f_\text{т})^2}{f_\text{т}},$$

где $f_\text{э}$ – эмпирические частоты;

$f_\text{т}$ – теоретические частоты.

Романовского:

$$C = \frac{\text{ч}^2 - \text{г}}{\sqrt{2\text{г}}},$$

где γ – число степеней свободы (при проверке гипотезы о нормальности распределения) равно числу групп минус три.

Ястремского:

$$L = \frac{\sum \frac{(f_\text{э} - f_\text{т}\text{Н})^2}{N \cdot pq} - K}{\sqrt{2K + 4Q}},$$

где N – объем совокупности;

pq – дисперсия альтернативного признака;

K – число вариантов или групп;

Q – принимает значение 0,6 при числе вариантов или групп от 8 до 20.

Колмогорова:

$$л = \frac{D}{\sqrt{\sum f}},$$

где D – максимальное значение разности между накопленными эмпирическими и теоретическими частотами;

Σf – сумма эмпирических частот.

ОШИБКА РЕПРЕЗЕНТАТИВНОСТИ
(большая выборка)

Предельная ошибка выборки:

$$\Delta = t\mu,$$

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

μ – средняя ошибка выборки.

Соотношение между генеральной и выборочной дисперсиями:

$$\sigma_{\bar{x}}^2 = \sigma_{\tilde{x}}^2 \cdot \frac{n}{n-1},$$

где $\sigma_{\bar{x}}^2$ – генеральная дисперсия;

$\sigma_{\tilde{x}2}$ – выборочная дисперсия;

n – численность выборки.

Средняя ошибка собственно-случайной выборки:

- повторный выбор

$$м = \sqrt{\frac{у_{\bar{x}}^2}{n}};$$

- бесповторный отбор

$$м = \sqrt{\frac{у_{\tilde{x}}^2}{n}(1 - \frac{n}{N})},$$

где N – численность генеральной совокупности.

Средняя ошибка механической выборки:

$$\text{м} = \sqrt{\frac{\text{у}_{\tilde{x}}^2}{n}\left(1 - \frac{n}{N}\right)}.$$

Средняя ошибка типической выборки:

при отборе, пропорциональном объему типических групп

- повторный отбор

$$\text{м} = \sqrt{\frac{\overline{\text{у}_i^2}}{n}},$$

где $\overline{\sigma_i^2}$ – средняя из внутригрупповых дисперсий;

- бесповторный отбор

$$\text{м} = \sqrt{\frac{\overline{\sigma_i^2}}{n}\left(1 - \frac{n}{N}\right)};$$

при отборе, пропорциональном дифференциации признака

- повторный отбор

$$\text{м} = \frac{1}{N}\sqrt{\sum \frac{\text{у}_i^2 N_i^2}{n_i}},$$

где σ_i^2 – дисперсия признака в i-й группе;
N_i – численность генеральной совокупности в i-й группе;
n_i – численность выборки в i-й группе;

- бесповторный отбор

$$\text{м} = \frac{1}{N}\sqrt{\sum \frac{\text{у}_i^2 N_i^2}{n_i}\left(1 - \frac{n_i}{N_i}\right)}.$$

Средняя ошибка серийной выборки:

- повторный отбор

$$\text{м} = \sqrt{\frac{\text{д}^2}{r}},$$

где δ_i^2 – межгрупповая (межсерийная) дисперсия;
r – число отобранных серий;

- бесповторный отбор

$$\text{м} = \sqrt{\frac{\text{д}_i^2}{r}(1 - \frac{r}{R})},$$

где R – число серий в генеральной совокупности.

Средняя ошибка малой выборки:

$$\text{м}_{\text{мв}} = \sqrt{\frac{\text{у}_{\text{мв}}^2}{n}},$$

где $\sigma^2_{\text{МВ}}$ – выборочная дисперсия (в малой выборке);
n – число отобранных в малую выборку единиц. При этом

$$\text{у}_{\text{мв}}^2 = \frac{\sum (x - \tilde{x})^2}{n - 1}.$$

НЕОБХОДИМЫЙ ОБЪЕМ ВЫБОРКИ ПРИ ОПРЕДЕЛЕНИИ СРЕДНЕГО ЗНАЧЕНИЯ ПРИЗНАКА

Необходимый объем собственно-случайной и механической выборки:

- повторная

$$n = \frac{t^2 \text{у}_{\tilde{x}}^2}{\text{Д}_{\tilde{x}}^2};$$

- бесповторная

$$n = \frac{t^2 \text{у}^2_{\tilde{x}} N}{\text{Д}^2_{\tilde{x}} N + t^2 \text{у}^2_{\tilde{x}}}.$$

Необходимый объем типической выборки:

при отборе, пропорциональном объему типических групп:

- повторная

$$n = \frac{t^2 \overline{\text{у}}^2_{\tilde{x}}}{\text{Д}^2_{\tilde{x}}};$$

- бесповторная

$$n = \frac{t^2 \overline{\text{у}}^2_{\tilde{x}} N}{\text{Д}^2_{\tilde{x}} N + t^2 \overline{\text{у}}^2_{\tilde{x}}};$$

при отборе, пропорциональном дифференциации признака:

$$n_i = \frac{n N_i \text{у}_i}{\sum N_i \text{у}_i}.$$

Необходимый объем серийной выборки:

- повторная

$$r = \frac{t^2 \text{д}^2_{\tilde{x}}}{\text{Д}^2_{x}};$$

- бесповторная

$$r = \frac{t^2 \text{д}^2_{\tilde{x}} R}{\text{Д}^2_{\tilde{x}} R + t^2 \text{д}^2_{\tilde{x}}}.$$

ПОКАЗАТЕЛИ АНАЛИЗА РЯДА ДИНАМИКИ

Темп роста:

- базисный

$$T_p = \frac{y_i}{y_1} \cdot 100;$$

- цепной

$$T_p = \frac{y_i}{y_{i-1}} \cdot 100,$$

где y_i – порядковый уровень ряда динамики;
y_1 – базисный уровень ряда динамики.

Абсолютный прирост:

базисный

$$\Delta_y = y_i - y_1.$$

Средний абсолютный прирост:

- по базисному абсолютному приросту

$$\bar{\Delta}_y = \frac{\Delta_{i/1}}{n-1} = \frac{y_n - y_i}{n-1},$$

где y_n – конечный уровень ряда динамики;
n – число уровней ряда динамики;

- по цепным абсолютным приростам

$$\bar{\Delta}_y = \frac{\sum_{i=1}^{n-1} \Delta_{i/i} - 1}{n-1}.$$

Относительный прирост ($T_{пр}$ – темп прироста):

- базисный

$$T_{пр} = \frac{\Delta_{i/1}}{y_1} = (k_{p_{i/1}} - 1) \cdot 100 = T_{p_{i/1}} - 100;$$

- цепной

$$T_{пр} = \frac{\Delta_{i/i-1}}{y_{i-1}} = \frac{y_i - y_{i-1}}{y_{i-1}} \cdot 100 = (k_{p_{i/i-1}} - 1) \cdot 100 = T_{p_{i/i-1}} - 100.$$

Абсолютное значение 1% прироста:

$$|\%| = \frac{\Delta_{i/i-1}}{T_{пр_{i/i-1}} \cdot \%} = \frac{y_i - y_{i-1}}{\frac{y_i - y_{i-1}}{y_{i-1}} \cdot 100} = \frac{y_{i-1}}{100} = 0,01 \cdot y_{i-1}.$$

СРЕДНИЙ УРОВЕНЬ РЯДА ДИНАМИКИ
($\bar{Y}$ – средняя хронологическая)

В моментном ряду динамики:

- с равноотстоящими уровнями

$$\bar{y} = \frac{\frac{1}{2} y_1 + y_2 + \ldots + y_{n-1} + \frac{1}{2} y_n}{n-1} = \frac{\frac{y_1 + y_n}{2} + \sum_{i=2}^{n-1} y_i}{n-1};$$

- с неравноотстоящими уровнями

$$\bar{y} = \frac{(y_1 + y_2)t_1 + (y_2 + y_3)t_2 + \ldots + (y_{n-1} + y_n)t_{n-1}}{2(t_1 + t_2 + t_3 + \ldots + t_{n-1})} = \frac{\sum (y_i + y_{i+1})t}{2\sum_{t=1}^{n-1} t},$$

где y_i, y_n – уровни ряда динамики;
t_i – длительность интервала времени между уровнями.

В интервальном ряду динамики:

- с равноотстоящими уровнями

$$\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n};$$

- с неравноотстоящими уровнями

$$\bar{y} = \frac{\sum_{i=1}^{n} y_i t_i}{\sum t_1}.$$

Средний темп роста ($\bar{T}_p$ – средняя геометрическая в рядах динамики):

по цепным коэффициентам (темпам) роста

$$\bar{T}_p = \sqrt[m]{k_{p2/1} \cdot k_{p3/2} ... k_{n/n-1}} = \sqrt[m]{\Pi k_{pi/i-1}},$$

где m – число темпов роста.

по абсолютным уровням ряда динамики

$$\bar{T}_\text{р} = \sqrt[n-1]{\frac{y_n}{y_1}};$$

по базисным темпам (коэффициентам) роста

$$\bar{T}_\text{р} = \sqrt[n-1]{T_{\text{р}_{n/1}}};$$

для рядов динамики с неравноотстоящими уровнями

$$\overline{T}_{\text{р}} = \sqrt[\Sigma t]{(k_{p_{2/1}})^{t_1} \cdot (k_{p_{3/2}})^{t_2} ... (k_{n/n-1})^{tn}},$$

где t – интервал времени, в течение которого сохраняется данный темп роста;
Σt – сумма отрезков времени периода.

Средний темп прироста:

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

ПРОВЕРКА ГИПОТЕЗЫ О СУЩЕСТВОВАНИИ ОСНОВНОЙ ТЕНДЕНЦИИ В РЯДАХ ДИНАМИКИ

Метод проверки существенности разности средних:

- t-критерий:

$$t = \frac{\overline{y}_1 - \overline{y}_2}{\text{у} \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}};$$

- среднее квадратическое отклонение разности средних:

$$\text{у} = \sqrt{\frac{(n_1 - 1)^2 \cdot \text{у}_1^2 + (n_2 - 1)^2 \cdot \text{у}_2^2}{n_1 + n_2 - 2}};$$

- дисперсии для первой и второй части ряда динамики:

$$\text{у}_i^2 = \frac{\sum (y_i - \overline{y}_i)^2}{n - 1}, \quad i = 1, 2, ..., n,$$

где $\overline{y}_1$ и $\overline{y}_2$ – средние для первой и второй половины ряда динамики $(\overline{y}_1)$;
n_1 и n_2 – число наблюдений в этих частях ряда;
y_i – фактические уровни ряда динамики;
n – число уровней ряда динамики.

Метод Фостсра-Стюарта:

t-критерий

$$t_s = \frac{S - м}{у_1}; \qquad t_d = \frac{d - 0}{у_2},$$

где

$$S = \sum_{i=2}^{n} S_i; \qquad d = \sum_{i=2}^{n} d_i ,$$

где $S_i = U_i + l_i$; $d_i = U_i - l_i$,

если $y_i > y_{i-1}; y_{i-2}, \ldots, y_1$. то $U_i = 1$, $l = 0$; если $y_i < y_{i-1}$, $y_{i-2}, \ldots, y_1$, *то* $U_i = 0$, $l = 1$; μ – среднее значение величины S, определенное для ряда, в котором уровни расположены случайным образом; σ_1 – стандартная ошибка величины S; σ_2 – стандартная ошибка величины d.

Метод наименьших квадратов при расчете параметров полиномов:

- по прямой

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

где a_0 и a_1 – параметры уравнения;
t – показатели времени.

При $\Sigma t = 0$ параметры a_0 и a_1 определяются:

$$a_0 = \frac{\sum y}{n}; \quad a_1 = \frac{\sum yt}{\sum t^2};$$

- по параболе

$$y_t = a_0 + a_1 t + a_2 t_2.$$

При $\Sigma t = 0$ параметры a_0, a_1 и a_2 определяются:

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

- по показательной кривой

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

При $\Sigma t = 0$ параметры a_0 и a_1 определяются

$$n\lg a_0 = \Sigma \lg y;$$

$$\lg a_1 \Sigma t^2 = \Sigma t \lg y;$$

Ряд Фурье:

$$\bar{y}_t = a_0 + \sum_{t=0}^{m}(a_k \cos kt + b_k \sin kt);$$

$$a_0 = \frac{1}{2}\sum y; \quad a_k = \frac{2}{n}\sum y\cos kt; \quad b_k = \frac{2}{n}\sum y \sin t,$$

где t – время, последовательные значение которого определяются от 0 с увеличением (приростом), равным $2\pi/n$;

n – число уровней ряда динамики.

Первая гармоника ряда Фурье:

$$\bar{y}_t = a_0 + a_1 \cos t + b_1 \sin t;$$

$$a_0 = \frac{\sum y}{12}; \; a_1 = \frac{\sum y \cos t}{6}; \; b_1 = \frac{\sum y \sin t}{6}.$$

Показатели «сезонной волны» (I_s – индекс сезонности):

- в стабильных рядах динамики

$$I_s = \frac{\overline{y}_i}{\overline{y}} \cdot 100\%,$$

где $\overline{y}_i$ – осредненные фактические данные по одноименным периодам;
y – общая (постоянная) средняя;

- в рядах динамики с тенденцией роста (уменьшения)

$$I_s = \frac{\left[\sum \frac{y_i}{\overline{y}_t} \cdot 100\%\right]}{n},$$

где $\overline{y}_t$ – выравненные уровни ряда динамики по уравнению тренда (переменная средняя);
n – число одноименных периодов.

СВЯЗНЫЙ АНАЛИЗ РЯДОВ ДИНАМИКИ

Коэффициент автокорреляции:

$$r_a = \frac{\overline{y_t \cdot y_{t+1}} - \overline{y}_t \cdot \overline{y}_{t+1}}{\sigma_{y_i} \cdot \sigma_{y_{t+1}}} \cdot 100,$$

где σ_{yt} и σ_{yt+1} – средние квадратические отклонения рядов y_t и y_{t+1} соответственно.

Коэффициент корреляции отклонений уровней ряда динамики от тренда:

$$r_{d_x d_y} = \frac{\sum d_x \cdot d_y}{\sqrt{\sum d_x^2 \cdot \sum d_y^2}},$$

где $d_x = x_i - \overline{x}_t$; $d_y = y_i - y_t$.

Коэффициент корреляции последовательных разностей уровней ряда динамики:

$$r_{\Delta_x \Delta_y} = \frac{\sum \Delta_x \cdot \Delta_y}{\sqrt{\Delta_x^1 \cdot \sum \Delta_y^2}},$$

где $\Delta_x = x_i - x_{i-1}$, $\Delta y = y_i - y_{i-1}$.

Экстраполяция тренда:

- по среднему абсолютному приросту

$$\hat{y}_{i+1} = \overline{y}_i + \overline{Д}t,$$

где $\hat{y}_{i+1}$ – экстраполируемый уровень;
$(i+t)$ – номер этого уровня (года);
i – номер последнего уровня (года) исследуемого периода, за который рассчитан $\overline{\Delta}$;
$\overline{\Delta}$ – средний абсолютный прирост;

- по среднему темпу роста

$$\hat{y}_{i+1} = y_i \cdot \overline{k}_p^t,$$

где y_i – последний уровень ряда динамики;
t – срок прогноза;
$\overline{k}_p$ – средний коэффициент роста;

- по аналитическому выражению тренда

$$\hat{y}_t \pm t_\alpha \cdot у_{\overline{y}_t},$$

где $\hat{y}_t$ – расчетное значение уровня ряда по уравнению тренда;
$\sigma_{\overline{y}_t}$ – средняя квадратическая ошибка тренда;
t_α – доверительная величина.

«Абсолютный» прирост удельного веса:

$$\Delta d_i = d_{ij} - d_{ij-1},$$

где d_{ij} – удельный вес i-й структурной части j-й период времени.

Темп роста удельного веса:

$$T_{p_{d_i}} = \frac{d_{ij}}{d_{ij-1}}.$$

Средний «абсолютный» прирост удельного веса:

$$\overline{\Delta} d_i = \frac{d_{in} - d_{i1}}{n-1},$$

где n – число периодов, за которые исследуются структурные сдвиги.

Средний темп роста удельного веса:

$$\overline{T}_{p_{d_i}} = \sqrt[n-1]{\frac{d_{in}}{d_{i1}}}.$$

Средний удельный вес:

$$\overline{d}_1 = \frac{\sum_{j=1}^{n} x_{ij}}{\sum_{j=1}^{n}\sum_{i=1}^{k} x_{ij}} \cdot 100\%,$$

где x_{ij} – абсолютная величина i-й структурной части в j-й период времени;
k – общее число структурных частей в исследуемой совокупности.

Линейный коэффициент «абсолютных» структурных сдвигов:

$$\overline{\Delta}_{d_1-d_0}=\frac{\sum\limits_{i=1}^{k}\left|d_{ij}-d_{ij-1}\right|}{k}.$$

Квадратический коэффициент «абсолютных» структурных сдвигов:

$$у_{d_1-d_0}=\sqrt{\frac{\sum\limits_{i=1}^{k}(d_{ij}-d_{ij-1})^2}{k}}.$$

Квадратический коэффициент «относительных» структурных сдвигов:

$$у_{\frac{d_1}{d_0}}=\sqrt{\sum_{i=1}^{k}\frac{(d_{ij}-d_{ij-1})^2}{d_{ij-1}}\cdot 100}.$$

Линейный коэффициент «абсолютных» структурных сдвигов за *n* периодов:

$$\overline{\Delta}^{(n)}_{d_1-d_0}=\frac{\sum\limits_{i=1}^{k}\left|d_{in}-d_{i1}\right|}{k(n-1)}.$$

Коэффициент Джини:

$$G=1-2\sum_{i=1}^{k}d_{xi}d_{yi}^{H}+\sum_{i=1}^{k}d_{xi}d_{yi},$$

где d_{xi} – доля i-й группы в объеме совокупности;
d_{yi} – доля i-й группы в объеме признака;
d_{yi}^{H} – накопленная доля i-й группы в объеме признака.

Коэффициент Лоренца:

$$L = \frac{\sum_{i=1}^{k} \left| d_{xi} - d_{yi} \right|}{2},$$

где d_{xi} – удельный вес *i*-й группы в объеме совокупности:;
d_{yi} – удельный вес *i*-й группы в объеме признака.

Показатель централизации:

$$I_z = \sum_{i=1}^{n} \left(\frac{x_i}{\sum_{i=1}^{n} x_i} \right)^2,$$

где x_i – значение признака *i*-й единицы совокупности.

ИНДЕКСЫ

Индивидуальные индексы:

- физического объема продукции

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

где q – количество продукции;

- цен

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

где p – цена единицы товара (продукта);

- себестоимости

$$i_z = \frac{z_1}{z_0},$$

где z – себестоимость единицы изделия (продукта);

- производительности труда

$$i_t = \frac{t_0}{t_1},$$

где t – затраты времени на изготовление единицы изделия (продукта);

- стоимости продукции

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

Общие индексы в агрегатной форме:

- стоимости продукции или товарооборота

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

- физического объема

$$I_q = \frac{\sum q_1 p_0}{\sum q_0 p_0};$$

- цен

$$I_p = \frac{\sum p_1 q_1}{\sum p_0 q_1};$$

- себестоимости продукции

$$I_z = \frac{\sum z_1 q_1}{\sum z_0 q_1};$$

- издержек производства

$$I_{zq} = \frac{\sum z_1 q_1}{\sum z_0 q_0};$$

- производительности труда

$$I_t = \frac{\sum t_0 q_1}{\sum t_1 q_1};$$

- затрат времени на производство продукции

$$I_{tq} = \frac{\sum t_1 q_1}{\sum t_0 q_0}.$$

ПРЕОБРАЗОВАНИЕ ФОРМЫ АГРЕГАТНОГО ИНДЕКСА

Средняя арифметическая форма:

- индекс физического объема продукции

$$I_q = \frac{\sum i_q p_0 q_0}{\sum p_0 q_0}, \text{ где } i_q = \frac{q_1}{q_0};$$

- индекс производительности труда

$$I_t = \frac{\sum i_t t_1 q_1}{\sum t_1 q_1} = \frac{\sum i_t T_1}{\sum T_1}, \text{ где } i_t = \frac{t_0}{t_1};$$

- индекс Струмилина

$$I_v = \frac{\sum \left(\frac{q_1}{T_1} : \frac{q_0}{T_0} \right) \cdot T_1}{\sum T_1}.$$

Средняя гармоническая форма:

- индекс себестоимости

$$I_z = \frac{\sum z_1 q_1}{\sum \frac{z_1 q_1}{i_z}}, \text{ где } i_z = \frac{z_1}{z_0};$$

- индекс цен

$$I_p = \frac{\sum p_1 q_1}{\sum \frac{p_1 q_1}{i_p}}, \text{ где } i_p = \frac{p_1}{p_0}.$$

Взаимосвязи индексов:

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

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

$$I_{td} = I_q : I_t = \frac{\sum q_1 t_0}{\sum q_0 t_0} : \frac{\sum t_0 q_1}{\sum t_1 q_1}.$$

Системы агрегатных индексов

индексы стоимости или товарооборота

- цепные индексы

$$\frac{\sum p_1 q_1}{\sum p_0 q_0}; \frac{\sum p_2 q_2}{\sum p_1 q_1}; \ldots; \frac{\sum p_n q_n}{\sum p_{n-1} q_{n-1}};$$

- базисные индексы

$$\frac{\sum p_1 q_1}{\sum p_0 q_0}; \frac{\sum p_2 q_2}{\sum p_0 q_0}; \ldots; \frac{\sum p_n q_n}{\sum p_0 q_0}.$$

Базисные индексы физического объема продукции с постоянными весами (p_0):

$$\frac{\sum q_1 p_0}{\sum q_0 p_0}; \frac{\sum q_2 p_0}{\sum q_0 p_0}; \ldots; \frac{\sum q_n p_0}{\sum q_0 p_0}.$$

Цепные индексы физического объема продукции с постоянными весами (p_0):

$$\frac{\sum q_1 p_0}{\sum q_0 p_0}; \frac{\sum q_2 p_0}{\sum q_1 p_0}; \ldots; \frac{\sum q_n p_0}{\sum q_{n-1} p_0}.$$

Базисные индексы цен с переменными весами:

$$\frac{\sum p_1 q_1}{\sum p_0 q_1}; \frac{\sum p_2 q_2}{\sum p_0 q_2}; \ldots; \frac{\sum p_n q_n}{\sum p_0 q_n}.$$

Общая формула индекса переменного состава:

$$I_x = \frac{\bar{x}_1}{\bar{x}_0} = \frac{\sum x_1 w_1}{\sum w_1} : \frac{\sum x_0 w_0}{\sum w_0},$$

где x – индексируемые величины;
w – веса индекса.

Общая формула индекса постоянного состава:

$$I_x = \frac{\sum x_1 w_1}{\sum w_1} : \frac{\sum x_0 w_1}{\sum w_1}.$$

Общая формула индекса изменения структуры (структурных сдвигов):

$$I_{\text{стр}} = \frac{\sum x_0 w_1}{\sum w_1} : \frac{\sum x_0 w_0}{\sum w_0}.$$

«Идеальная» формула Фишера:

$$I_p = \sqrt{\frac{\sum p_1 q_1}{\sum p_0 q_1} \cdot \frac{\sum p_1 q_0}{\sum p_0 q_0}}.$$

Индекс цен Ласпейреса:

$$I_p = \frac{\sum p_1 q_0}{\sum p_0 q_0}.$$

Индекс цен Пааше:

$$I_p = \frac{\sum p_1 q_1}{\sum p_0 q_1}.$$

ИЗМЕРЕНИЕ СВЯЗИ

Однофакторные модели связи *(x)*:

- прямолинейная

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

где a_0, a_1 – коэффициенты регрессии, определяются:

$$\begin{cases} na_0 + a_1 \sum x = \sum y; \\ a_0 \sum x + a_1 \sum x^2 = \sum xy, \end{cases}$$

или по сгрупированным данным

$$\begin{cases} na_0 + a_1 \sum x f_x = \sum y f_y; \\ a_0 \sum x f_x + a_1 \sum x^2 f_x = \sum xy f_y, \end{cases}$$

где $f_x f_y$ – число единиц совокупности согласно распределению соответственно по факторному и результативному признакам;

- по параболе второго порядка

$$\bar{y}_x = a_o + a_1 x + a_2 x^2,$$

где a_0, a_1 и a_2 определяются:

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

- по гиперболе

$$\bar{y}_x = a_0 + a_1 \frac{1}{x},$$

где a_0, a_1 определяются:

$$\begin{cases} na_0 + a_1 \sum \frac{1}{x} = \sum y; \\ a_0 \sum \frac{1}{x} + a_1 \sum \frac{1}{x^2} = \sum \frac{1}{x} y. \end{cases}$$

Многофакторные модели связи (x_1, x_2. …, x_k):

- прямолинейная

$$\bar{y}_{1,2,...,k} = a_0 + a_1 x_1 + a_2 x_2 + ... + a_k x_k;$$

- степенная

$$\bar{y}_{1,2,...,k} = a_0 x_1^{a_1} \ x_2^{a_2} ... x_k^{a_k};$$

- показательная

$$\bar{y}_{1,2,...,k} = e^{a_0 + a_1 x_1 + a_2 x_2 + ... + a_k x_k};$$

- парабола

$$\bar{y}_{1,2,...,k} = a_0 + a_1 x_1^2 \ + a_2 x_2^2 + ... + a_k x_k^2;$$

- гипербола

$$\overline{y}_{1,2,...,k} = a_0 + \frac{a_1}{x_1} + \frac{a_2}{x_2} + ... + \frac{a_k}{x_k};$$

- в стандартизированном масштабе

$$\overline{t}_{1,2,...,k} = \text{в}_1 t_1 + \text{в}_2 t_2 ... + \text{в}_k t_k,$$

где $t_{1,2,...,k}$ – стандартизованные значения признаков $x_{1,2,...,k}$, определяемые по формуле:

$$t_{x_i} = \frac{x_i - \overline{x}_i}{\text{у}_i},$$

где $\overline{t}_{1,2,...,k}$ – среднее значение стандартизованной переменной соответствующего результативного признака, полученного по уравнению регрессии;
$\beta_{1,2,...,k}$ – стандартизованные коэффициенты регрессии.

Оценка существенности связи:

t-критерий Стьюдента

$$t_p = \frac{|a_i|}{\sqrt{\text{у}^2_{a_i}}}, \ t_{\text{кр}} : (\text{б}; \text{н} = n - k - 1),$$

где α – заданный уровень значимости;
ν – число степеней свободы;
n – объем совокупности;
k – число факторных признаков;
$\sigma^2_{a_i}$ – дисперсия коэффициента регрессии, определяется по формуле

$$\text{у}^2_{a_i} = \frac{\text{у}^2_y}{k},$$

где σ^2_y – дисперсия результативного признака.

Проверка значимости уравнения регрессии:

- F – критерий Фишера

$$F_p = \frac{\frac{1}{k+1}\sum \bar{y}_k^2}{\frac{1}{n-k-1}\sum\left(y_i - \bar{y}_k\right)^2},$$

где $\bar{y}_{1,2,\dots,k}$ – теоретические значения результативного признака, полученные по уравнению регрессии;
n – объем совокупности;
k – число факторных признаков в модели;

- средняя ошибка аппроксимации

$$\bar{e} = \frac{1}{n}\sum\frac{\left|y - y_{1,2,\dots,k}\right|}{y}\cdot 100.$$

АНАЛИТИЧЕСКИЕ ПОКАЗАТЕЛИ

Коэффициент эластичности:

$$Э_{x_i} = a_i \cdot \frac{\bar{x}_i}{\bar{y}},$$

где $\bar{x}_i$ – среднее значение соответствующего факторного признака;
$\bar{y}$ – среднее значение результативного признака;
a_i – коэффициент регрессии при соответствующем факторном признаке.

Частный коэффициент детерминации:

$$d_{x_i} = r_{yx_i} \cdot \beta_{x_i},$$

где r_{yxi} – парный коэффициент корреляции между результативным и i-м факторным признаками;
β_{xi} – соответствующий коэффициент уравнения множественной регрессии в стандартизованном масштабе.

***Q*-коэффициент:**

$$Q_{x_i} = Э_{x_i} \cdot V_{x_i}$$

где V_{xi} – коэффициент вариации соответствующего факторного признака.

ПАРАМЕТРИЧЕСКИЕ ПОКАЗАТЕЛИ СВЯЗИ

Линейный коэффициент корреляции:

$$r = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\sigma_x \sigma_y}.$$

Теоретическое корреляционное отношение:

$$\eta = \sqrt{\frac{\sigma^2 - \sigma^2_{\text{ост}}}{\sigma^2}} = \sqrt{1 - \frac{\sigma^2_{\text{ост}}}{\sigma^2}},$$

где σ^2 – общая дисперсия результативного признака;
$\sigma^2_{\text{ост}}$ – остаточная дисперсия, рассчитанная по формуле

$$\sigma^2_{\text{ост.}} = \frac{\sum (y - \bar{y}_x)^2}{n}.$$

Множественный коэффициент корреляции:

$$R_{y/x_1, x_2 \ldots x_k} = \sqrt{1 - \frac{\sigma^2_{\text{ост}}}{\sigma^2}},$$

где $\sigma^2_{\text{ост}}$ – остаточная дисперсия;
σ^2 – общая дисперсия результативного признака.

Совокупный (общий) коэффициент корреляции:

$$R_{y/x_1, x_2} = \sqrt{\frac{r^2_{yx_1} + r^2_{yx_2} - 2 r_{yx_1} r_{yx_2} r_{x_1 x_2}}{1 - r^2_{x_1 x_2}}},$$

где r_{xy} – линейные коэффициенты парной корреляции,

или

$$R_{y/x_1,x_2...x_k} = \sqrt{\text{в}_1 r_{yx_1} + \text{в}_2 r_{yx_2} + ... + \text{в}_k r_{yx_k}},$$

где r_{xy} – парные коэффициенты корреляции;
β_{xi} – коэффициенты регрессии в стандартизованном масштабе.

НЕПАРАМЕТРИЧЕСКИЕ ПОКАЗАТЕЛИ СВЯЗИ

Коэффициент корреляции рангов Спирмена:

- нет связных рангов

$$\text{с}_{x/y} = 1 - \frac{6\sum d_i^2}{n(n^2-1)},$$

где d_i – разность рангов вариантов факторного и результативного признаков;
n – число наблюдений (число пар рангов).

Связные ранги (упрощенная формула):

$$\text{с}_{x/y} = 1 - \frac{\sum_{i=1}^{n} d_i^2}{\frac{1}{6}(n^3-n)-(T_x+T_y)},$$

где $T_{x,y} = \frac{1}{12}\sum_{j=1}^{k}\left(t_j^3 - t_j\right); t_j$ – число одинаковых рангов в j-м ряду,

Коэффициент корреляции рангов Кендалла:

- нет связных рангов

$$\text{ф} = \frac{2S}{n(n-1)},$$

где S – сумма разностей между числом последовательностей и числом инверсий по второму признаку;
n – число наблюдений;

- есть связные ранги

$$ф = \frac{S}{\sqrt{\left[\frac{n(n-1)}{2} - V_x\right] \cdot \left[\frac{n(n-1)}{2} - V_y\right]}},$$

где $V_{x/y} = \frac{1}{2}\sum_{j=1}^{k} t_j\left(t_j - 1\right); t_j$ – число одинаковых рангов *в j*-м ряду.

Коэффициент конкордации:

- нет связных рангов

$$W = \frac{12S}{m^2(n^3 - n)},$$

где m – количество факторов;
n – число наблюдений;
S – отклонение суммы квадратов рангов от средней квадратов рангов;

- есть связные ранги

$$W = \frac{S}{\frac{1}{12}m^2(n^3 - n) - m\sum_{j=1}^{m} T_j},$$

где $T_j = \frac{1}{12}\sum_{j=1}^{m}\left(t_j^3 - t_j\right); t_j$ – количество связных рангов по отдельным показателям.

Коэффициент ассоциации:

$$К_а = \frac{ad - bc}{ad + bc},$$

где a, b, c, d – градации признаков.

Коэффициент контингенции:

$$K_k = \frac{ad - bc}{\sqrt{(a+b)(b+d)(a+c)(c+d)}}.$$

Коэффициент взаимной сопряженности Пирсона:

$$\text{Кп} = \sqrt{\frac{\varphi^2}{1+\varphi^2}},$$

где φ^2 – показатель взаимной сопряженности;
φ – сумма отношений квадратов частот каждой клетки таблицы к произведению итоговых частот соответствующего столбца и строки. Вычитая из этой суммы 1, получаем величину:

$$\varphi^2 = \sum \frac{n_{xy}^2}{n_x \cdot n_y} - 1.$$

Коэффициент взаимной сопряженности Чупрова:

$$K_{\text{ч}} = \sqrt{\frac{\varphi^2}{\sqrt{(k_1 - 1)(k_2 - 1)}}},$$

где k_1 – число значений (групп) первого признака;
k_2 – число значений (групп) второго признака.

Биссериальный коэффициент корреляции:

$$r = \frac{|\bar{y}_2 - \bar{y}_1|}{\text{у}_y} \cdot \frac{pg}{z},$$

где $\bar{y}_1, \bar{y}_2$ – средние в группах;
σ_y – среднее квадратическое отклонение фактических значений признака от среднего уровня;
p – доля первой группы;
g – доля второй группы;
z – табличные значения распределения в зависимости от p.

Частные коэффициенты корреляции:

$$r_{1,2,...,k} = \frac{r_{1,2,...,k-1} - r_{1,k,3,...,k-1} \cdot r_{2,k,3,...,k-1}}{\sqrt{\left(1 - r^2_{1,k,3,...,k-1}\right)\left(1 - r^2_{2,k,3,...,k-1}\right)}};$$

$$r_{yx_1/x_2} = \frac{r_{yx_1} - r_{x_1x_2} \cdot r_{yx_2}}{\sqrt{\left(1 - r^2_{x_2y}\right)\left(1 - r^2_{x_1x_2}\right)}};$$

$$r_{yx_2/x_1} = \frac{r_{yx_1} - r_{x_1x_2} \cdot r_{yx_2}}{\sqrt{\left(1 - r^2_{x_2y}\right)\left(1 - r^2_{x_1x_2}\right)}},$$

где r – парные коэффициенты корреляции между указанными в индексе переменными.

Выявление аномальных наблюдений:

- на основе q-статистики

$$g_t = \frac{\left|y_i - \bar{y}\right|}{\text{у}_y},$$

где y_i – уровни ряда динамики;
y – средний уровень ряда;
σ_y – среднее квадратическое отклонение значений ряда от их среднего уровня;

- методом Ирвина

$$\text{л}_i = \frac{\left|y_i - y_{i-1}\right|}{\text{у}_y},$$

где y_{i-1} – значение предыдущего уровня ряда динамики.

ПРЕДМЕТНЫЙ УКАЗАТЕЛЬ

Учебное издание

Шмойлова Римма Александровна
Минашкин Виталий Григорьевич
Садовникова Наталья Алексеевна
Шувалова Елена Борисовна

ТЕОРИЯ СТАТИСТИКИ

Заведующая редакцией *Л.А. Табакова*
Редактор *Л.В. Сергеева*
Младший редактор *Н.А. Федорова*
Технический редактор *В.Ю. Фотиева*
Корректо *Г.В. Хлопцева*
Компьютерная верстка *Е.Ф. Тимохиной*
Обложка художника *О.В. Толмачева*

ИБ № 4456

Подписано в печать 20.08.2013. Формат 60×90/16
Гарнитура «Таймс». Печать офсетная
Усл. п.л. 40,18. Уч. изд.-л. 37,28
Тираж 400 экз. Заказ «С» 213

Издательство «Финансы и статистика»
101000, Москва, ул. Покровка, 7
Телефон (495) 625-35-02, факс (495) 625-09-57
E-mail: mail@finstat.ru http: // www.finstat.ru

ООО «Великолукская городская типография»
182100, Псковская область, г. Великие Луки,
ул. Полиграфистов, 78/12
Тел./факс: (811-53) 3-62-95
E-mail: zakaz@veltip.ru